Introduction
This unit deals with research methods used in psychology. In any discipline, knowledge of research methods greatly increases our ability to understand a topic. Psychology is not an exception. The ability to evaluate psychological knowledge critically on the basis of how it was obtained is essential to avoid misconceptions.

Speaking of misconceptions, there are plenty of them in this field. Psychology is a popular discipline which makes it vulnerable to numerous popular interpretations. So it is important to clearly understand what psychology is and what it is not.

Knowledge of methods also allows you to see the whole research process clearly, with all its strengths and limitations. When studying the material in this chapter, you will no longer take statements like “British psychologists have discovered ...” at face value. You will read between the lines and understand what was done by the “British psychologists” and to what extent their inferences are justified.

Psychology is a special discipline. On the one hand, it is scientific, which means that psychologists, just like physicists or chemists, rigorously test hypotheses and eliminate competing explanations in an attempt to achieve objective knowledge. On the other hand, unlike natural sciences that study “nature”, psychology studies humans, inherently subjective creatures. So psychology is an attempt to study the subjective (for example, the mind) objectively. Not an easy task, if you think about it.

This unit may seem a little abstract to you but it builds an important foundation for the understanding of the material in all other chapters. Applying the knowledge and skills related to research methodology, you will be able to critically evaluate knowledge in specific areas of psychology and arrive at balanced conclusions, avoiding misconceptions and unjustified generalizations. We will keep referring you back to this chapter so that you can apply and reinforce the concepts related to research methodology.

We start by discussing the definition of psychology, what it is and what it is not. Then we introduce two broad groups of research methods: quantitative and qualitative. These two groups of methods differ dramatically in their rationale and objectives, but at the same time can be combined to complement each other in a holistic investigation. Following this, we discuss four overarching concepts that apply to both quantitative and qualitative research: sampling, credibility, generalizability and bias. Next, we look at the application of these concepts separately in quantitative (experiments, correlational studies) and qualitative research. Finally, any discipline that involves research with living beings needs to adhere to the principles of ethics. We discuss ethical considerations in psychological research.
What is psychology?

“Psychology is the scientific study of behaviour and mental processes”. This is the definition we are going to use throughout this book. Although it is quite a short definition, there are a lot of implications in it. Let’s try and uncover them one by one.

*Psychology is the scientific study…* This part of the definition excludes such areas as pop psychology, that is, simple and appealing explanations that are not backed up by empirical evidence. What makes a theory or a study scientific, or where is the line between science and non-science? This is largely a TOK question and you will return to it throughout the book, but here are some major points:

1. It should be supported by empirical evidence and be based on this evidence.
2. It should be falsifiable, that is, it should be possible for the theory or study to be proven wrong.
3. There should be a history of independent attempts to test the theory or replicate the study.

<table>
<thead>
<tr>
<th>TOK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science versus non-science demarcation is one of the key topics in TOK. The following concepts are important in the discussion of demarcation criteria:</td>
</tr>
<tr>
<td>1. empirical evidence</td>
</tr>
<tr>
<td>2. falsification/falsifiability</td>
</tr>
<tr>
<td>3. replication.</td>
</tr>
</tbody>
</table>

While reading this chapter, take a note of examples that illustrate these three concepts.

Think of other similar examples from such areas of knowledge as human sciences, natural sciences and mathematics.
Exercise

Look at the following research questions and pick one that you find interesting:

1. Do children who watch more violent TV shows become more violent?
2. Does extrasensory perception exist?
3. Are women attracted to men by the smell of their body?
4. Is abuse experienced differently in heterosexual and gay relationships?
5. Are breathing exercises effective for reducing test anxiety?
6. What emotions do people experience when watching horror movies in a cinema?
7. Are people in arranged marriages happier than people who married by choice?

If you were to conduct a research study to answer the question that you picked, how would you go about it? Think about details such as who your participants would be, what they would be required to do, how you would measure results and how you would ensure that the results are believable.

Figure 1.1 Wilhelm von Osten and Clever Hans

In the early twentieth century, under the influence of Charles Darwin’s theory of evolution, the public was very interested in animal intelligence: if humans evolved from animals, animals must be at least partially intelligent, so what exactly are they capable of? The case of Clever Hans sparked a lot of interest. Hans was a horse. Its owner Wilhelm von Osten, a mathematics teacher, claimed that he had taught Hans to solve arithmetic problems (addition, subtraction, multiplication, division, fractions), read, spell and understand some German. Questions could be asked verbally or in writing, and Hans would respond by tapping his hoof a certain number of times. Von Osten exhibited the horse frequently and gained a lot of public attention. A special committee was formed in Germany (called the Hans Commission). They ran a series of tests and concluded that the performance was not a fraud. So Hans’s abilities were officially recognized as phenomenal!

However, another independent investigation carried out later by Oskar Pfungst, a psychologist, yielded different results. It demonstrated that Hans could not actually perform mental operations like multiplication, but the horse was very responsive to clues that were provided by unsuspecting humans. To arrive at these conclusions, Pfungst successively tested a number of alternative hypotheses:

1. What if spectators give the horse hints or clues? He tested the horse and the questioner in the absence of spectators, but the horse continued to solve tasks correctly anyway.
2. What if von Osten himself gives the horse some clues? Another questioner was used during several trials, but the horse’s performance did not worsen.
3. What if something in the questioner gives the correct answer away and the horse can feel that? Blinders were used to test this hypothesis. It turned out that when Hans was wearing blinders responses (the number of hoof taps) were incorrect most of the time. So, it was something in the questioner after all.
4. But did the questioners consciously let the horse know the correct answer? Additional trials were organized so that the questioner either knew or did not know the answer to the questions. It turned out that Clever Hans could only answer the questions correctly when the questioner knew the answer in advance.

This changed the focus of research from the horse to the questioner. When Pfungst carried out his observations, it was concluded that questioners who knew the answers had a tendency to become more tense as the hoof tapping approached the correct answer which would be reflected in their posture and facial expressions without them
realizing it. This was probably the clue that the horse was using. This makes sense evolutionarily, as detection of small postural changes is important as a survival skill for horses in the wild. Thus, Clever Hans certainly was clever, but the nature of his abilities was not mathematical (Goodwin, 2010)!

Von Osten himself, however, was never convinced of Pfungst’s findings and he continued to exhibit the horse throughout Germany, gaining as much popularity as before. Nonetheless, scientifically, this was one of the starting points for designing rigorous experimental methodology in psychology and other human sciences. It was recognized that experiments, if not carefully controlled, could produce artifacts—results that are associated with the effect of unforeseen factors.

But this whole story shows how claims can and should be tested scientifically, that is, by conducting a systematic evidence-based investigation that puts forward one hypothesis after another and tests them in a rigorous fashion. Note also how the whole investigation attempted to falsify the existing theory rather than support it.

... study of behaviour and mental processes. A scientific investigation requires an empirical approach to research, that is, relying on observation as a means of data collection. On the other hand, psychology (which comes from the Greek psyche = soul and logos = study, “the study of the soul”) concerns itself with a wealth of phenomena, many of which are not directly observable. The first step in solving this dilemma is to identify something that can be observed directly. That’s behaviour. Behaviour is everything that can be registered by an independent observer: it includes overt actions as well as gestures, facial expressions, verbal responses, endocrine reactions and so on. What stays “behind the scene” are the mental processes such as attention, perception, memory and thinking. We cannot observe them directly (which led some psychologists to say that they represent a “black box” and cannot be studied scientifically), but we can observe the indirect effects mental processes have on one’s behaviour. Thus we can infer something about the mental world as well.

**ATL skills: Thinking**

How does Pfungst’s investigation illustrate the concepts of empirical evidence, falsification and replication?

Brainstorm some behavioural indicators of the following:

- attention
- anxiety
- embarrassment.

To what extent do you think it is possible to use behavioural indicators to infer these “internal” phenomena? Would the inference be reliable?

Throughout this book we will use the term “behaviour” to refer to external, observable manifestations while the term “mental processes” will be used to denote internal patterns of information processing. However, you need to be aware of the fact that the term “behaviour” is often used in a more general sense, as an umbrella term for everything psychological. So sometimes you will encounter references to mental processes as types of “behaviour”. This is not exactly accurate, but acceptable.

Note that the definition of psychology does not specify human behaviour or mental processes. This is because research with non-human animals is also an integral part of psychology. Since humans are just a stage in the continuous process of evolution, the study of animals may inform our understanding of human behaviour (and mental processes).

IB psychology is an academic discipline with an emphasis on rigorous research and scientific knowledge, but psychology is broader than pure academics and research. When people think about psychology many imagine counselors and psychotherapists, practitioners who work with individual clients. University workers in lab coats conducting research is not the first thing that comes to mind. However, IB psychology focuses on academic knowledge and scientific research rather than counseling skills. This is because thorough understanding of psychological concepts and being able to think critically about psychological phenomena is of paramount importance in all spheres of psychology.
including counseling. It makes perfect sense to start with building these skills, much like the need to study aerodynamics before you are allowed to pilot an airplane.

**Research methodology: quantitative and qualitative methods**

All research methods used in psychology can be categorized as either quantitative or qualitative. Data in quantitative research comes in the form of numbers. The aim of quantitative research is usually to arrive at numerically expressed laws that characterize behaviour of large groups of individuals (that is, universal laws). This is much like the aim of the natural sciences in which it has been the ideal for a long time to have a set of simple rules that describe the behaviour of all material objects throughout the universe (think about laws of gravity in classic Newtonian physics, for example). In philosophy of science such orientation on deriving universal laws is called the nomothetic approach.

Quantitative research operates with variables. A variable (“something that can take on varying values”) is any characteristic that is objectively registered and quantified. Since psychology deals with a lot of “internal” characteristics that are not directly observable, they need to be operationalized first. Hence there’s an important distinction between constructs and operationalizations.

A construct is any theoretically defined variable, for example, violence, aggression, attraction, memory, attention, love, anxiety. To define a construct, you give it a definition which delineates it from other similar (and dissimilar) constructs. Such definitions are based on theories. As a rule constructs cannot be directly observed: they are called constructs for a reason—we have “constructed” them based on theory.

To enable research, constructs need to be operationalized. Operationalization of a construct means expressing it in terms of observable behaviour. For example, to operationalize verbal aggression you might look at “the number of insulting comments per hour” or “the number of swear words per 100 words in the most recent Facebook posts”. To operationalize anxiety you might look at a self-report score on an anxiety questionnaire, the level of cortisol (the stress hormone) in the bloodstream or weight loss. As you can see, there are usually multiple ways in which a construct may be operationalized: the researcher needs to use creativity in designing a good operationalization that captures the essence of the construct and yet is directly observable and reliably measurable. As you will see throughout examples in this book, it is often a creative operationalization that makes research in psychology outstanding.

**ATL skills: Research and communication**

In small groups brainstorm operationalizations of the following constructs: belief in God, assertiveness, shyness, pain, love, friendship, prejudice, tolerance to uncertainty, intelligence, wisdom.

Is it equally easy to operationalize them?
Discuss each other’s operationalizations and outline their strengths and limitations

There are three types of quantitative research.

1. **Experimental studies.** The experiment in its simplest form includes one **independent variable** (IV) and one **dependent variable** (DV), while the other potentially important variables are controlled. The IV is the one manipulated by the researcher. The DV is expected to change as the IV changes. For example, if you want to investigate the effect of psychotherapy on depression, you might randomly assign participants to two groups: the experimental group will receive psychotherapy while the control group will not. After a while you might measure the level of depression by conducting a standardized clinical interview (diagnosis) with each of them. In this case the IV is psychotherapy. You manipulate the IV by changing its value: yes or no. The DV is depression; it is operationalized through the standardized diagnostic procedure. If the DV is different in the two groups, you may conclude that a change in the IV “caused” a change in the DV. This is why the experiment is the only method that allows cause-and-effect inferences.
2. **Correlational studies.** Correlational studies are different from experiments in that the researcher does not manipulate any variables (there are no IVs or DVs). Variables are measured and the relationship between them is quantified. For example, if you want to establish if there is any relationship between violent behaviour of adolescents and how much time they spend watching violent TV shows, you may recruit a sample of adolescents and measure their violent behaviour (by self-report, by ratings from classmates or even by observation in a natural setting) and the average number of hours per day spent watching violent TV shows. Then you can correlate these two variables using a formula. Suppose you obtained a large positive correlation. This means that there’s a trend in the data: the more time an adolescent spends watching violent shows, the more violent he or she is. However, you cannot make cause-and-effect inferences from correlational studies. Since you did not manipulate one of the variables, you do not know the direction of influence. It could be the case that watching violence influences violent behaviour (this would probably be the most popular, intuitive assumption). But it is also possible that adolescents who behave violently choose to watch violent TV programmes. Or there could even be a third variable (for example, low self-esteem) that influences both violent behaviour and watching violence on TV. What you observe “on the surface” is just that—“co-relation”, the fact that one variable changes as the other one changes.

**ATL skills: Communication and social**

In small groups brainstorm results of fictitious studies that would demonstrate either correlation or causation. For example:

1. In a group of adults we measured their attitudes to horror films and the number of siblings they have. We found that the more siblings you have, the more you like horror films.

2. We told one group of astronauts that their mission would start in a month and the other group that the mission would start in a year. We measured anxiety and found that it was higher in the group of astronauts who expected the mission to start in a month.

As you go through your list of fictitious studies, the other groups will have to say whether the study shows correlation or causation.

3. **Descriptive studies.** In descriptive studies relationships between variables are not investigated, and the variables are approached separately. An example of a descriptive quantitative study would be a public opinion survey. We ask questions (for example, “Do you support the current policies of the President?”) and we are interested in the distribution of answers to this particular question. Descriptive studies are often used in sociology and they are sometimes used in psychology to conduct a broad investigation of a phenomenon before “delving deeper” into the specifics.

**Qualitative research** is different. Its main focus is an in-depth study of a particular phenomenon. “In-depth” entails going beyond what can be objectively measured and quantified into the realm of human experiences, interpretations and meanings. Qualitative research makes use of such data collection methods as interviews or observations. Data comes in the form of texts: interview transcripts, observational notes, and so on. Interpretation of data involves a degree of subjectivity, but analysis is deeper than we can usually achieve through quantitative approaches. In philosophy of science such
orientation on an in-depth analysis of a particular case or phenomenon (without trying to derive universally applicable laws) is called the idiographic approach.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Quantitative research</th>
<th>Qualitative research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aim</td>
<td>Nomothetic approach: derive universally applicable laws</td>
<td>Idiographic approach: in-depth understanding of a particular case or phenomenon</td>
</tr>
<tr>
<td>Data</td>
<td>Numbers</td>
<td>Texts</td>
</tr>
<tr>
<td>Focus</td>
<td>Behavioural manifestations (operationalizations)</td>
<td>Human experiences, interpretations, meanings</td>
</tr>
<tr>
<td>Objectivity</td>
<td>More objective (the researcher is eliminated from the studied reality)</td>
<td>More subjective (the researcher is included in the studied reality)</td>
</tr>
</tbody>
</table>

Qualitative research methods that we will discuss in this chapter are:
- observation
- interview
- focus group
- case study
- content analysis.

### Sampling, credibility, generalizability and bias in research

Sampling, credibility, generalizability and bias are some of the characteristics used to describe a research study and make a judgment of its quality. These characteristics are universal for social sciences, but they can be approached very differently by quantitative and qualitative researchers, sometimes even with distinctly different sets of terms to express the same ideas. So it is important that you understand both these overarching concepts and the way they are broken down in quantitative as compared to qualitative research. Let us start with the overarching concepts.

**A sample** is the group of individuals taking part in the research study. **Sampling** is the process of finding and recruiting individuals for the study. There are different sampling techniques, and it is important to be aware of their strengths and limitations as sampling may affect the results of the study. For example, if the aim of your research is to see if anxiety correlates with aggression in teenagers (in general), but you only sample teenagers from one school in a criminal neighbourhood, your sampling technique will have important implications for the conclusions you will be able to make. Similarly, if you study political views of unemployed people and you recruit your sample by asking a small number of participants to bring their friends (and possibly friends of friends), you might end up with a limited sample because people of similar political views are more likely to be friends with each other.

**Credibility** refers to the degree to which the results of the study can be trusted to reflect the reality. It is closely linked to bias, because the results of the study do not reflect reality if there was some sort of bias in it. There are a lot of “traps” that a researcher can walk into. For example, in an interview, while the researcher believes the interviewee’s responses to be true, participants may actually guess the aim of the study and respond in a way that they think the researcher is expecting them to. Or researchers themselves, being interested in confirming their hypothesis, may selectively notice supporting evidence and unintentionally ignore contradicting evidence. If there is indication that potential sources of bias were, to the best of our knowledge and abilities, controlled or eliminated, credibility of the research study is believed to be high. Quantitative and qualitative research approaches to credibility and bias are distinctly different, although they overlap in some aspects.

**Generalizability** refers to the extent to which the results of the study can be applied beyond the sample and the settings used in the study itself. Sometimes, especially in quantitative research, you want to generalize findings from the sample to a much wider group of people (called “population”) because your aim is to discover universal laws of behaviour. Sometimes the research study is conducted in artificial settings...
(for example, a laboratory experiment), but you want to believe that people will behave the same way in their natural setting in daily life too. In any case generalizability is an important aspect in the interpretation of findings. Again, the ways quantitative and qualitative research studies approach generalizability of findings is distinctly different.

The table below gives you an overview of the main concepts used to characterize sampling, generalizability, credibility and bias in experimental, correlational and qualitative research. As you read on, you will understand these concepts better. Refer to this table from time to time so that you place them clearly in the general framework.

### Overview table:

**Sampling, generalizability, credibility and bias in qualitative and quantitative research**

<table>
<thead>
<tr>
<th>Overarching concepts</th>
<th>Quantitative research</th>
<th>Qualitative research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sampling</strong></td>
<td>Experimental studies</td>
<td>Correlational studies</td>
</tr>
<tr>
<td></td>
<td>Same</td>
<td>Quota sampling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Purposive sampling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theoretical sampling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Snowball sampling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Convenience sampling</td>
</tr>
<tr>
<td><strong>Generalizability</strong></td>
<td>External validity:</td>
<td>Population validity</td>
</tr>
<tr>
<td></td>
<td>– Population validity</td>
<td>Construct validity</td>
</tr>
<tr>
<td></td>
<td>– Ecological validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Construct validity</td>
<td></td>
</tr>
<tr>
<td><strong>Credibility</strong></td>
<td>Internal validity:</td>
<td>No special term used:</td>
</tr>
<tr>
<td></td>
<td>– To what extent is the DV used: “validity” and</td>
<td>“validity” and</td>
</tr>
<tr>
<td></td>
<td>influenced by the IV and</td>
<td>“credibility” can be used</td>
</tr>
<tr>
<td></td>
<td>not some other variable?</td>
<td>interchangeably</td>
</tr>
<tr>
<td></td>
<td>Controlling confounding</td>
<td>Credibility is high if</td>
</tr>
<tr>
<td></td>
<td>variables: eliminating or</td>
<td>no bias occurred</td>
</tr>
<tr>
<td></td>
<td>keeping constant in all conditions</td>
<td></td>
</tr>
<tr>
<td><strong>Bias</strong></td>
<td>Threats to internal validity:</td>
<td>On the level of</td>
</tr>
<tr>
<td></td>
<td>– Selection</td>
<td>measurement of</td>
</tr>
<tr>
<td></td>
<td>– History</td>
<td>variables: depends</td>
</tr>
<tr>
<td></td>
<td>– Maturation</td>
<td>on the method of</td>
</tr>
<tr>
<td></td>
<td>– Testing effect</td>
<td>measurement</td>
</tr>
<tr>
<td></td>
<td>– Instrumentation</td>
<td>On the level of</td>
</tr>
<tr>
<td></td>
<td>– Regression to the mean</td>
<td>interpretation of</td>
</tr>
<tr>
<td></td>
<td>– Experimental mortality</td>
<td>findings:</td>
</tr>
<tr>
<td></td>
<td>– Experimenter bias</td>
<td>– Curvilinear</td>
</tr>
<tr>
<td></td>
<td>– Demand characteristics</td>
<td>relationships</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– The third variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>problem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Spurious correlations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant bias:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Acquiescence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Social desirability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Dominant respondent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Sensitivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Researcher bias:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Confirmation bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Leading questions bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Question order bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Sampling bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Biased reporting</td>
</tr>
</tbody>
</table>
Confounding variables

As we mentioned, the experiment is the only method that allows researchers to make cause-and-effect inferences. This is achieved by defining the independent variable (IV) and the dependent variable (DV), manipulating the IV and observing how the DV changes in response to this manipulation.

Psychological reality is very complex and the trick is to isolate the IV so that when you manipulate it, nothing else changes. Imagine, for example, that you manipulate X and observe the resulting changes in Y. But every time you manipulate X, you also unintentionally change Z. In reality it is Z that causes a change in Y, but you incorrectly conclude that X (your independent variable) is the cause of Y, thus incorrectly confirming your hypothesis. If this sounds too abstract, think about the following example: X is sleep deprivation (which you manipulate by waking up one group of participants every 15 minutes when they sleep, while the control group sleeps normally) and Y is memory performance (which you measure by a simple memory test in the morning). But without realizing that this might be an important factor, you let the control group sleep at home while the experimental group sleeps in a laboratory being
supervised by an experimenter. So there’s another variable, variable Z: stress caused by the unfamiliar environment. It could be the case that in this experiment it was the unfamiliar environment (Z) that caused a reduction in memory performance (Y), rather than sleep deprivation (X).

Variables that can potentially distort the relationship between the IV and the DV (like Z in the example above) are called confounding variables. They contribute to bias. These variables need to be controlled, either by eliminating them or keeping them constant in all groups of participants so that they do not affect the comparison.

Discussion

How could the researchers have controlled the confounding variable in this example?

Exercise

Imagine you are investigating the influence of praise on the school performance of teenagers. For this experiment you need to have a sample of participants that you would split into two groups (experimental and control). In the experimental group the teacher is instructed to praise every student three times a week while in the control group the teacher is told to only praise the students once every week. At the end of the research period performance grades in the two groups are compared.

Suppose that the participants in this experiment are high school students from one of the schools in your city. Will you be able to generalize the findings to the target population, that is, teenagers in general? This depends on how representative your sample is. For this you need to take into account your target population and the aim of the research.

1. The aim of the research links to the participant characteristics that are essential. Whatever can theoretically influence the relationship between the IV and the DV is essential. For example, cultural background may be essential for how a teenager reacts to praise (depending on their cultural attitudes to adults, teachers and authority in general). Socio-economic background may be important as well: theoretically there may be a connection between the socio-economic status of a teenager’s family and their value of education. The type of school is another potentially important factor: in top schools where children pursue quality education and prestigious college placements teachers’ praise may be a point of pride, whereas in public schools in criminal neighbourhoods it may lead to bullying from classmates.

2. If the sample is representative, it must reflect the essential characteristics of the target population. Is the sample of teenagers from one school in our example sufficient to reflect all these characteristics? No, because it does not represent the variation of cultural backgrounds, socio-economic backgrounds and types of schools found in the population.

3. If the sample is not representative of the essential characteristics of the target population, there are two ways to fix it: either keep sampling or narrow down the target population and do not claim that the research findings are more generalizable than they really are.

Given the aim of the study, how would you increase representativeness of your sample?
There is no quantitative way to establish representativeness of a sample and it is always the expert decision of a researcher to say whether a particular characteristic is essential or not. This is done on the basis of prior knowledge from published theories and research studies. In any case the choice of the target population needs to be well justified and explicitly explained.

Several sampling techniques can be used in an experiment. The choice depends on the aim of the research, available resources and the nature of the target population.

1. **Random sampling.** This is the ideal approach to make the sample representative. In random sampling every member of the target population has an equal chance of becoming part of the sample. With a sufficient sample size this means that you take into account all possible essential characteristics of the target population, even the ones you never suspected to play a role. Arguably, a random sample of sufficient size is a good representation of a population, making the results easily generalizable. However, random sampling is not always possible for practical reasons. If your target population is large, for example, all teenagers in the world, it is impossible to ensure that each member of this population gets an equal chance to enter your sample. Being based in Europe, you can’t just create a list of all teenagers in the world, randomly select a sample and then call Lynn from Fiji to come and join your experiment. In this case you either believe that cross-cultural differences are not essential (for your hypothesis) or narrow down your target population. On the other hand, if your target population is students from your school, it is perfectly possible to create the full list of students and select your participants randomly from this list. An example of random sampling strategy is a pre-election telephone survey where participants are selected randomly from the telephone book (or a random selection of Facebook profiles). But even in this case you have to admit that the target population is not all the citizens of a particular country; it is all the citizens of the country who own a telephone (or have a Facebook profile).

2. **Stratified sampling.** This approach is more theory-driven. First you decide on the list of essential characteristics that the sample has to reflect. Then you study the distribution of these characteristics in the target population (for this you may use statistical data available from various agencies). Then you recruit your participants in a way that keeps the same proportions in the sample as is observed in the population. For example, imagine that your target population is all the students of your school. Participant characteristics that you decide are important for the aim of the study are age (primary school, middle school, high school) and GPA (low, average, high). You study your school records and find out the distribution of students across these categories:

<table>
<thead>
<tr>
<th></th>
<th>Low GPA</th>
<th>Average GPA</th>
<th>High GPA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary school</td>
<td>0%</td>
<td>10%</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>Middle school</td>
<td>5%</td>
<td>30%</td>
<td>15%</td>
<td>50%</td>
</tr>
<tr>
<td>High school</td>
<td>5%</td>
<td>20%</td>
<td>5%</td>
<td>30%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10%</td>
<td>60%</td>
<td>30%</td>
<td>100%</td>
</tr>
</tbody>
</table>

For a stratified sample you need to ensure that your sample has the same proportions. For every cell of this table you can either sample randomly or use other approaches (see below). In any case, what makes stratified sampling special is that it is theory-driven and it ensures that theory-defined essential characteristics of the population are fairly and equally represented in the sample. This may be the ideal choice when you are certain about essential participant characteristics and when available sample sizes are not large.

3. **Convenience (opportunity) sampling.** For this technique you recruit participants that are more easily available. For example, university students are a very popular choice because researchers are usually also university professors so it is easy for them to find samples there. Jokingly, psychology has been sometimes referred to as a study of “US college freshmen and white rats”. There could be several reasons for choosing convenience sampling. First, it is the technique of choice when financial resources and time are limited. Second, there
could be reasons to believe that people are not that different in terms of the phenomenon under study. For example, if you study the influence of caffeine on attention, there are reasons to believe that results will be similar cross-culturally, and it might be a waste of time to use a stratified or a random sample. Finally, convenience sampling is useful when wide generalization of findings is not the primary goal of your research, for example, if you are conducting an exploratory study and you are not sure the hypothesis will be supported by evidence. If the hypothesis will not “work” in a small sample, why waste time testing it in a representative sample? Or you are replicating someone else’s research and your aim is to see if the universal law (that was discovered by this someone) will hold true in your specific sample, thus trying to falsify prior theory. The limitation of convenience sampling is, of course, lack of representativeness.

4. **Self-selected sampling.** This refers to recruiting volunteers. An example of this approach is advertising the experiment in a newspaper and using the participants who respond to the advert. The strength of self-selected sampling is that it is a quick and relatively easy way to recruit individuals while at the same time having wide coverage (many different people read newspapers). The most essential limitation, again, is representativeness. People who volunteer to take part in experiments may be more motivated than the general population, or they may be looking for the incentives (in many studies participants are financially rewarded for their time).

**Exercise**

Now that you know what sampling strategies can be used in an experiment, how would you change your approach to recruiting a sample for the investigation of the influence of praise on school performance of teenagers?

**Experimental designs**

Experiments always involve manipulating some variables and measuring the change in others. But the specific ways in which this can be organized differ depending on the aims of the research. The organization of groups and conditions in an experiment is known as the experimental design, and there are three basic types of experimental design.

**Independent measures design** involves random allocation of participants into groups and a comparison between these groups. In its simplest form, you randomly allocate participants from your sample into the experimental group and the control group. Then you manipulate the experimental conditions so that they are the same in the two groups except for the independent variable. After the manipulation you compare the dependent variable in the two groups.

**ATL skills: Research**

Consider the difference between random sampling (selecting the sample from the target population) and random group allocation (dividing your sample into groups). It is possible to have random group allocation in non-random samples and vice versa.

The rationale behind random group allocation is that all potential confounding variables cancel each other out. If the groups are not equivalent at the start of the experiment, you will be comparing apples to oranges. Imagine that you are testing the hypothesis that praise at school improves students’ performance and for this you take two existing groups of students, with one being rarely praised by their teachers and the other one often praised. Arguably, the groups might not be equivalent: they have different experiences with the teachers, different in-group values and habits, and so on. But to account for all these potentially important factors is impossible.

Conversely, when the group sizes are sufficiently large and allocation is completely random, chances are that groups will be equivalent—the larger the sample, the higher the chance.

Of course, there could be more than two groups, depending on how many IVs you use and how many levels each variable has. In the above example, you could use more than one IV: the influence of praise and the allocation of homework on school performance. With two levels for each of these IVs you would need to randomly allocate participants into four groups:
This experimental design with two IVs, each with two levels, is quite frequently used in psychological experiments. It is known as a $2 \times 2$ experimental design. Of course you can think of other combinations: $2 \times 3$ (two IVs, three levels in each), $3 \times 2$ (three IVs, two levels in each), $4 \times 4$ (four IVs, four levels in each). But the more cells you have in this table, the larger the sample you need, so at some point it becomes impractical to increase the number of groups.

To summarize, regardless of the number of IVs and their levels, an experiment follows an independent measures design when the IV is manipulated by randomly allocating participants into groups. This allows us to assume that the groups are equivalent from the start so whatever difference we observe at the end of the experiment must have been caused by our experimental manipulation.

**Matched pairs design** is similar to independent measures. The only difference is that instead of completely random allocation, researchers use matching to form the groups.

To illustrate matching, let’s consider an example. Suppose you are conducting a study of the effect of sleep deprivation on memory. For this you need two groups of participants. One of the groups will sleep peacefully in the laboratory and the other group will be woken up every 15 minutes. In the morning you will give both groups a memory test and compare their performance. You suspect that there is one confounding variable that may influence the results: memory abilities. Some people generally have better memory than others, therefore it is important to you that the two groups at the start of the experiment are equivalent in their memory abilities. Random allocation will usually make that happen, but you only have 20 participants (10 in each group). With a small sample like this there is a chance that random allocation will not work. So you want to control the equivalence of memory abilities “manually” while leaving everything else to random chance. For this you test memory abilities in your participants prior to the experiment. Then you rank participants according to their memory abilities (for example, from the highest to the lowest). Then you take the first two participants from the top of the list and randomly allocate one of them to the experimental group and the other one to the control group. You take the next two participants and repeat the procedure for the rest of the list. The two resulting groups are certainly equivalent in terms of memory abilities and probably (due to random chance) equivalent in all other characteristics.

![Figure 1.3 Matched pairs design](image)

The variable which is controlled (memory abilities in the example above) is called the **matching variable**. Matched pairs designs are preferred when:
- the researcher finds it particularly important that the groups are equivalent in a specific variable
- the sample size is not large, therefore there is a chance that random allocation into groups will not be sufficient to ensure group equivalence.

**Repeated measures design** is used when the goal is to compare conditions rather than groups of participants. The same group of participants is exposed to two (or more) conditions, and the conditions are compared. For example, imagine your aim is to investigate the effect of classical music on learning. You ask your participants to learn a list of trigrams (meaningless combinations of three letters such as HPX, LJW) for 10 minutes in silence and register the number of trigrams correctly recalled. Then you ask the same participants to learn a different list of trigrams for another 10 minutes, but this time with classical music playing in the background. You compare results from the first and the second trial.

The problem with repeated measures designs is that they are vulnerable to **order effects**: results...
may be different depending on which condition comes first (for example, silence then classical music or classical music then silence). Order effects may appear due to various reasons, such as:

- practice: participants practise, improve their on-task concentration and become more comfortable with the experimental task during the first trial, thus increasing performance in the second trial
- fatigue: participants get tired during the first trial, and their concentration decreases, thus reducing performance in the second trial.

To overcome order effects researchers use **counterbalancing**. Counterbalancing involves using other groups of participants where the order of the conditions is reversed. For our example two groups could be used: one given the sequence “silence then music” and one given the sequence “music then silence”. It is important to note that comparison will still be made between conditions, not between groups. Data from Group 1 Condition 1 will be collated with data from Group 2 Condition 2, and vice versa. These two collated data sets will be compared.

![Diagram](chart.png)  
**Figure 1.4 Counterbalancing**

An advantage of repeated measures designs is that people are essentially compared to themselves, which overcomes the influence of **participant variability** (differences between the groups prior to the start of the experiment). It makes the comparison more reliable. Another advantage that follows from this is that smaller sample sizes are required.

**Credibility and generalizability in the experiment: types of validity**

As you have seen, credibility and generalizability are overarching terms that are used to characterize the quality of research studies. When it comes to experiments specifically, these terms are very rarely used. Instead the quality of experiments is characterized by their construct, internal and external validity.

**Construct validity** characterizes the quality of operationalizations. As you know, the phenomenon under study is first defined theoretically as a construct and then expressed in terms of observable behaviour (operationalization). Operationalization makes empirical research possible. At the same time when results are interpreted research findings are linked back to constructs. Moving from an operationalization to a construct is always a bit of a leap. Construct validity of an experiment is high if this leap is justified and if the operationalization provides sufficient coverage of the construct. For example, in some research studies anxiety was measured by a fidgetometer, a specially constructed chair that registers movements at various points and thus calculates the amount of “fidgeting”. Subjects would be invited to the laboratory and asked to wait in a chair, not suspecting that the experiment has already started. The rationale is that the more anxious you are, the more you fidget in the chair. Are the readings of a fidgetometer a good operationalization of anxiety? On the one hand, it is an objective measure. On the other hand, fidgeting may be a symptom of something other than anxiety. Also the relationship between anxiety and increased fidgeting first has to be demonstrated in empirical research.

**Internal validity** characterizes the methodological quality of the experiment. Internal validity is high when confounding variables have been controlled and we are quite certain that it was the change in the IV (not something else) that caused the change in the DV. In other words, internal validity links directly to bias: the less bias, the higher the internal validity of the experiment. Biases in the experiment (threats to internal validity) will be discussed below.

**External validity** characterizes generalizability of findings in the experiment. There are two types of external validity: population validity and ecological validity. **Population validity** refers to the extent
to which findings can be generalized from the sample to the target population. Population validity is high when the sample is representative of the target population and an appropriate sampling technique is used. **Ecological validity** refers to the extent to which findings can be generalized from the experiment to other settings or situations. It links to the artificiality of experimental conditions. In highly controlled laboratory experiments subjects often find themselves in situations that do not resemble their daily life. For example, in memory experiments they are often asked to memorize long lists of trigrams. To what extent can findings from such studies be applied to everyday learning situations?

There is an inverse relationship between internal validity and ecological validity. To avoid bias and control for confounding variables, you make the experimental procedures more standardized and artificial. This reduces ecological validity. Conversely, in an attempt to increase ecological validity you may allow more freedom in how people behave and what settings they choose, but this would mean that you are losing control over some potentially confounding variables.

### Figure 1.5 Validity of experiments

<table>
<thead>
<tr>
<th>Internal</th>
<th>External</th>
<th>Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>To what extent is the change in DV caused by IV?</td>
<td>Population</td>
<td>To what extent do the operationalizations reflect the construct?</td>
</tr>
<tr>
<td>Credibility</td>
<td>Ecological</td>
<td>Generalizability (to theory)</td>
</tr>
<tr>
<td>To what extent can the findings be generalized to the wider population?</td>
<td>To what extent can the findings be generalized to real-life settings?</td>
<td>Generalizability (to other people)</td>
</tr>
<tr>
<td>Generalizability (to other people)</td>
<td>Generalizability (to other situations)</td>
<td></td>
</tr>
</tbody>
</table>

### Exercise

- Leaf through this book (consider the chapters on the biological, cognitive or sociocultural approach to behaviour), find a description of any experimental study and analyse its construct, internal and external validity. If you feel like you do not have enough detail, you could find more information on the study online, or even read the original article.
- Present the results of your analysis in class.

### Bias in experimental research: threats to internal validity

Bias in experimental research comes in the form of confounding factors that may influence the cause-and-effect relationship between the IV and the DV, decreasing internal validity. Below you will find a description of several common sources of threat to internal validity, based on Campbell (1969).

### Figure 1.6 Sources of threat to internal validity

- Selection
- Demand characteristics
- Mortality
- Regression to the mean
- Instrumentation
- History
- Maturation
- Testing effect
- Experimenter bias
- Regression to the mean

▲ Figure 1.6 Sources of threat to internal validity
1. **Selection.** This occurs if for some reason groups are not equivalent at the start of the experiment: apart from the planned IV-related difference, they differ in some other variable. As a result, we cannot be sure if the post-experiment differences between groups reflect the influence of the IV or this other variable. Selection occurs in independent measures and matched pairs designs in case group allocation was not completely random.

2. **History.** This refers to outside events that happen to participants in the course of the experiment. These outside events become a problem when they can potentially influence the DV or are not evenly distributed in the comparison groups. History is especially important in lengthy experiments where the DV is measured sometime after the onset of the study. For an example of history-related bias think of a memory experiment where participants are required to memorize long lists of words and the experiment is conducted in two groups (experimental and control) simultaneously in two different rooms on the opposite sides of a school. As the experiment begins, there is some noise coming from road construction outside. The control group is closer to the construction site so the noise in their room is louder. Since distracting noise can affect memory performance and levels of noise were not equal in the two groups, resulting differences in the DV may reflect the influence of the IV as well as the confounding variable (noise). To counteract history as a threat to internal validity such confounding variables should be either eliminated or kept constant in all comparison groups (for example, change the rooms so that they are both on the same side of the school building).

3. **Maturation.** In the course of the experiment participants go through natural developmental processes, such as fatigue or simply growth. For example, suppose you are piloting a psychological training programme to increase assertiveness in middle school students. You measure assertiveness at the start, conduct the training programme for several months and measure assertiveness again. The resulting increase of assertiveness may be due to either the IV (the training) or simply to the fact that the middle school students grew up a little and naturally became more assertive. The counteracting strategy would be using a control group (the same time period, the same measurements but no training sessions).

4. **Testing effect.** The first measurement of the DV may affect the second (and subsequent) measurements. For example, suppose you are investigating the effectiveness of a video to reduce test anxiety in primary school children. For this your participants take an ability test preceded by a self-report anxiety measure at time A. They then watch your specially designed video and repeat the procedure (test and self-report anxiety measure) at time B. The difference in anxiety between time A and time B may be the result of both the video and the fact that it is their second time taking the test—they are more familiar with the format and therefore may be naturally less anxious. A solution to this is to use a control group where you show a neutral video of the same duration. Suppose you get the following results:

<table>
<thead>
<tr>
<th>Group</th>
<th>Before Test 1</th>
<th>Before Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental (specially</td>
<td>90</td>
<td>55</td>
</tr>
<tr>
<td>designed video)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control (neutral video)</td>
<td>90</td>
<td>70</td>
</tr>
</tbody>
</table>

Analysis of these results can reveal that a reduction of anxiety by 20 points is probably due to the testing effect, however over and above that there is a 15-point anxiety effect of the specially designed video.

In repeated measures designs testing effect is a special case of order effects, and counterbalancing is used to control for it.

5. **Instrumentation.** This effect occurs when the instrument measuring the DV changes slightly between measurements. For psychology this becomes relevant when you consider that an “instrument of measurement” is often a human observer. Suppose you are investigating bullying on school campus during breaks. You are looking at two groups of students who are exposed to different experimental conditions. If you observe group 1 in the morning and
group 2 in the afternoon, you might be more tired in the afternoon and miss some important behaviours. If you observe one of the groups during a short break and the other one during the lunch break, observations during the lunch break may be less accurate because it is more crowded. To avoid this researchers should try to standardize measurement conditions as much as possible across all comparison groups and all observers.

6. **Regression to the mean.** This is an interesting source of bias that becomes a concern when the initial score on the DV is extreme (either low or high). Extreme scores have a purely statistical tendency to become more average on subsequent trials. Suppose you have designed anxiety reduction training for students. To test its effectiveness, you administer an anxiety questionnaire in a group of students and select a sample of students who have the largest score (for example, 80–100 on a 100-point scale). With these students you then conduct your training session and measure their anxiety again. Even if we assume that testing effects are not an issue, we would expect extremely anxious students to naturally become less anxious even without the training session. To put it more precisely, the probability that extremely anxious students will become even more anxious is less than the probability that they will become less anxious. Hence statistically a reduction of anxiety should be expected. A counter-measure is a control group with the same starting average anxiety level and measurements at the same point of time, but without the intervention.

7. **Experimental mortality.** This refers to the fact that some participants drop out during an experiment which may become a problem if dropouts are not random. Suppose you are investigating the influence of emotion on ethical decision-making. For this you give your participants a number of scenarios of the type “Would you kill 1 person to save 1000?” In the control group the description of this “one person” is neutral, but in the experimental group this is someone they know personally, so there is more emotional involvement. You hypothesize that people will be less likely to be utilitarian in their decision-making when they are personally involved (note that this research would create distress among participants and thus raises ethical issues; it is quite possible such a study would not be approved by the ethics committee). Suppose that several participants in the experimental group refuse to continue participation and drop out, more so than in the control group. Ethical issues aside, this presents a methodological issue as well: even if the two groups were equivalent at the start of the experiment, they may be non-equivalent now. There appears a confounding variable (sensitivity) which is disproportionally represented in the two groups. There is no reliable way to counteract experimental mortality other than designing experimental conditions in such a way that participants would not feel the need to drop out.

8. **Demand characteristics.** This refers to a situation in which participants understand the purpose of the experiment and change their behaviour subconsciously to fit that interpretation. In other words, they behave in ways that they think the experimenter expects. This can happen for various reasons, for example, participants may feel that they will somehow be evaluated and thus behave in a socially desirable way. To avoid demand characteristics, deception may be used to conceal the true purpose of the study (however, deception raises ethical issues, see below). You can consider using post-experimental questionnaires to find out to what extent demand characteristics may have influenced the results (this strategy does not prevent demand characteristics but just estimates their impact). Note that in repeated measures designs demand characteristics are a larger threat because participants take part in more than one condition and hence have greater opportunities to figure out or guess the aim of the study.

9. **Experimenter bias.** This refers to situations in which the researcher unintentionally exerts an influence on the results of the study, for example, the Clever Hans case discussed above. Existence of this bias was first rigorously supported by Rosenthal and Fode (1963). In this experiment rats were studied for their maze-running performance. Rats were split into two groups completely at random, but the lab assistants (psychology students) were told that one of the groups was “maze-bright” and
the other one was “maze-dull” and that this difference in ability was genetic. Lab assistants had to follow a rigorous and standardized experimental procedure in which rats were tested on their performance in learning the maze task. This was supposed to be an identical study conducted with identical rats, but results actually showed that the rats with the “maze-dull” label performed significantly worse than their counterparts labeled “maze-bright”. It was concluded that the result was an artifact: it was caused by experimenter bias rather than any genuine differences between the groups of rats. Post-experiment investigations revealed that experimenter bias was not intentional or conscious. The results were induced by subtle differences in the way lab assistants handled the rats. For example, when working with “maze-bright” rats, lab assistants, without realizing it, handled them slightly longer and thus reduced their stress more. A counter-measure against experimenter bias is using so-called double-blind designs where information that could introduce bias is withheld both from the participants and from the people conducting the experiment. The study of Rosenthal and Fode would have been double-blind if the lab assistants had not been told which group of rats had which label.

Exercise

Once again leaf through this book and find a description of any experimental study.

- To what extent was this experimental study susceptible to one of the sources of threat to internal validity? What does it tell you about credibility of the study?
- If you do not have enough detail, find more information on the study online, or even read the original article.
- Present the results of your analysis in class.

ATL skills: Self-management

Athabasca University has a great learning resource on threats to internal validity. One tutorial consists of two parts, where part 1 is the theoretical background and definitions and part 2 is a practical exercise involving the analysis of 36 hypothetical experiments.

If you want to practise identifying potential sources of bias in experiments, you can access the tutorial here: https://psych.athabascau.ca/open/validity/index.php

Quasi-experiments versus true experiments

Quasi-experiments are different from “true” experiments in that the allocation into groups is not done randomly. Instead some pre-existing inter-group difference is used. “Quasi” is a prefix meaning “almost”. The major limitation of a quasi-experimental design is that cause-and-effect inferences cannot be made. This is because we cannot be sure of the equivalence of comparison groups at the start of the study: pre-existing differences in one variable may be accompanied by a difference in unexpected confounding variables.

Suppose your hypothesis is that anxiety influences test performance. You have an opportunity sample of high school students. An intuitively obvious way to test this hypothesis would be to administer an anxiety questionnaire, divide the sample into two groups (anxious and non-anxious) based on the results, and then model a testing situation and compare test performance. The IV in this study is anxiety (it has two levels) and the DV is test performance. However, the researcher does not really manipulate the IV in this study. Pre-existing differences in anxiety are used, so we cannot be sure that anxiety is the only variable that differs in the two groups. For example, it is possible that high school students with high levels of anxiety also tend to have unstable attention, and it is actually attention that influences test performance. The bottom line is that we will be able to conclude that “anxiety is linked to test performance”, but strictly speaking we will not be able to say “anxiety influences test performance”.

To test the “influence” hypothesis a true experiment would be required, so we would have to manipulate the IV. How can you manipulate anxiety? One example is splitting participants randomly into two groups and telling one of the
groups that they should expect results of their college applications later today. Anticipation of these results would probably increase anxiety in the experimental group. Then the test can be given. (Note that such an experiment would have ethical issues since it involves major deception and creates distress among participants.)

Other examples of pre-existing differences are age, gender, cultural background and occupation. Formation of experimental groups based on these variables implies a quasi-experiment. Sometimes a “true” experiment cannot be conducted because it is impossible to manipulate the IV (for example, how do you manipulate age or gender?) so quasi-experiments are justified.

In the way they are designed (superficially) quasi-experiments resemble “true” experiments, but in terms of the possible inferences (essentially) they are more like correlational studies.

**Field experiments and natural experiments**

Field experiments are conducted in a real-life setting. The researcher manipulates the IV, but since participants are in their natural setting many extraneous variables cannot be controlled. The strength of field experiments is higher ecological validity as compared to experiments in a laboratory. The limitation is less control over potentially confounding variables so there is lower internal validity. An example of a field experiment is Piliavin et al’s (1969) subway study in which the researchers pretended to collapse on a subway train and observed if other passengers would come to help. To manipulate the IV, some researchers were carrying a cane (the cane condition) while others were carrying a bottle (the drunk condition).

Natural experiments, just like field experiments, are conducted in participants’ natural environment, but here the researcher has no control over the IV—the IV occurred naturally. Ecological validity in natural experiments is an advantage and internal validity is a disadvantage due to there being less control over confounding variables. Another advantage of natural experiments is that they can be used when it is unethical to manipulate the IV, for example, comparing rates of development in orphans that were adopted and in those who stayed in the orphanage. Since researchers do not manipulate the IV, all natural experiments are quasi-experiments.

<table>
<thead>
<tr>
<th>Type of experiment</th>
<th>Independent variable</th>
<th>Settings</th>
<th>Can we infer causation?</th>
</tr>
</thead>
<tbody>
<tr>
<td>True laboratory experiment</td>
<td>Manipulated by the researcher</td>
<td>Laboratory</td>
<td>Yes</td>
</tr>
<tr>
<td>True field experiment</td>
<td>Manipulated by the researcher</td>
<td>Real-life</td>
<td>Yes (but there may be confounding variables)</td>
</tr>
<tr>
<td>Natural experiment</td>
<td>Manipulated by the nature</td>
<td>Real-life</td>
<td>No</td>
</tr>
<tr>
<td>Quasi-experiment</td>
<td>Not manipulated; pre-existing difference</td>
<td>Laboratory or real-life</td>
<td>No</td>
</tr>
</tbody>
</table>

**Exercise**

Go online and find examples of quasi-experiments, natural experiments and field experiments in psychology.
What is a correlation?

Correlational studies are different from experiments in that no variable is manipulated by the researcher, so causation cannot be inferred. Two or more variables are measured and the relationship between them is mathematically quantified. The way it is done can be illustrated graphically through scatter plots. Suppose you are interested in investigating if there is a relationship between anxiety and aggressiveness in a group of students. For this you recruit a sample of students and measure anxiety with a self-report questionnaire and aggressiveness through observation during breaks. You get two scores for each participant: anxiety and aggressiveness. Suppose both scores can take values from 0 to 100. The whole sample can be graphically represented with a scatter plot.

What you will learn in this section

- What is a correlation?
  - Effect size
  - Statistical significance
- Limitations of correlational studies
  - Causation cannot be inferred
  - The third variable problem
- Curvilinear relationships
- Spurious correlations
- Sampling and generalizability in correlational studies
- Credibility and bias in correlational studies

Inquiry questions

- What does it mean for two variables to correlate with each other?
- What should be avoided when interpreting correlations?
- Can two correlating variables be unrelated in fact?
- Can correlations show curvilinear relationships?
Each dot on the scatter plot represents one person. The coordinates of each dot give you the scores obtained for each of the variables. For example, Jessica’s score on anxiety is 70 (the x-axis coordinate) and her score on aggressiveness is 50 (the y-axis coordinate). The whole scatter plot looks like a “cloud” of participants in the two-dimensional space of the two variables. A correlation is a measure of linear relationship between two variables. Graphically a correlation is a straight line that best approximates this “cloud” in the scatter plot.

In the example above, the correlation is positive because the cloud of participants is oblong and there is a tendency: as X increases, Y increases, so if an individual got a high score on variable X, they probably also got a high score on variable Y, and vice versa. This is where the name “correlation” comes from: the two variables “co-relate”. Remember that correlation does not imply causation: we cannot say that X influences Y, nor can we say that Y influences X. All we know is that there is a link between them.

A correlation coefficient can vary from −1 to +1. The scatter plots below demonstrate some examples:

A positive correlation demonstrates the tendency for one variable to increase as the other variable increases. A negative correlation demonstrates the inverse tendency: when one variable increases the other variable decreases. The steeper the line, the stronger the relationship. A perfect correlation of 1 (or −1) is a straight line with the slope of 45 degrees: as one variable increases by one unit, the other variable increases (or decreases) by exactly one unit. A correlation close to zero is a flat line. It shows that there is no relationship between the two variables: the fact that a person scored high or low on variable X tells us nothing about their score on variable Y. Graphically such scatter plots are more like a circle or a rectangle.

**Effect size and statistical significance**

The absolute value of the correlation coefficient (the number from −1 to 1) is called the effect size. How do you know if a correlation is small or large? There are widely accepted guidelines based on Cohen’s (1988) suggestions to interpret the effect size of correlations in social sciences.

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 0.10</td>
<td>Negligible</td>
</tr>
<tr>
<td>0.10–0.29</td>
<td>Small</td>
</tr>
<tr>
<td>0.30–0.49</td>
<td>Medium</td>
</tr>
<tr>
<td>0.50 and larger</td>
<td>Large</td>
</tr>
</tbody>
</table>

The effect size is not the only parameter that is important when interpreting a correlation coefficient. Another is the level of statistical significance. Statistical significance shows the likelihood that a correlation of this size has been obtained by chance. In other words, what is the probability that you will replicate the study with a different sample and the correlation will turn to zero? It depends on the sample size: with small samples you cannot be sure that an obtained correlation, even if it is relatively large, has not been obtained due to random chance. With large samples correlation estimates are more reliable and you can be more confident that the correlation is not a product of random chance but a genuine reflection of a relationship between the two variables in the population. The probability
that a correlation has been obtained due to random chance can be estimated. Again, there are conventional cut-off points when results are considered to be “statistically significant” or not.

<table>
<thead>
<tr>
<th>The probability that the result is due to random chance</th>
<th>Notation</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than 5%</td>
<td>p = n.s.</td>
<td>Result is non-significant</td>
</tr>
<tr>
<td>Less than 5%</td>
<td>p &lt; .05</td>
<td>Result is statistically significant (reliably different from zero)</td>
</tr>
<tr>
<td>Less than 1%</td>
<td>p &lt; .01</td>
<td>Result is very significant</td>
</tr>
<tr>
<td>Less than 0.1%</td>
<td>p &lt; .001</td>
<td>Result is highly significant</td>
</tr>
</tbody>
</table>

Thus, the conventional cut-off point for statistical significance is 5%. Whatever result you obtained, if the probability that this result is pure chance occurrence is less than 5%, we assume that the result is statistically significant, reliably different from zero and thus would be replicated in at least 95 out of 100 independent samples drawn from the same target population.

**TOK**

As you see, the nature of knowledge in psychology, just like the other social sciences, is probabilistic. We only know something with a degree of certainty and there is a possibility this knowledge is a product of chance.

How does that compare to the nature of knowledge in other areas such as natural sciences (physics, chemistry, biology), ethics or indigenous knowledge systems?

What can we do to increase the degree of certainty in social sciences (for example, think about replication of studies)?

When interpreting correlations one needs to take into account both the effect size and the level of statistical significance. If a correlation is statistically significant, it does not mean that it is large, because in large samples even small correlations can be significant (reliably different from zero). So, scientists are looking for statistically significant correlations with large effect sizes.

**ATL skills: Research**

Correlations are denoted by the letter r. These are some conventional notations for results of correlational studies. See if you can interpret them using your knowledge of Cohen’s effect size guidelines and levels of statistical significance:

- \( r = 0.14, p = \text{n.s.} \)
- \( r = 0.10, p < .05 \)
- \( r = 0.25, p < .05 \)
- \( r = 0.34, p < .01 \)
- \( r = 0.61, p < .001 \)

**Limitations of correlational studies**

Correlational studies have several major limitations.

1. As already mentioned, correlations cannot be interpreted in terms of causation.

2. “The third variable problem”. There is always a possibility that a third variable exists that correlates both with X and Y and explains the correlation between them. For example, cities with a larger number of spa salons also tend to have more criminals. Is there a correlation between the number of criminals and the number of spa salons? Yes, but once you take into account the third variable, the size of the city, this correlation becomes meaningless.

3. Curvilinear relationships. Sometimes variables are linked non-linearly. For example, a famous Yerkes-Dodson law in industrial psychology states that there is a relationship between arousal and performance: performance increases as arousal increases, but only up to a point. When levels of arousal surpass that
point, performance begins to decrease. Thus, optimal performance is observed when levels of arousal are average. This can be seen in the scatter plot below.

![Arousal and performance](image)

Figure 1.9 Arousal and performance

However, this relationship can only be captured by looking at the graph. Since correlation coefficients are linear, the best they could do is to find a straight line that fits best to the scatter plot. So, if we were using correlational methods to find a relationship between arousal and performance, we would probably end up obtaining a small to medium correlation coefficient. Psychological reality is complex and there are a lot of potentially curvilinear relationships between variables, but correlational methods reduce these relationships to linear, easily quantifiable patterns.

4. **Spurious correlations.** When a research study involves calculating multiple correlations between multiple variables, there is a possibility that some of the statistically significant correlations would be the result of random chance. Remember that a statistically significant correlation is the one that is different from zero with the probability of 95%. There is still a 5% chance that the correlation is an artifact and the relationship actually does not exist in reality. When we calculate 100 correlations and only pick the ones that turned out to be significant, this increases the chance that we have picked spurious correlations.

**Sampling and generalizability in correlational studies**

Sampling strategies in correlational research are the same as in experiments. First the target population is identified depending on the aims of the study and then a sample is drawn from the population using random, stratified, opportunity or self-selected sampling.

Generalizability of findings in correlational research is directly linked with sampling and depends on representativeness of the sample. Again, this is much like population validity in experiments.

**Credibility and bias in correlational research**

Bias in correlational research can occur on the level of variable measurement and on the level of interpretation of findings.

On the level of measurement of variables, various biases may occur and they are not specific to correlational research. For example, if observation is used to measure one of the variables, the researcher needs to be aware of all the biases inherent in observation. If questionnaires are used to measure variables, biases inherent in questionnaires become an issue. The list goes on.

On the level of interpretation of findings, the following considerations represent potential sources of bias:

1. Curvilinear relationships between variables (see above). If this is suspected, researchers should generate and study scatter plots.
2. “The third variable problem”. Correlational research is more credible if the researcher considers potential “third variables” in advance and includes them in the research in order to explicitly study the links between X and Y and this third variable.
3. Spurious correlations. To increase credibility, results of multiple comparisons should be interpreted with caution. Effect sizes need to be considered together with the level of statistical significance.

**ATL skills: Self-management**

Go back to the overview table on page 8. Compare and contrast sampling, generalizability, credibility and bias in correlational research with those in experimental research.

- In what aspects are the approaches different?
- In what aspects are they the same?
- Are there any aspects where the ideas are similar but the terminology differs?
Credibility in qualitative research

Credibility in qualitative research is an equivalent of internal validity in the experimental method. As you have seen, internal validity is a measure of the extent to which the experiment tests what it is intended to test. To ensure internal validity in experimental research we need to make sure that it is the IV, not anything else, that causes the change in the DV. To do this, we identify all the possible confounding variables and control them, either by eliminating them or by keeping them constant in all groups of participants.

In a similar fashion, credibility in qualitative research is related to the question, “To what extent do the findings reflect the reality?” If a true picture of the phenomenon under study is being presented, the study is credible.

The term “trustworthiness” is also used to denote credibility in qualitative research.
How do we know if the picture of a phenomenon presented in the findings from a qualitative study is “true”? If we had a way to know that, we wouldn’t need a research study in the first place!

One of the popular definitions of knowledge is “justified true belief”. A similar problem, however, arises with this definition: other than through “knowledge”, we do not have a way of establishing if something is true. So, knowledge depends on truth but truth is a result of knowledge.

To solve this paradox, it has been suggested to substitute “true” in this definition to “beyond reasonable doubt”. So, to ensure that a qualitative research study is credible we need to demonstrate that its findings are “true beyond reasonable doubt”.

How do you understand that? What do you think is “reasonable doubt” in this context?

To ensure that what is presented in the findings of a qualitative study is true, several types of measures can be taken.

### Credibility (trustworthiness) in qualitative research

#### Figure 1.10 Trustworthiness

1. **Triangulation.** This refers to a combination of different approaches to collecting and interpreting data. There are several types of triangulation all of which can be used to enhance the credibility of a study.

   a. Method triangulation. The use of different methods in combination can compensate for their individual limitations and reinforce their strengths. If the same results are obtained using various methods (for example, interviews and observations), credibility increases.

   b. Data triangulation. This refers to using data from a variety of accessible sources. For example, if participants during an interview refer to certain documents, these documents may be studied in order to gain a clearer understanding of the participants’ experiences. Observations may be supported by studying documented biographical data, and so on.

   c. Researcher triangulation. As follows from the name, this refers to combining observations/interpretations of different researchers. Undoubtedly, if two people see the same thing, this increases credibility of their findings.

   d. Theory triangulation. This refers to using multiple perspectives or theories to interpret the data.

2. **Establishing a rapport.** Researchers should ensure that the participants are being honest. For example, the researcher should remind participants about voluntary participation and the right to withdraw so that responses are only obtained from participants who are willing to contribute. It should be made clear to the participants that there are no right or wrong answers and in general a good rapport should be established with the participants so that they alter their behaviour in the presence of the researcher as little as possible.

3. **Iterative questioning.** In many research projects, especially those involving sensitive data, there is a risk that participants will distort data either intentionally (lying) or unintentionally to try to create a certain impression on the researcher. Spotting ambiguous answers and returning to the topic later while at the same time rephrasing
the question might help researchers to gain a deeper insight into the sensitive phenomenon.

4. Reflexivity. Researchers should reflect on the possibility that their own biases might have interfered with the observations or interpretations. Arguably, due to the nature of qualitative research that requires the involvement of the researcher in the studied reality, a certain degree of bias is unavoidable. However, researchers need to be able to identify the findings that might have been affected by these biases the most, and if they were affected, how. There are two types of reflexivity:

a. **Epistemological reflexivity**, linked to knowledge of the strengths and limitations of the method used to collect data (“the following behaviours were observed … however, they should be interpreted with caution because participants were aware that they were being observed and hence might have modified their behaviour”)

b. **Personal reflexivity**, linked to the personal beliefs and expectations of the researcher (“I noticed that overcoming trauma was particularly emphasized in their conversations, however, since I myself have a history of overcoming childhood trauma, this observation could have been influenced by my personal beliefs and should be cross-checked by an independent interviewer”).

5. Credibility checks. This refers to checking accuracy of data by asking participants themselves to read transcripts of interviews or field notes of observations and confirm that the notes/transcripts are an accurate representation of what they said (meant) or did. This is often used in interviews with the interviewees receiving the transcript and being asked to correct any inaccuracies or provide clarifications.

6. “**Thick descriptions**”. This refers to explaining not just the observed behaviour itself, but also the context in which it occurred so that the description becomes meaningful to an outsider who never observed the phenomenon firsthand. Essentially it boils down to describing the phenomenon in sufficient detail so that it can be understood holistically and in context. For example, imagine a stranger smiled at you.

This behaviour out of context can be reported “ thinly”, just stating the fact, or it can be placed in a context (who, where, in what circumstances) thus making it meaningful. To provide thick descriptions researchers should reflect anything that they observe and hear including their own interpretations, even if some of these details do not seem significant at the time. Thick descriptions are also referred to as “rich” descriptions; these terms are interchangeable.

---

**ATL skills: Research**

To what extent is this similar to the way internal validity is ensured in experimental research? What are the differences?

---

**Bias in qualitative research**

In quantitative research we deal with potential bias by trying to eliminate it completely or keeping the potentially confounding variables constant in all comparison groups. In qualitative research this approach is not possible, and bias is actually an integral part of the research process because the researcher is a tool through which data is collected. So, while some types of bias may be avoided, other types of bias are inevitable and need to be reflected on and accounted for.

Sources of bias in qualitative research may be associated both with the researcher and the participant. Let’s look at the major sources of bias.

**Participant bias**

1. **Acquiescence bias** is a tendency to give positive answers whatever the question. Some people are acquiescent by nature, and in some others acquiescence may be induced by the nature of the questions or the researcher’s behaviour. To avoid this bias, researchers should be careful not to ask leading questions, making their questions open-ended, neutral and focused on the opinions of the participant.

2. **Social desirability bias** is a participant’s tendency to respond or behave in a way that they think will make them liked or accepted. Participants may guess (or at least have a vague idea about) the aim of the study and try to look better than they really are. This may be done intentionally or unintentionally. Research into sensitive topics is especially vulnerable to social
desirability. To reduce this bias, questions should be phrased in a non-judgmental way that suggests that any answer is acceptable. Another trick that researchers use is to ask questions about a third person (for example, what do your friends think about ...?). This helps participants to disengage from the sensitive topics and provide more honest answers.

3. **Dominant respondent bias** occurs in a group interview setting when one of the participants influences the behaviour and responses of the others. Dominant respondents may “hijack” talking time or intimidate others by demonstrating their assertiveness or superior knowledge of the subject. Researchers should be trained to keep dominant respondents in check and make sure that all participants are provided with equal opportunities to speak and are in a safe and comfortable environment to voice their opinions.

4. **Sensitivity bias** is a tendency of the participant to answer regular questions honestly, but distort their responses to questions on sensitive subjects. They may even give incorrect information to hide secrets. The solution to this problem is to build a good rapport with the participant and create trust between the participant and the researcher. To build trust, the researcher needs to behave professionally, make ethical guidelines regarding issues such as confidentiality absolutely clear to the participant and increase the sensitivity of the questions gradually while being responsive to the participant’s concerns.

**Researcher bias**

1. **Confirmation bias** occurs when the researcher has a prior belief and uses the research in an unintentional attempt to confirm that belief. Confirmation bias may influence the way questions are worded, the small nuances in the researcher’s non-verbal behaviour, and selectivity of attention while observing behaviour or interpreting the data. Information that supports the prior belief is attended to, while information that contradicts it is disregarded. Reflexivity is the solution to confirmation bias. Confirmation bias is such a deeply grounded error in human information processing that it is largely unavoidable in qualitative research where data can only be collected “through” a human observer. So rather than avoiding it, researchers should be trained to recognize it and take it into account. If the possibility of bias is recognized, research can then be repeated with another observer to corroborate the findings (or not).

2. **Leading questions bias** occurs when respondents in an interview are inclined to answer in a certain way because the wording of the question encourages them to do so. Even if an interview is carefully planned in advance, researchers often ask additional follow-up or clarification questions, and these may potentially cause distortions in the responses. Interviewers should be rigorously trained in asking open-ended, neutral questions that do not suggest a particular answer. Also they should avoid paraphrasing the participant’s response to make sure they understood it correctly. Questions should be worded in the participant’s own language.

3. **Question order bias** occurs when responses to one question influence the participant’s responses to the following questions. This bias stems from the human tendency to be consistent in our beliefs and actions. For example, if the first question on the interview asked if you liked sports and you hesitated but said yes, you would probably be inclined later to give more positive answers about your attitudes to gym membership. To minimize this bias, general questions should be asked before more specific ones, positive questions before negative ones, and behaviour questions before attitude questions.

4. **Sampling bias** occurs when the sample is not adequate for the aims of the research. For example, the selection of people who are not “the best fit” in terms of the research purposes may be the result of convenience sampling. Also there are “professional participants” who look for opportunities to take part in research that provides financial incentives for participation. Although they can be accessed quickly and recruited easily, samples consisting entirely of “professional participants” should be used with caution.

5. **Biased reporting** occurs when some findings of the study are not equally represented in the research report. For example, the researcher
might choose to only briefly mention pieces of evidence that do not “fit”. Reflexivity, integrity and training of researchers are the means to counteract biased reporting.

### Bias in qualitative research

<table>
<thead>
<tr>
<th>Participant bias</th>
<th>Researcher bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquiescence</td>
<td>Confirmation bias</td>
</tr>
<tr>
<td>Social desirability</td>
<td>Leading questions bias</td>
</tr>
<tr>
<td>Dominant respondent</td>
<td>Question order bias</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>Sampling bias</td>
</tr>
<tr>
<td></td>
<td>Biased reporting</td>
</tr>
</tbody>
</table>

▲ Figure 1.11 Types of bias in qualitative research

To sum up, some types of bias in qualitative research may be eliminated while some others need to be recognized and taken into account. Reflexivity and triangulation are the two most important instruments that allow the researcher to reduce the influence of bias in qualitative research.

With regards to researcher bias, special attention needs to be paid to incorporating all data in the report and acknowledging the limitations of the research study, as well as asking independent researchers to review the results and procedure followed. With regards to participant bias, it is important to ask carefully crafted, indirect and open-ended questions and maintain neutrality.

The presence of biases is directly linked to both credibility and generalizability of research findings.

### ATL skills: Thinking and self-management

The sources of bias in experimental and qualitative research appear in the table below. See if you can find any overlaps and discuss in class.

<table>
<thead>
<tr>
<th>Experimental research</th>
<th>Qualitative research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection</td>
<td>Acquiescence bias</td>
</tr>
<tr>
<td>History</td>
<td>Social desirability bias</td>
</tr>
<tr>
<td>Maturation</td>
<td>Dominant respondent bias</td>
</tr>
<tr>
<td>Testing effect</td>
<td>Sensitivity bias</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>Researcher bias</td>
</tr>
<tr>
<td>Regression to the mean</td>
<td>Confirmation bias</td>
</tr>
<tr>
<td>Experimental mortality</td>
<td>Leading questions bias</td>
</tr>
<tr>
<td>Experimenter bias</td>
<td>Question order bias</td>
</tr>
<tr>
<td>Demand characteristics</td>
<td>Sampling bias</td>
</tr>
<tr>
<td></td>
<td>Biased reporting</td>
</tr>
</tbody>
</table>

### Sampling and generalizability in qualitative research

Generalization is a broad inference from particular observations. It is “an inference about the unobserved based on the observed” (Polit and Beck, 2010).

Traditionally generalizability has been the focus of debate between supporters of quantitative and qualitative methods. The main argument against generalizability in qualitative research is that samples are not statistically representative of the target population. As you know, representativeness in quantitative research is a necessary requirement for findings to be applied beyond the sample to the target population it represents. A “weak” counter-argument to that is to say that qualitative methods do not aim to apply research findings to a wider population. In other words, the purpose of qualitative methods is the study of a particular sample but not the population it “represents”. However, some scholars make a stronger argument and claim that generalizability is in fact achievable, to a certain extent, in qualitative research.

There are other arguments too, less popular, but no less valid. Some scientists doubt that generalizability is possible in principle, even in quantitative studies. They argue that every research study is embedded in a certain context (sample, setting, time, and so on), and generalization of findings would always include a degree of unsubstantiated speculation. Some other scholars argue that qualitative research is in fact more generalizable. They claim that rich data obtained in qualitative studies allows us to gain a deeper understanding of the phenomenon and thus make more accurate inferences about its nature.

### Sampling

In quantitative research, representativeness of the sample (and therefore the ability to generalize results to a wider population) is ensured through random sampling. In random sampling each member of the target population has an equal chance of being included in the sample. In other words, random sampling is probabilistic. However, sampling in qualitative research is non-probabilistic. These are the
most commonly used types of sampling in qualitative research.

1. **Quota sampling.** In quota sampling it is decided prior to the start of research how many people to include in the sample and which characteristics they should have. This decision is driven by the research question—researchers look for people whose experiences would most likely provide an insight into the topic. Using various recruitment strategies, researchers then recruit participants until the quotas are met. Quota sampling is similar to stratified sampling in quantitative research in that both the important participant characteristics and the necessary sample proportions are pre-defined.

2. **Purposive sampling.** This is similar to quota sampling in the sense that the main characteristics of participants are defined in advance and then researchers recruit participants who have these characteristics. However, the proportions and the sample size are not defined.

3. **Theoretical sampling.** This is a special type of purposive sampling that stops when the point of data saturation is reached. **Data saturation** means that no new information is obtained from new participants added to the sample. Whether information is “new” or not is defined on the basis of a background theory: if no new evidence (or counterevidence) for the claims of the theory emerges, data saturation is reached. Generalization in this case is made from the data to the theory.

4. **Snowball sampling.** In this approach a small number of participants is invited and asked to invite other people they know who also are of interest for the purposes of the research. This approach is mostly used in pilot research studies (when there are insufficient resources to carefully select participants) or in research with groups of people who are very difficult to reach (for example, drug users, youth gang members).

5. **Convenience sampling.** The most superficial approach where you just use the sample that is easily available/accessible (for example, professors might conduct research with university students simply because it is time- and cost-efficient).

Generalizability of research findings in qualitative research may depend on the type of sampling used—studies using quota, purposive or theoretical sampling are more generalizable.

**Types of generalizability**

Firestone (1993) distinguished between three types of generalizability which provide a convenient framework for comparing quantitative and qualitative studies.

1. **Sample-to-population generalization.** The researcher starts by identifying the target population and then selects a sample that is representative of this population. The best approach to achieve this is to use random sampling. The concept that is used to describe sample-to-population generalizability in experiments is “population validity” (part of “external validity”). Due to the non-probabilistic nature of samples in qualitative research, this type of generalization is difficult.

2. **Theoretical generalization.** Generalization is made from particular observations to a broader theory. In quantitative research theoretical generalization takes the form of construct validity: it is the leap we make from directly observable operationalizations to the unobservable construct. In qualitative research theoretical generalization is achieved through rigorous analysis and interpretation of research findings: we can generalize to a wider theory if data saturation was achieved, thick descriptions provided, analysis was in-depth and free of biases, and so on. Theory plays a much greater role in qualitative research than in quantitative.

3. **Case-to-case generalization, also known as transferability.** Generalization is made to a different group of people or a different setting/context. In qualitative research transferability is the responsibility of both the researcher and the reader of the research report. The researcher’s responsibility is to ensure that thick descriptions are provided so that the reader has sufficient information and details about the context of the study. The reader’s responsibility is to decide whether or not the context described in
the report is similar to a new situation (Polit and Beck, 2010). A rough and pretty distant equivalent of transferability in quantitative research would probably be “ecological validity” (another part of “external validity”).

**ATL skills: Research and self-management**

Compare the sampling techniques used in experiments and in qualitative research studies. Use any kind of visual representation to demonstrate the results of this comparison and present it in class.

How are the three types of generalizability approached in experiments and qualitative research studies? Which of these do you think are better achieved in qualitative research as compared to experimental research?

Go back to the overview table on page 8 and see if it reflects your current knowledge of generalizability.
Observation

There are several common reasons for choosing the method of observation.

1. The focus of the research is on how people interact, interpret each other’s behaviour and act upon these interpretations in a natural setting. For example, if you observe a group of primary school children in a typical enrichment class you may understand a lot about their everyday school life. Most other research methods are artificial in the sense that they place the participant in a specially constructed research context.

2. The researcher believes that meaningful knowledge in the research area cannot be generated without observation, for example, because it cannot be articulated. For example, if you want to gain an insight into the behaviour of your classmates during a fire drill at your school, it will probably be more meaningful to observe an actual fire drill than to conduct an interview and analyse verbal responses.

3. Observation allows the researcher to become immersed deeply into the studied phenomenon, sometimes even becoming part of it. Arguably, this is a strength because you gain almost first-hand experiences.

Observation is “experiential” and the researcher is strongly involved in the process of data generation. All generated data is the product of his or her selective attention and interpretations. This makes reflexivity especially important.

So, the main advantage of observation is the ability to generate diverse data about the behaviour of participants in a naturally occurring setting. The
major limitation would be susceptibility to biases, so reflexivity and other methods of ensuring credibility and generalizability of qualitative research need to be used extensively.

There are several types of observation, and the particular type chosen will have broad implications in terms of credibility, reflexivity, generalizability and ethics.

1. **Laboratory versus naturalistic observation.** Naturalistic observation is carried out in naturally occurring settings, that is, a place that has not been arranged for the purposes of the study. Sometimes naturalistic observation would be the only choice, for example, in situations where it is unethical to arrange settings for the behaviour of interests to occur. If you wanted to study inter-group discrimination and violence, it would be unethical to encourage violence in a research setting. However, you may observe naturally occurring violence. A drawback is that it may be time-consuming because the behaviour of interest only occurs at certain times.

2. Observation may be overt or covert. **Overt observation** occurs when participants are aware of the fact that they are being observed. Clearly the ethics of this approach are a strength as participants give their informed consent, but there are methodological limitations—biases related to participant expectations. When people know that they are being observed, they can intentionally or unintentionally change their behaviour. In contrast, in **covert observation** the researcher does not inform the members of the group about the reasons for his or her presence. An advantage of covert observation is gaining access to groups that would not normally agree to participate in research (for example, socially isolated or violent groups). Another strength is the avoidance of participant bias—subjects do not know they are being observed, so they behave naturally. The ethics here are a disadvantage. Participants do not give their consent to take part in the study. One way to avoid this issue is to debrief participants after the observation session and ask for their consent prior to using the data for research purposes.

3. **Participant observation.** In this method the observer becomes part of the observed group. For example, many anthropologists spend time living among members of an indigenous society in order to study their culture “from the inside”. For a great example of this, watch the BBC documentaries *Tribe* and *Amazon* with Bruce Parry. The advantage of participant observations is that they allow the researcher to gain first-hand experiences with the phenomenon of interest, gaining valuable insights. However, the drawbacks include the risk that the observer will lose objectivity as he or she becomes too involved with the studied group of individuals. This may happen because the researcher begins to identify himself or herself with the group. And, of course, there is the ethical issue: if participants do not realize that one of the members of their group is in fact an observer collecting information, this may be ethically questionable, especially in sensitive research topics.

4. **Structured versus unstructured observation.** In structured observation information is recorded systematically and in a standardized way. For example, structured observation may be conducted with a checklist of behaviours of interest where the observer is required to note the occurrence of these specific behaviours in pre-defined time intervals. Rosen, Carrier and Cheever (2013) conducted structured observations of the use of technology among school students. Observers were equipped with a checklist of behaviours related to the use of technology (using a browser, using a telephone, and so on) and they had to fill out this checklist minute-by-minute. Unstructured observations do not have a pre-defined structure and the observer simply registers whatever behaviours they find noteworthy.

**Exercise**

- Suppose your aim is to study ways in which destructive cults brainwash their new members, and observation is your method. What type of observation would you use and why?
- Describe how you would set up your research procedure both in terms of preparation and the actual observation process.
**Interview**

In-depth interviews are one of the most popular qualitative research methods for several reasons.

1. This may be the only way to get an insight into the nature of subjective experiences and interpretations. Since attitudes, values, patterns of interpretation and other subjective phenomena are unobservable, the most straightforward way to study them is to rely on the participants’ verbal reports.

2. Interviews may be used to understand the meanings participants attach to certain events and their points of view. Again, this is not directly achievable by most other methods.

3. In-depth individual interviews are useful when the topic is too sensitive for people to discuss in a group setting.

Interviews are a very personal form of research because there is direct contact between the interviewer and the interviewee. At the same time, interviews can, and often do, touch upon sensitive topics like coping with a terminal illness, experiencing phobias, daily routines related to internet addiction and drug use.

Interviewing techniques are driven by the goal of learning as much as possible about the interviewee’s opinions and experiences. The interviewer tries to build a rapport with the participant and then engage them by asking neutral and carefully phrased questions, listening carefully to their responses and asking follow-up questions. The interviewer is the main research instrument. Tiny nuances in verbal and non-verbal behaviour of the interviewer may affect the interviewee’s responses. For example, it is common in everyday conversations to ask leading questions, but interviewers must avoid doing it. This is why interviewers receive intensive training.

Interview data comes in the form of an audio or video recording which is subsequently converted to an interview transcript. Sometimes data also includes interview notes, accompanying observations about the participant and the interview context. Transcripts are later coded and analysed in line with the aims of the research.

There are three types of interview, depending on how fixed the list and the sequence of the questions is.

1. **Structured interviews** include a fixed list of questions that need to be asked in a fixed order. It is most useful when the research project involves multiple interviewers and it is essential that they all conduct the sessions in a similar way. This allows many participants to be interviewed and some comparisons to be made (for example, comparing responses from male and female participants, across age groups, across cultures).

2. **Semi-structured interviews** do not specify an order or a particular set of questions. They are somewhat like a checklist: the researcher knows that certain questions must be asked, but beyond that he or she can ask follow-up questions to get clarifications. If it better fits the natural flow of the conversation, the researcher can change the question order. Semi-structured interviews are better suited for smaller research projects, but they are also more effective in studying the unique experiences of each participant.

3. **Unstructured interviews** are mostly participant-driven, and every next question is determined by the interviewee’s answer to the previous one. Of course, the researcher still has to keep in mind the overall purpose of the research and stay focused on exploring a particular topic. However, two different interviewees may end up getting very different questions.

**Exercise**

Suppose you are interested in studying the reasons why teenagers join criminal groups. You used snowball sampling techniques to recruit 10 participants. Would you use a structured, semi-structured or unstructured interview? Why?

What do you think are the factors that need to be considered in conducting an interview with teenage gang members?

**Focus group**

The focus group is a special type of semi-structured interview that is conducted simultaneously with a group of 6–10 people. The key factor is that participants are encouraged to interact with each other and the interviewer serves as a facilitator. Participants discuss responses to every question
and react to each other’s statements. This provides additional data because they use their own language, agree and disagree with each other, enrich each other’s perspectives and demonstrate a variety of opinions. The focus group facilitator can observe group dynamics and make use of it by directing their interaction so that they stay focused on the research topic.

The advantages of a focus group include:

- it is a quick way to get information from several participants at the same time
- it creates a more natural and comfortable environment than a face-to-face interview, ensuring less participant bias
- it is easier to respond to sensitive questions when you are in a group
- multiple perspectives are discussed hence a more holistic understanding of the topic is achieved.

However, there are several “new” limitations that come as a cost for including group dynamics into the research process.

- If one of the participants is especially dominant, this may distort the responses of the other participants (for example, if they feel a need to conform), and it is the facilitator’s responsibility to ensure that each participant contributes freely to the conversation.
- It is more difficult to preserve anonymity and confidentiality.
- Focus groups are especially demanding in terms of sampling and creating interview transcripts.

**Content analysis**

Interview recordings need to be transcribed and then analysed. But how do you analyse a text in a systematic and rigorous way while minimizing researcher bias? The widely used approach to analysing texts produced by participants is known as **inductive content analysis**, or thematic analysis. The goal of inductive content analysis is to derive a set of recurring themes. When extracting the themes the researcher has to maintain a balance between description and interpretation in the sense that the text needs to be interpreted, but these interpretations must be backed up by evidence from the text.

**TOK**

What is the difference between induction and deduction? If you do not remember, look it up.

Inductive content analysis follows a series of steps (Elo and Kungäs, 2008).

1. Writing the transcript. There are two types of transcript: verbatim or post-modern. Verbatim transcripts are word-for-word accounts of everything the participant said. Post-modern transcripts include notes about the intonation, gestures and other non-verbal elements in the participant’s behaviour.

2. Reading the raw material several times and identifying initial themes. This is done iteratively. Researchers start with low-level themes, trying to stay as close to the text as possible. When the first reading is done, a set of initial themes is identified and may be written on the margins. The second reading is done and the themes are confirmed (and revised); also new themes may be added. This is done several times. Sometimes independent coders are used to check the credibility of deriving low-level themes from the text.

3. Low-level themes are grouped into a smaller number of high-level themes. This grouping involves an element of interpretation on the part of the researcher: they need to decide if X, Y and Z belong to category A. As a credibility check, other researchers may be involved in the process so that results of grouping can be compared across researchers. The result of this stage of analysis is a manageable set of high-level meaningful units that summarize the transcript.

4. A summary table of themes is prepared. The table lists all the high-level emergent themes, all the lower-level themes within them, and supporting quotations from the raw transcript. The structure of themes can also be revised slightly at this point to account for parts of the transcript that are still unexplained. Data saturation is reached when subsequent readings of the transcript do not lead to identifying any new themes.
5. Finally, conclusions are formulated based on the summary table. These conclusions link the emergent themes to the theory. As a credibility check, participants may be shown the results of the analysis and asked to confirm the emergent themes as well as the derived interpretations.

The resulting analysis may be accompanied by “memos” that explain to the reader how and why certain analysis decisions were made, increasing the “thickness” of descriptions (which, as you know, increases credibility).

Inductive content analysis can also be applied to observational data. In this case the raw material for analysis comes in the form of field notes describing a participant’s behaviour rather than interview transcripts.

If a theory emerges from the data, it is referred to as a “grounded theory”. The name suggests that grounded theory “grows out of” empirical data as opposed to prior beliefs.

Exercise

- Find an example in this book of a study that used the interview or the focus group as the primary research method. What type of interview or focus group was it? How was content analysis organized?
- What can you say about generalizability and credibility of the findings?

Case study

A case study is an in-depth investigation of an individual or a group. You might say that this is not a proper definition because other research methods can also be defined this way, and you would be right. In fact, case studies can involve a variety of other methods (observations, interviews, and so on), anything that deepens our understanding of an individual or a group of interest. There are several reasons why case studies are referred to as a separate research method, even though they are actually a combination of other methods.

1. The individual or group that is the object of a case study is unique in some way. As a result, the purpose is to gain a deep understanding of this particular individual or group.

2. Sampling is not an issue: you are interested in this particular case, not the population this case “represents”.

3. There is less focus on generalizability. Findings do get generalized, but this is a by-product of the in-depth description and explanation of the case (case-to-case and theoretical generalization).

4. The case is studied thoroughly, using a combination of different methods, and often longitudinally. This is why we defined a case study as an “in-depth investigation”.

What are the reasons for choosing a case study as the preferred method?

First, case studies are useful to investigate phenomena that could not be studied otherwise. For example, it is a group that is hard to get access to and you may only get a chance to study one individual (think about studying the personality of a serial killer).

Second, case studies can contradict established theories and help develop new theories. Why is this a good thing? According to the principle of falsification in science (Karl Popper), the proper way to test a theory is to find one case that contradicts it. If you can’t, the theory stands, but if you succeed, the theory needs to be rejected or modified, and this is how science develops. To test the theory that “all swans are white” you need to try and find one black swan. In a similar fashion, universal theories of memory in cognitive psychology can be tested by studying individuals with unusual or unique memory abilities. If in these individuals memory proves to function differently, then the universal theory of memory is not as universal as we thought. So, “boundary” cases are interesting, and since they are quite rare, we want to study them thoroughly.

Case studies have several limitations. Researcher bias can be a problem as, due to the longitudinal nature of the study, researchers might get too involved. Participant bias is also a potential problem for the same reason: the participant interacts with the researcher for a long period and it is easier for the participant to become susceptible to acquiescence, social desirability,
and so on. The generalization of findings is especially problematic from a single case to other settings or to a wider population. Generalization depends on thickness of descriptions and triangulation (other researchers, other case studies, and so on).

Apart from the ethical considerations involved in qualitative research in general, case studies are especially demanding in terms of anonymity and confidentiality—it is difficult to preserve anonymity of unique cases. In case studies of patients with brain damage it may be difficult to obtain informed consent because they might not fully realize the terms of the document. It is debatable how “informed” this informed consent is exactly.
Ethics is an integral part of psychological research because it is research with living beings (humans and animals). This is one of the things that distinguishes the human sciences from the natural sciences—ethically, the study of human beings is not the same as the study of material objects.

All around the world the activities of psychologists are regulated by codes of ethics. These codes outline the ethical principles and procedures to be followed in all aspects of a psychologist’s professional activities: counseling, testing and research. If a psychologist breaches the code, their professional license may be discontinued. Codes of ethics have been developed by international as well as national psychological associations, and there is a lot of overlap in their content as the ethical considerations in psychology are pretty much universal. As an example of a well-established code of ethics you may want to look through the one issued by APA (American Psychological Association).

---

**Inquiry questions**

- Since psychology is a study of living beings, what ethical issues does it raise?
- How can we decide what is ethical and what is not in psychology?

---

**What you will learn in this section**

- Ethical considerations in conducting the study
  - Informed consent
  - Protection from harm
  - Anonymity and confidentiality
  - Withdrawal from participation
  - Deception
  - Debriefing
  - Cost-benefit analysis in ambiguous cases
- Ethics committees
- The Little Albert experiment

- Ethical considerations in reporting the results
  - Data fabrication
  - Plagiarism
  - Publication credit
  - Sharing research data for verification
  - Handling of sensitive personal information
  - Social implications of reporting scientific results
- The controversy around Cyril Burt

---

**Psychology in real life**


Since IB psychology is an academic subject (involving no counseling), we will only focus on ethical considerations related to research. We will also break them into two large groups:

1. ethical considerations in conducting the study
2. ethical considerations in reporting the results.
Ethical considerations in conducting the study

The following list outlines the main ethical considerations to be addressed when conducting a research study in psychology.

1. **Informed consent.** Participation in a study must be voluntary, and participants must fully understand the nature of their involvement, including the aims of the study, what tasks they will be exposed to and how the data will be used. Researchers should provide as much information as possible and in the clearest possible way, hence the name “informed” consent. If the participant is a minor, consent should be obtained from parents or legal guardians.

2. **Protection from harm.** At all times during the study participants must be protected from physical and mental harm. This includes possible negative long-term consequences of participating in a research study.

3. **Anonymity and confidentiality.** These two terms are often used interchangeably, but they refer to slightly different things. Participation in a research study is confidential if there is someone (for example, the researcher) who can connect the results of the study to the identity of a particular participant, but terms of the agreement prevent this person from sharing the data with anyone. So, the participant provides personal data, but the data stays confidential under the research agreement. Participation in a study is anonymous if no one can trace the results back to a participant’s identity because they have not provided any personal details. An example of anonymity would be filling out an online survey without providing your name.

4. **Withdrawal from participation.** It must be made explicitly clear to participants that, since their participation is voluntary, they are free to withdraw from the study at any time they want. Researchers must not prevent participants from withdrawing or try to convince them to stay.

5. **Deception.** In many cases the true aims of the study cannot be revealed to the participants because it would change their behaviour (for example, due to social desirability). So a degree of deception needs to be used. In some research methods deception is part of the process, for example, covert observation. Researchers must be careful and if deception is used, it must be kept to the necessary minimum.

6. **Debriefing.** After the study participants must be fully informed about its nature, its true aims, how the data will be used and stored. They must be given an opportunity to review their results and withdraw the data if they want to. If deception was used, it must be revealed. Care must be taken to protect participants from any possible harm including long-term effects such as recurring uncomfortable thoughts. In some cases psychological help must be offered to monitor the psychological state of the participant for some time after the study (for example, in sleep deprivation studies).

**ATL skills: Self-regulation**

To memorize short lists, it is useful to use acrostics—phrases in which the first letter of each word stands for one of the elements on the list. For example, the ethical considerations in conducting a study may be combined in the following acrostic:

- Can (consent)
- Do (debriefing)
- Cannot (confidentiality)
- Do (deception)
- With (withdrawal)
- Participants (protection from harm)

Try making such acrostics of your own with other lists in this chapter: threats to internal validity, types of bias in qualitative research, and so on.

Display the results in your classroom to share with others and gradually you will pick out the ones that are most easily memorized.

Very often ethical decisions prior to conducting a study are not easy, and a **cost-benefit analysis** needs to be conducted. For example, sometimes participants should not know the true aim of the study for their behaviour to be more natural. Sometimes it is difficult to preserve confidentiality (for example, in unique cases). Sometimes there is a risk that participants could get mentally or physically harmed. For example, in the famous Stanford Prison Experiment (Zimbardo *et al.*, 1973) participants were led to believe that they were
imprisoned and were kept in harsh conditions, being humiliated and dehumanized by other participants (who were randomly assigned the role of guards). Studies of such phenomena as obedience, conformity, compliance, violence and prejudice can rarely be designed so that they are harmless to the participants. So can we make the decision to relax some of the ethical standards for a particular study?

Such decisions can be made in some circumstances, including:

1. if potentially the study can reveal scientific information that will benefit a lot of people
2. if there is no way the study of a phenomenon can be conducted without relaxing an ethical standard.

In all countries professional bodies of psychologists have ethics committees that resolve ambiguous issues and approve research proposals. Research proposals with a full description of the aims, procedures and anticipated results are submitted to the committee and reviewed. In some cases, when research is potentially useful, ethically ambiguous research studies may get the “green light”. Then the researchers will need to be extra careful in making sure that participant harm is minimized and long-term follow-up after the study is provided. Failure to cooperate with an ethics committee is itself a violation of ethics.

Psychology in real life

If you want to know more about the Stanford Prison Experiment, explore this website: http://www.prisonexp.org/.


If you want to look at the current composition of the APA Ethics Committee, explore their page on the APA website: http://www.apa.org/about/governance/bdcmte/ethics-committee.aspx.

You might also want to look at their annual reports.

Research in focus: The Little Albert experiment

The Little Albert experiment was carried out by John B Watson (Watson and Rayner, 1920). The study provided evidence of classical conditioning in humans. Similar to Ivan Pavlov’s experiments with his dogs (salivating at the sound of a bell), Watson was trying to form a certain reaction in response to a certain stimulus in a human baby. Watson observed that a baby’s fearful reaction to loud noises was an innate, automatic response. When they hear a loud noise, little children always display behavioural signs of fear (tears and so on). So he set out to form a fearful reaction to a neutral stimulus, furry objects, using the classic Pavlovian techniques.
Ethical considerations in reporting the results

The following list gives the main ethical considerations to be addressed when reporting results.

1. **Data fabrication.** This is a serious violation of ethical standards and psychologists may lose their license if they fabricate data. If an error is found in already published results, reasonable measures should be taken to correct it (for example, retraction of an article or publication of an erratum).

2. **Plagiarism.** It is unethical to present parts of another’s work or data as one’s own.

3. **Publication credit.** Authorship on a publication should accurately reflect the relative contributions of all the authors. For example, the APA Code of Ethics states specifically that if a publication is based primarily on a student’s work, the student must be listed as the first author, even though his or her professors co-authored the publication.

4. **Sharing research data for verification.** Researchers should not withhold the data used to derive conclusions presented in the publication. The journey from raw data (in the form of a matrix with numbers for quantitative research or a text/transcript for qualitative research) to inferences and conclusions is full of intermediate decisions, interpretations and inevitable omissions. It is healthy scientific curiosity to want to replicate the analysis, and any request from an independent researcher to share raw data should be satisfied, provided both parties use the data ethically and responsibly. This entails, for example, making the shared data set anonymous (deleting the names or other identifiers) and only using the shared data set for the stated purposes.

5. **Handling of sensitive personal information.** This refers to how the results of the study are conveyed to individual participants.

   a. **Handling of information obtained in genetic research.** Research into genetic influences on human behaviour, such as twin, adoption or family studies, can sometimes lead to revealing private information to one individual about other members of their family. Examples include misattributed parentage or health status. In twin studies one may discover that he
or she has a twin that they never met. Information of this sort may be disclosed accidentally during interviews, inferred by the participants in the debriefing session or in the report of results. All these considerations imply certain requirements in the way results should be relayed to the participants. Such information must be handled with care and sensitivity, and if detrimental consequences are suspected, subjects should be monitored for some time after the end of the study, and psychological counseling may be offered.

b. Handling of information related to mental disorders. Some studies may result in revealing the presence of illness that was previously unknown (for example, a study of depressive symptoms in response to life stress requires carrying out a diagnosis of depression for all participants). This knowledge may have a lot of unwelcome consequences like a change in self-esteem or a change in family perceptions and expectations for a child. On the other hand, research may reveal that some family members do not have the disease now, but they are at higher risk of developing it in the future. People may not want to know that.

6. Social implications of reporting scientific results. Researchers must keep in mind potential effects of the way research conclusions are formulated on the scientific community and society in general. For example, imagine you conducted a research study that supported the idea that homosexuality is inherited. Where should you publish the results? Should it be a narrowly specialized scientific journal or a more popular journal that targets a wider audience including non-scientists? Stating that homosexuality is inherited (and bluntly believing in this statement because it “came from the scientists”) may have deep effects on society. At the same time, you can never be sure of the results of a single research study—there might have been bias; measurements might have been inaccurate; findings may later turn out to be false. Science is a very meticulous (and often inconclusive) process, and care must be taken to report results precisely and accurately, recognizing all potential limitations of the research study, especially if the findings are of social significance.

Case study: The controversy around Cyril Burt

There is much controversy about the work of Cyril Burt, a British psychologist who became famous for his contributions to intelligence testing. In 1942 he became president of the British Psychological Society. He was responsible for administration and interpretation of mental ability tests in London schools. In one of his most famous studies he conducted research with 42 identical twins reared apart. His results showed that the IQ scores of identical twins reared apart were much more similar than that of non-identical twins reared together. He concluded that genetic inheritance in intelligence plays a much greater role than environmental factors (such as education).

In 1956 Burt reported on another study, this time with 53 pairs of identical twins raised apart, where he found a high correlation (0.771) between the IQ scores of the twins. This was exactly the same correlation (to the third decimal place) that he had reported in an earlier study with a smaller sample size. Burt’s research was very influential in forming educational policies in the country, for example, the belief that intelligence is fixed and hereditary led to the practice of using standardized tests to measure intelligence in school children and allocate them to schools based on the results.

After his death in 1971 the British Psychological Society found him guilty of publishing a series of fraudulent articles and fabricating data to support the theory that intelligence is inherited. The case was built on several details that were considered to be highly suspicious.

- There was a very unlikely coincidence of the same correlation coefficient in the two studies.
Case study (continued)

- Some factors that should theoretically influence intelligence (such as mental illness or childhood influences) were suspiciously unimportant in Burt’s data sets, almost a statistical impossibility.

- Identical twins reared apart is an extremely rare sample; there were only three other studies at that time using this kind of sample and none of them had more than 20 pairs of twins as participants.

- Burt’s two female collaborators who worked for him collecting and processing data could not be found, their contact with Burt could not be traced and it was even suspected that these people never existed!

However, some scholars have recently re-examined the claims made earlier and found that evidence of Burt’s fraud is not conclusive, or at least he deserved the benefit of doubt.

In any case, data sets and publications that raise questions regarding their credibility are in themselves an ethical concern, even if they are not falsified intentionally. This is especially true for settings where research findings are used to inform social (for example, educational) policies.

Exercise

At the beginning of this unit you came up with a research proposal related to a research question. Go back and review that proposal. Now that you are equipped with more knowledge about research methodology in psychology, what would you change in your original proposal and why?
Bibliography


Introduction

Psychology in real life

Let’s begin this journey with a thought experiment. Imagine you live in a society of knowledge, a city of dreams called Humanborough. It is a society of rational people who live to maximize their well-being and happiness and who value knowledge over most other things. The most prestigious career is that of a researcher. Crime is rare, and there are no wars. People are modest in their material needs. They would not buy a new phone if the old one still worked. The most popular pastime is learning (taking online courses, attending weekend schools, reading, and so on). Of course, this society faces all the regular human problems: illness and death, disabilities, interpersonal conflicts, jealousy, individuals’ inability to always live up to their potential. Everything as usual, except people of Humanborough are ready to use knowledge as the basis to find a solution.

These people have elected you as their leader. They trust your judgment immensely. Your job is to manage research programmes and their applications to contribute to the well-being of this society. The slogan of your campaign was no less than “Make Humans Better”. The question is, how?

While you are contemplating the scope of the task, note that elements of Humanborough can be seen in today’s real-world popular culture. Here are some examples of films that were built around the idea of using scientific knowledge to “make humans better”.

1. *Limitless* (2011) is a film based on the novel *The Dark Fields* by Alan Glynn. The main character discovers a pill that allows him to use 100% of his brain potential, becoming a meta-human with superb cognitive abilities.

2. *Lucy* (2014): after absorbing special drugs in her bloodstream the main character gains psychokinetic abilities and turns into an invincible warrior.

3. *Avatar* (2009): a special apparatus enables a physically disabled marine to control the body of his “avatar”—an alien life form exploring the planet Pandora.

4. *Robocop* (1987 and 2014) features a cyborg that is a blend of a human and a machine. A human brain controls the immense power of its mechanical body.

5. *The Island* (2005): a powerful corporation is growing clones of rich clients to be used for organ transplantation.

As you know from Unit 1, “Research methodology”, psychology is the scientific study of behaviour and mental processes, and science pursues four goals: description, explanation, prediction and control. Each subsequent goal supersedes the previous one: you need to describe in order to explain, you need to explain in order to predict, and you need to predict to be able to control.

When it comes to explanation, you need to identify a cause of a phenomenon. This is why researchers want to make cause-and-effect inferences and why experiments are so valued as a research method. Knowledge of causes allows you to predict and, in the long run, control the phenomenon under study.

Identifying causes in human sciences has some issues we have to consider. Human behaviour is complex and multi-determined. This means that at any given point of time behaviour is influenced by a whole system of various factors. Some of these factors influence behaviour directly, others indirectly. Some have immediate effects, and the effects of some others only manifest in the long term. So, to study behaviour holistically all these various factors need to be taken into account. However, as you remember, the experiment is the only method that allows cause-and-effect inferences, and the experiment requires that one variable is manipulated and all other variables are carefully controlled (eliminated or kept constant). The dilemma is: we understand that behaviour is influenced by multiple factors simultaneously, but to study it scientifically we have to isolate factors one by one. Research therefore inevitably becomes reductionist.

Holism versus reductionism

Reductionism is an attempt to explain a complex phenomenon by its constituent parts. It may be understood as reducing the whole to its parts. Holism is a methodological position that attempts to gain understanding of the whole in all its complexity. It claims that the whole is bigger than the sum of its parts.

Discuss examples of holism and reductionism from various areas of knowledge.

IB Psychology broadly divides all factors influencing human behaviour into three groups: biological, cognitive and sociocultural. For example, love is a psychological phenomenon—how do we explain it? Some scientists would claim that love is a chemical reaction of the brain. Others would say that love is a mental process, the product of information processing.

TOK

Holism versus reductionism

We know that behaviour is influenced by multiple factors

However, to study them scientifically, we have to isolate them one by one

Figure 2.1 Holism versus reductionism

If you were to decide on your first big project as a leader of Humanborough, what would it be? You may use the films to give you ideas, but you may also be creative. As you read this chapter, you will explore more possibilities.

Discussion

Can you recall any other films or fiction stories based on similar ideas? Share what you have watched with your classmates. To what extent do you think these ideas are real?
Yet others would emphasize the influence of social and cultural norms. All of these claims are reductionist—they are reducing the complex phenomenon of love to its simple constituents. Clearly, the truth lies in combining all these claims. However, one needs to understand the parts before one understands the whole. Experimental research is often reductionist by necessity because it attempts to isolate the effect of one variable.

Biological approach to behaviour looks at behaviour as a product of evolution, genetic inheritance, brain structure or chemical processes in the body. It rests on the following principles.

1 **Behaviour is the product of physiology** (the structure and function of the nervous and endocrine systems). The structure is how a system is constructed; for example, brain damage is a structural problem. The function is how the system operates; for example, low activity in certain parts of the brain is a functional problem of the nervous system, while abnormal levels of hormones are a functional problem of the endocrine system.

2 **Behaviour can be genetically inherited.** The idea that characteristics such as eye colour are inherited raises no objection, but inheritance of behaviours such as perfectionism or preference in movie genres is not so obvious. However, this assumption is made in the biological approach: patterns of behaviour can be inherited as well as physical characteristics. This principle follows from principle 1, because the structure and function of the nervous and endocrine systems are to a large extent genetically determined.

3 **Animal research may inform our understanding of human behaviour.** This principle follows from principle 2: we share a large portion of the genotype with our animal ancestors, and since genotype determines behaviour, animal behaviour in some aspects may be very similar to that of humans. This justifies animal research in psychology.

Note that in the three principles above the term “behaviour” is used broadly and also includes mental processes.

**TOK**

**What is a principle?**

A principle is a broad assumption that guides research in a certain area. What makes a principle different from all other assumptions is its breadth and its fundamental nature. It is fundamental in the sense that if this assumption was not true, research in the area would not make any sense. For example, if we did not assume behaviour to be the product of physiology, the biological approach to behaviour would be meaningless.

Can you name similar principles in some other areas of knowledge?
What you will learn in this section

- Localization of function is the idea that every behaviour is associated with a specific brain region
- It rests on the first principle of the biological approach (behaviour may be the product of brain structure)
- Brain structure
  - cortex
  - cerebellum
  - limbic system
  - brain stem
- Each of these structures and substructures is associated with certain functions, but the term “associated with” implies mild localization only
- Research supporting strict localization
  - Early studies showed that a person with damage to a very specific brain area may demonstrate a very specific malfunction in behaviour
  - Paul Broca (1861): the case study of “Tan”. Broca’s area and Broca’s aphasia—the loss of articulated speech
  - Carl Wernicke (1874): Wernicke’s area, Wernicke’s aphasia—a general impairment of language comprehension, while at the same time speech production is intact
  - Wilder Penfield used the method of neural stimulation in treating patients with severe epilepsy; created a map of sensory and motor cortex known as the cortical homunculus
- Research opposing the idea of strict localization
  - Karl Lashley: the method of induced brain damage in experiments with rats in a maze. The principle of mass action (there is a correlation between learning abilities and the percentage of cortex removed, but not the location of removed cells), equipotentiality (one part of the cortex can take over the functions of another part). Conclusion: memory is distributed rather than localized.
- There needs to be a converging position. Currently, neuroscience supports relative localization: it admits localization for some functions under some conditions, but it also clearly outlines limits of localization.

Inquiry questions

Now that we have assumed that the brain controls behaviour, it would be useful to know what behaviours are controlled by which parts of the brain. Isn’t the idea that all our behaviours can be mapped onto certain brain regions exciting? This way we may be able to “repair” human behaviour by repairing the brain. For example, we might be able to treat criminal inclinations by removing the brain region that controls them.

Is it true, though, that behaviours can be mapped onto specific brain regions? How would you design a research study to find this out?

Think of a short research proposal and explain how your research will contribute to answering the question. Discuss your proposals in class.
Relativity of localization: the split-brain study

- Gazzaniga (1967) and Sperry (1968): research into lateralization—a special case of localization. Language is mostly lateralized in the left hemisphere, but there are exceptions.
  - The right hemisphere can comprehend simple words such as “pencil”: the patient correctly picks a pencil from behind the screen with the left hand
  - The right hemisphere can spell simple words such as “love”, but this differs from person to person
  - Visuospatial abilities are better controlled by the right hemisphere
  - Emotional responses are not lateralized

Conclusions

- Some functions are localized, and brain damage will lead to a loss of function

Localization is limited in the following ways

- Some functions are localized weakly, that is, several brain areas may be responsible for it but some areas are dominant
- Some functions are widely distributed
- Some components of a function may be localized while other components of the same function are distributed in the brain
- Localization is not static (neuroplasticity)

This section also links to:

- Unit 3 “Cognitive approach to behaviour”: Sharot et al (2007) flashbulb memory
- Unit 8, “Developmental psychology”: Saxe and Kanwisher (2003) theory of mind

Brain structure

The nervous system is a system of neurons—cells that perform the function of communication in the body. The central nervous system consists of the spinal cord and the brain.

The major parts of the human brain are the:

- cortex
- cerebellum
- limbic system
- brain stem.

The cortex is the layer of neurons with a folded surface covering the brain on the outside. It is the largest part of the human brain associated with higher-order functions such as abstract thought or voluntary action. Evolutionarily, this part of the brain developed the latest. The cortex is divided into four sections called “lobes”.

- The frontal lobe is associated with reasoning, planning, thinking and decision making, voluntary action, complex emotions, and so on.
- The **parietal lobe** is associated with movement, orientation, perception and recognition.
- The **occipital lobe** is associated with visual processing.
- The **temporal lobes** are associated with processing auditory information, memory and speech.

Note that we say “associated with”: this implies a mild form of localization; that is, the lobe has been demonstrated to be involved in a certain function, but it is not necessarily the only brain structure that influences the function. For example, although temporal lobes are involved in memory processes, a lot of other brain regions play their role in memory too.

There is a deep furrow along the cortex that divides it into the left and right hemispheres. A structure of neurons that connects these two hemispheres is known as the **corpus callosum**.

The cerebellum (“the little brain”) got this name because it looks somewhat like the cortex: it has two hemispheres and a folded surface. It is associated with coordination of movement and balance.

The **limbic system** is an evolutionarily older subcortical structure. It is sometimes it is referred to as the “emotional brain”. It includes several structures, some of which are as follows.

- The **thalamus** has mostly sensory functions. Nerves from almost all sensory organs reach thalamus as a final “hub” before they are connected to the cortex.
- The **hypothalamus** is “below” thalamus in the brain and it is involved in such functions as homeostasis, emotion, thirst and hunger.
- The **amygdala** is involved in memory, emotion and fear.

The **hippocampus** is important for such functions as learning, memory and transferring short-term memory to a more permanent store, spatial orientation.

The **brain stem** is underneath the limbic system and its main function is to regulate the basic vital processes such as breathing or heartbeat. It connects the brain to the spinal cord. This part of human brain is very much like the entire brain found in lower animals such as reptiles.

**Exercise**

There are many websites and even apps that take you on a human brain tour. Search for “brain atlas” or “brain map”. Many apps provide 3-D models. An example is the app called 3DBrain. You may want to download the app and explore.

3DBrain Android: https://tinyurl.com/lfllr9d

Apple: https://tinyurl.com/ac5bw6
Research supporting strict localization

The first research studies that inspired psychologists to investigate the idea of strict localization of function were performed with patients with brain damage. Some of these studies showed that a person with damage to a very specific brain area may demonstrate a very specific malfunction in behaviour. One of the earliest discoveries in this sphere was the discovery of a speech centre by Paul Broca (1861) in the case study of “Tan”.

Case study: Louis Leborgne (“Tan”)

Louis Leborgne, now better known as “Tan”, lost the ability to speak when he was 30. Later, Tan developed gangrene and was admitted to surgery which was to be performed by Paul Broca, a French physician who also specialized in language. By that time “tan” was the only syllable that Leborgne could pronounce. He typically repeated it twice ("tan-tan") and accompanied it with quite expressive hand gestures. His condition remained the same until his death. His inability to speak (or write) was the only malfunction: his intelligence was intact, he understood everything he was asked and tried to communicate back… he just couldn’t utter anything other than “tan”.

Broca carefully described Tan’s condition, which is now known as Broca’s aphasia (the loss of articulated speech). When Tan died, at the age of 51, an autopsy of his brain was carried out and it revealed a lesion in the frontal area of the left hemisphere, in particular a region in the posterior inferior frontal gyrus. This region is now known as Broca’s area.

Broca was cautious about rushing to publish his conclusions. He described 25 additional patients with the same problem before finally asserting that speech articulation is controlled by the left frontal lobe.

However, the area responsible for articulate speech may be more complex than we would like to think.

Tan’s brain (which was carefully preserved by Broca) was re-examined more than 100 years later with the use of modern technology, and it turned out that the lesion had actually been broader than documented by Broca. He did not notice this detail because he decided to preserve the brain intact rather than dissecting it.

ATL skills: Thinking and research

There are some details about Broca’s research that may make you wonder about his methods.

- With a discovery of such significance for its time, why didn’t Broca rush to publish the findings and assert the existence of a brain centre for speech articulation? Why was one case not enough? What do you think is the significance of corroborating your findings by additional research in human sciences, and how does that compare to natural sciences?

- Broca did not cut Tan’s brain open, so did not notice some lesions inside it. If you were Broca, would you make the same decision and preserve the brain for later generations rather than dissecting it and studying it yourself?

Broca’s finding was inspiring. It suggested that other functions can also be mapped onto specific brain areas. Wernicke’s area was discovered by Carl Wernicke in 1874. It is an area located in the temporal lobe of the dominant hemisphere (which is the left hemisphere for most individuals). Wernicke’s
area is responsible for the comprehension of written and spoken language. People with Wernicke’s aphasia have a general impairment of language comprehension, while at the same time speech production is intact. As a result, when they speak they sound really fluent and natural, but what they say is in fact largely meaningless.

Mapping of brain functions was done on a larger scale by Wilder Penfield (1891–1976), a Canadian neurosurgeon. He used the method of neural stimulation. As part of his work he was treating patients with severe epilepsy by destroying nerve cells that initiated the seizure. Before conducting the surgery, though, he would stimulate various parts of the brain while the patient was still conscious, and observe the effects this stimulation had on behaviour. This allowed him to create a map of sensory and motor cortex known as the cortical homunculus. The cortical homunculus is an original model of the body within the brain: it shows the relative representation of various parts of the body in the sensory cortex. As you see, such body parts as hands, tongue and lips are very widely represented in the cortex of the human brain.

One commonality between research of Broca and Wernicke is the method they used: studying a patient with a naturally occurring brain lesion and conducting brain autopsy after the patient’s death. This method has a number of drawbacks, including the following.

- A naturally occurring brain lesion is rarely neat or confined to a specific area.
- As cynical as this sounds, you have to wait until the patient dies.

ATL skills: Communication

In small groups, use brainstorming to answer the following question: What alternative methods can you suggest that would not involve waiting for the patient to die to conduct autopsy?
Remember about ethics of research in psychology!
As a large group, discuss the relative pros and cons of the alternative methods you suggested.

See video

Patients with aphasias sometimes agree to be recorded. Videos of interviews with these patients can be found online, and these give you a good insight into the nature of the malfunction. Here are two examples that you may want to see. Note that aphasias differ in their severity.

Broca’s aphasia: Sarah Scott developed Broca’s aphasia after she suffered a stroke at age 18. https://www.youtube.com/watch?v=1apITvEQ6ew. You can also search for videos that show her progress over the years as she attended speech therapy.
Wernicke’s aphasia: Byron Peterson, a stroke survivor, can be seen at: https://www.youtube.com/watch?v=3oe6YabD0

Mapping of brain functions was done on a larger scale by Wilder Penfield (1891–1976), a Canadian neurosurgeon. He used the method of neural stimulation. As part of his work he was treating patients with severe epilepsy by destroying nerve cells that initiated the seizure. Before conducting the surgery, though, he would stimulate various parts of the brain while the patient was still conscious, and observe the effects this stimulation had on behaviour. This allowed him to create a map of sensory and motor cortex known as the cortical homunculus. The cortical homunculus is an original model of the body within the brain: it shows the relative representation of various parts of the body in the sensory cortex. As you see, such body parts as hands, tongue and lips are very widely represented in the cortex of the human brain.
All these discoveries (by Broca, Wernicke and Penfield) suggest that psychological functions have directly corresponding regions in the brain. We might want to conclude that psychological functions are strictly localized—but it is not that simple.

**Research opposing the idea of strict localization**

**Karl Lashley** (1890–1958) used the technique of measuring behaviour before and after a specific carefully controlled induced brain damage in the cortex of rats. In a typical study he would train a rat to run through a maze without errors in search of food. After learning occurred, he would remove an area from the cortex. Then he would place the rat back at the start of the maze and register the change in behaviour. He removed varying portions of the cortex in different rats, ranging from 10% to 50%. The idea was that if memory of the maze is localized somewhere, then by removing area after area you will finally be able to pinpoint the specific region in the cortex responsible for it. This search turned out to be a failure, so Lashley abandoned his own initial hypothesis. He concluded that memory was distributed rather than localized: a conclusion supported by the following observations.

- **The principle of mass action** based on a correlation observed between the percentage of cortex removed and learning abilities. The less cortex, the slower and more inefficient the learning. The key idea here is that performance deterioration depends on the percentage of cortex destroyed but not on the location of the destroyed cells.

- **Equipotentiality**—this refers to the ability of one part of the cortex to take over the functions of another part of the cortex.

These observations led Lashley to conclude that memory is widely distributed across the cortex. This conclusion is mostly supported today. However, it has been shown that memory is not as evenly (and uniformly) distributed in the cortex as Lashley thought.

**ATL skills: Thinking**

To what extent would you say Lashley’s research was ethically justified? Induced brain damage is a very invasive technique and the research design required the use of many rats.

Conduct a cost-benefit analysis: list the costs as well as potential benefits of this research study. If you were a member of the Ethics Committee, would you approve it?

To some extent the difference in the two extreme positions (the localizationism of Broca, Wernicke and Penfield versus the holism of Lashley) may be explained by the methods they used in their studies. Localizationists relied on aphasia resulting from brain damage. Holists investigated maze-running behaviour. However, learning to run through a maze is in itself a highly complex behaviour that
involves motor and sensory functions, so it may not be suitable enough for the study of localization.

There needs to be a converging position that is a more accurate reflection of localization in the brain. Currently, neuroscience supports relative localization: it admits localization for some functions under some conditions, but it also clearly outlines limits of localization.

Before we formulate these limits let’s look at another research study that demonstrates relativity of localization of function (that is, localization and distribution at the same time): split-brain research by Gazzaniga (1967) and Sperry (1968).

Relativity of localization: the split-brain research

It has to be noted that split-brain studies represent research into lateralization—the division of functions between the two hemispheres of the cortex. Lateralization is a special case of localization.

Research in this area was pioneered by Roger Sperry. Initially, the studies were conducted with animals, for example cats.

An opportunity to replicate the studies with humans emerged when it was discovered that surgically cutting corpus callosum was an effective measure against severe epilepsy with uncontrollable seizures. Roger Sperry was joined by Michael Gazzaniga, and in 1967 Gazzaniga published results of the first research with human split-brain patients. Four of the ten patients who had undergone this surgical procedure by that time agreed to participate. The patients were examined thoroughly over a long time period with various tests.

The aims of the study were to test the theory of lateralization and to see if the two hemispheres have uniquely different functions.

Initial observations showed that patients seemed to be remarkably unaffected by the surgery. There was no change in their personality and intelligence, and one of the patients on awakening from the surgery joked that he had a “splitting headache” and recited a tongue-twister.

The authors devised a technique where the participant had to sit in front of a board and look at the dot in the middle of it. Visual stimuli would then be presented for one tenth of a second either to the left or the right visual field (the far left or far right on the board). Optic nerves from the left eye are connected in our brain to the right hemisphere, and vice versa. So, by presenting the stimulus to the left visual field the researcher “sends it” to the right hemisphere, and stimuli from the right visual field goes to the left hemisphere. Also, a variety of objects were placed behind the screen so that participants could feel them with their hands.
Here are some results obtained in a typical test with split-brain patients.

- When shown the picture of a spoon to the left visual field (connected to the right hemisphere) and asked to name or describe what they saw, the patients said nothing. However, when asked to pick a corresponding object from a group of objects behind the screen, they felt around and picked a spoon (with their left hand, because it is controlled by the right hemisphere). Patients could not explain why they picked the spoon. The right hemisphere saw the spoon and picked it from behind the screen using the left hand, but the centre of speech is in the left hemisphere so the patients were unable to explain what they saw and what they did. This supports lateralization of language in the left hemisphere.

- However, when a simple word, such as “pencil”, was flashed to the right hemisphere, the patients were able to pick a pencil from a group of objects behind the screen with their left hand. This shows that the right hemisphere does have some amount of language comprehension and language is not exclusive to the left hemisphere. Language production, however, is confined to the left hemisphere.

- When the word “heart” was flashed on the screen so that “he” was presented in the left visual field and “art” in the right visual field, the patients said that they saw “art” but pointed (with their left hand) to the card with the word “he” on it. This corroborates the previous findings, but also shows that the two hemispheres process stimuli independently.

- Some patients were able to spell simple words with their left hand, although it differed from person to person. For example, when researchers placed four plastic letters in a pile behind the screen, one patient was able to spell “love” with his left hand (the instruction was simply “spell a word”). Of course, after completing the task the patient was not able to say what word he had just spelled! This shows that the right hemisphere is even capable of language production, but only in the simplest form and only in some patients.

The four findings listed above demonstrate the dominance of the left hemisphere for language. The left hemisphere produces speech and makes a person consciously aware of something. However, this lateralization is not strict: some forms of language production and comprehension can be performed by the right hemisphere also. Moreover, it differs somewhat from person to person.

- The right hemisphere performs better in tasks that involve visual construction. A redrawn picture was a much closer match to the original when done by the left hand (controlled by the right hemisphere) than the right hand (controlled by the left hemisphere), even in right-handed people. So, if you think you cannot draw, do you think switching hands might help?

- Both hemispheres independently are capable of forming an emotional response. In other words, emotions are not lateralized. In one series of tests, various objects were flashed on the screen and among them the picture of a
A nude woman. This immediately evoked an emotional response irrespective of where on the screen the picture was flashed. If it was flashed to the left hemisphere, the patient laughed and identified it. Whereas, the researchers reported of a female patient: “When it was later presented to the right hemisphere, she said in reply to a question that she saw nothing, but almost immediately a sly smile spread over her face and she began to chuckle. Asked what she was laughing at, she said: ‘I don’t know... nothing... oh – that funny machine’.” (Gazzaniga 1967: 29).

Exercise

To strengthen your knowledge of Sperry and Gazzaniga’s procedure and findings, investigate this interactive online game from Nobleprize.org, the official website for the Nobel Prize: https://www.nobelprize.org/educational/medicine/split-brain/

See if you can predict all the findings and explain all the results.

Conclusions

Summarizing split-brain studies and other research accumulated over the years, does all this mean that localization of function is relative?

Some functions are indeed localized in very specific parts of the brain, and damage to this part will lead to loss of the function. Examples that have been discussed here include Broca’s area for the production of articulate speech and Wernicke’s area for language comprehension. As you read this book you will come across a lot of examples and research studies that support localization of some other functions. For example, Sharot et al (2007) demonstrated that selective activation of left amygdala is responsible for the formation of flashbulb memories—a special memory mechanism when situations that are unexpected and emotionally laden get “imprinted” in the brain with perceptual clarity (see Unit 3, “Cognitive approach to behaviour”). Saxe and Kanwisher (2003) showed that understanding when another person’s belief is false is localized in tempo-parietal junction (see “Theory of mind” in Unit 8, “Developmental psychology”). Maguire et al (2000) found that spatial memory in London taxi drivers is localized in the hippocampus (see later in this unit). Most of modern discoveries in this area are made with the use of non-invasive methods—brain imaging technology.

However, the quest for a complete map of localized functions has reached its limits. Here are the most prominent points.

- Some functions are localized weakly; that is, several brain areas may be responsible for a function but some areas are dominant. There
were many examples of weakly localized (lateralized) functions in Sperry and Gazzaniga’s research. Although the left hemisphere was consistently shown to be dominant for language, the right hemisphere was also shown to be capable of understanding some simple language. Scientists have been generally more successful in establishing strict localization for sensory and motor functions than for higher-order cognitive functions such as memory, thinking and learning.

- Some functions are **widely distributed**. An example from Sperry and Gazzaniga’s research would be the ability of both the hemispheres to form an emotional reaction independently of each other. Karl Lashley demonstrated that maze-running memory was widely distributed in the cortex of the brain of rats.

- Some **components of a function** may be localized while other components of the same function are distributed in the brain. One example involves language: we cannot say that language (as a holistic function) is localized somewhere in the brain, but some specific components of language (such as the production of articulate speech in Broca’s area) are.

- Localization is **not static**. Functional areas move about. For example, people with damage to a functional brain area may learn to “respecialize” other brain areas to perform this function. This is known as **neuroplasticity**: many neurons can potentially perform a variety of functions, and the brain can take over the functions from the damaged parts. Like localization, neuroplasticity is relative—there are limits. You will learn more about this in the next section.

These considerations outline the modern views on localization of function. Function mapping is far from being over, and a lot of amazing things are still out there to be discovered.

**TOK**

Some of the concepts you encountered in this chapter are used in other areas of knowledge too. Discuss how these concepts are used elsewhere and make sure to ask classmates whose subject choices are different from your own. Focus on the concepts of:

- function (for example in mathematics, history, art)
- structure (for example in physics, chemistry, TOK)
- localization (for example in geography, astronomy)
- system (for example in natural sciences, mathematics)
- the relative and absolute (for example in physics)
- the weak and strong (for example in religious knowledge systems)
- the static and dynamic (for example in history, TOK).

**Psychology in real life**

Think about the potential practical applications of the idea of localization. How can it be used to improve the life of people in Humanborough? Now that you know that localization is relative and some functions are more localized than others, consider these questions.

- Would you fund a scientific programme to search for all strictly localized functions?
- Would you study the brains of children to determine their future abilities?
- Would you research the possibility to relocalize psychological functions in the brain?
- Localization of which psychological functions would you like to establish as a priority? Would it be aggression or attraction, for example? Would it be something else?
- Would you authorize animal research in this area?
- Which methods of research would you most invest into?

Give your reasoned arguments and present your vision in class.
Neuroplasticity

Inquiry questions

Behaviour is a product of brain structure, but to what extent can the brain itself be changed? For example, if a patient’s Broca’s area is damaged, would it be possible to re-grow this area in another region of the brain?

Apart from brain injury, do you think the brain can change itself in the course of our life? For example, what happens when you learn something, such as a new skill?

If the brain does change itself, what practical applications using this process can you think of? Share your ideas in class.

What you will learn in this section

- Definitions
  - Neuroplasticity is the ability of the brain to change through the making and braking of synaptic connections between neurons. Causing factors are both genetic and environmental.
  - Different scales of neuroplasticity—from synaptic plasticity to cortical remapping
  - Synaptic plasticity depends on the activity of neurons

- Remapping of the sensory cortex
  - Merzenich et al (1984): cortical remapping of sensory inputs from the hand occurs within 62 days in owl monkeys—adjacent areas spread and occupy parts of the now unused area for the amputated digit.

- Neuroplasticity as a mechanism of learning
  - Neuroplasticity occurs on a regular basis in our daily lives. When you learn your brain gradually reshapes itself.
  - Example 1—Draganski et al (2004): learning a simple juggling routine increases the volume of grey matter in the mid-temporal area in both hemispheres. Lack of practice makes this area shrink, but not to the original size.
  - Example 2—Draganski et al (2006): learning large amounts of abstract material leads to an increase of grey matter in the parietal cortex and the posterior hippocampus
  - Example 3—Maguire et al (2000): neuroplasticity is observed in natural settings too. London taxi drivers, as compared to controls, have redistributed brain matter in the hippocampus: more in the posterior, less in the anterior

- Practical applications
  - Bach-y-Rita et al (1969): sense substitution
  - Human echolocation
  - Brain-machine interfaces

This section also links to:

- localization of function, brain scanning technology, the use of animal research
- Unit 3 “Cognitive approach to behaviour”: cognitive processing in the digital world
- Unit 8 “Developmental psychology”: brain development
Definitions

Neuroplasticity is the ability of the brain to change throughout the course of life. The change occurs through the making and breaking of synaptic connections between neurons. In this process neural networks in the brain literally change their shape. The reasons for such changes are both genetic (normal pre-programmed development of the brain) and environmental (for example injury, brain damage or simply learning new skills).

Neuroplasticity can be observed on different scales. On the smallest scale, at the level of a single neuron, it takes the form of synaptic plasticity: the ability of the neuron to form new synaptic connections and break up the old ones. On the largest scale, neuroplasticity takes the form of cortical remapping: the phenomenon when brain area X assumes the functions of brain area Y, for example due to injury.

Synaptic plasticity depends on the activity of neurons. If two nearby neurons are frequently activated at the same time, a synaptic connection between them may gradually form. Similarly, if two neurons are rarely activated together, the existing connection may gradually fall apart. This has been summarized like this: “neurons that fire together, wire together” (which was originally said by Carla Shatz and is quoted in Doidge 2007) and “neurons that fire out of sync, fail to link” (Doidge 2007: 63–64).

Remapping of the sensory cortex

One of the early studies of neuroplasticity on the level of cortical remapping was done by Merzenich et al (1984). Researchers studied the cortical representation of the hand in eight adult owl monkeys. The procedure involved three steps.

1. Sensory inputs from all the hand digits (fingers) were mapped in the cortex. To do this, electrodes were inserted in the cortical area known to be responsible for sensation from the hand, then researchers stimulated various areas on all the fingers one by one and noted which electrode was responding to the stimulation. Monkeys were anesthetized before this procedure.
2. The third digit (the middle finger) on the monkey’s hand was amputated.
3. Sixty-two days later a remapping was done to see how the cortical area responsible for sensitivity from the hand changed after amputation.

Results of the first mapping showed that there were five distinct areas in the brain, each responsible for one finger, and adjacent fingers were represented in adjacent areas in the cortex. What happens to the area responsible for the third finger after this finger is amputated? It was found that the adjacent areas (those responsible for sensitivity from digits 2 and 4) spread and occupied parts of the now unused area. The areas responsible for digits 2 and 4 became larger while the areas responsible for digits 1 and 5 stayed the same. It was concluded that cortical remapping of sensory inputs from the hand occurs within 62 days in owl monkeys.
Neuroplasticity as a mechanism of learning

Neuroplasticity is not confined to making up for damage. It occurs on a regular basis in our daily lives. For example, neuroplasticity is thought to be the brain mechanism of learning. When you learn, your brain gradually reshapes itself.

Draganski et al. (2004) conducted a study to find out whether the human brain can really change structure in response to environmental demands. The researchers used a random sampling design and a self-selected sample—they randomly allocated a sample of volunteers into one of two groups: jugglers and non-jugglers. They made sure that both groups had no experience of juggling before the start of the experiment. The first brain scan was performed at this point. Participants in the juggler group subsequently spent three months learning a classic juggling routine with three balls. The second brain scan was performed. Then the participants spent another three months where they were instructed not to practise juggling. Finally, the third brain scan was performed after this non-practice period. The control group (the non-jugglers) just lived their daily lives and had their brains scanned three times on the same schedule as the jugglers.

Figure 2.15 The procedure followed in Draganski et al. (2004)
Comparison of brain scans in the two groups prior to the start of the experiment showed no differences in brain structure. At the second scan, however, the juggling group had significantly more grey matter in some areas of the cortex, most notably the mid-temporal area in both hemispheres. These areas were known to be implicated in coordination of movement. At the time of the third scan these differences decreased, but the amount of grey matter in these areas in jugglers was still greater than at the time of the first scan. Also, there was a correlation between juggling performance and the amount of change: brain changes in participants who trained better were more pronounced. In other words, as you learn a simple juggling routine, certain areas of your brain grow. When you fail to practise, they shrink back significantly (perhaps not to the initial state, though).

**ATL skills: Research**

The study of Draganski *et al* (2004) is an example of how different research methods can be combined in a single study. Is this an experiment or a correlational study?

On the one hand, there is random allocation into groups (juggling versus non-juggling). On the other hand, the researchers computed a correlation between amount of juggling and grey matter growth. In cases like this you need to determine the main research method and the supplementary methods. The main research method is the one used to test the hypothesis. In Draganski *et al* (2004) the aim was to see whether environmental demand (juggling) leads to a change in brain structure (grey matter volume). This implies causation. This hypothesis was tested in an experiment. When the researchers got their answer (which was yes), they additionally looked at the relationship between amount of learning and rate of grey matter growth. Correlation was therefore a supplementary method used to clarify the main finding.

Knowing what method was used, what can you say about the methodological quality of the study (sampling, credibility, generalizability and bias)?

Do you want to know what areas of your brain are growing as you are reading this book? We have some evidence for learning a large amount of abstract material in preparation for a medical exam, and that is close enough. Draganski *et al* (2006) looked at 38 medical students and 12 control subjects matched for age and sex. The first scan was obtained three months before the medical exam, the second scan on the first or second day after the exam, and the third scan three months later (after the exam the students had a break). Results showed that, although there were no differences in regional grey matter at baseline, there were two major changes occurring in the brains of the medical students.

- There was an increase of grey matter in the parietal cortex in both hemispheres. The volume of grey matter in this region did not decrease by the time of the third scan. Studying for a medical exam has a more lasting impact on the brain than learning a juggling routine!
- The changes stay with you even after a study break.
- There was an increase of grey matter in the posterior hippocampus. The pattern was different here. Grey matter gradually increased from the first scan to the third; that is, grey matter in the hippocampus continued to grow even after the exam.

The results of the study were in line with prior knowledge that these areas were involved in the formation of new memories. However, the changes in posterior hippocampus after the exam were surprising. They contradicted the hypothesis because any increase in grey matter after the exam could not be induced by learning. Based on previously discovered properties of the hippocampus, the researchers suggest the following explanation. Stress is known to reduce grey matter volume in hippocampal regions. This may have resulted in two opposite effects.
on hippocampus simultaneously between the first and the second scan: learning increased grey matter in the posterior hippocampus but exam stress decreased it. After the exam this negative influence of the exam stress was corrected and the lost hippocampal volume was restored, while the grey matter that was formed due to learning remained.

Exercise

A very useful exercise to develop critical thinking in psychology is coming up with alternative explanations. Review the results of Draganski et al (2006)—the researchers saw that grey matter increase in the parietal cortex was abrupt: it increased from the first measurement to the second and stayed on the same level by the third measurement; grey matter increase in the hippocampus was gradual from the first measurement to the third. Go online and read more about the known functions of the hippocampus (for example, start with Wikipedia). What alternative explanations of Draganski et al’s (2006) findings can you think of? How plausible are they?
Another well-known study showing both localization of function and neuroplasticity for spatial memory is that of London taxi drivers. The value of this study is that it looks at human neuroplasticity in a natural setting.

Maguire et al (2000) investigated the brains of London taxi drivers, a group chosen for their extensive navigation experience. The researchers hypothesized that the structure of the hippocampus would be different because hippocampus from prior animal studies was known to be involved in spatial abilities. Taxi drivers in London undergo an intensive training programme on how to navigate in the city, and have to pass a set of stringent exams to be licensed. Participants in this study were 16 right-handed male licensed taxi drivers. Their average pre-licensing training time was 2 years and the average experience as a taxi driver was 14.3 years (with a range from 1.5 to 42 years of experience). All taxi drivers had healthy medical profiles.

Brain scans of control subjects were taken from the database of brain scans at the same unit where brain scans were performed with taxi drivers. The scans were obtained by magnetic resonance imaging (MRI). It was important to make the comparison groups as equivalent as possible in terms of potential confounding variables, so some exclusion criteria were applied to the control subjects. Subjects below 32 and above 62 years old were excluded, as well as subjects who were female, left-handed or had any health issues. This resulted in the selection of (brain scans of) 50 healthy right-handed male subjects who did not drive a taxi.

Results indicated an increased brain matter volume in the brains of taxi drivers as compared to control subjects in the posterior hippocampus. At the same time, control subjects had greater volumes of grey matter in the anterior hippocampus. This meant there was no difference between the groups in terms of the general volume of the hippocampus, but there was significant redistribution of grey matter from the anterior to posterior hippocampus in the brains of taxi drivers. Brain matter “shifted” from the back to the front.

However, this study is a quasi-experiment. The researchers did not randomly assign people to be either taxi drivers or controls. The results obtained are therefore essentially correlational and cause-and-effect inferences cannot be made. It seems plausible to suggest that driving a taxi in London leads to redistribution of grey matter in the hippocampus, but we cannot be certain. One of the possible alternative explanations for this finding is that people with larger grey matter volumes in the posterior hippocampus (and lesser volumes in the anterior) are naturally predisposed to choose professions that depend on navigational skills. In other words, people become taxi drivers because they have a special brain, not the other way around.

To test this alternative explanation, Maguire et al examined the correlation between hippocampal volume and amount of time spent as a taxi driver. Grey matter volume in the posterior hippocampus correlated positively and significantly with experience as a taxi driver ($r = 0.6, p < 0.05$), and there was a reverse relationship with grey matter volume in the anterior hippocampus ($r = -0.6, p < 0.05$). Note that the effect size is exactly the same but with a different sign. This led researchers to conclude that, since grey matter volumes change as taxi drivers’ experience increases, these differences are indeed the result of neuroplasticity.

Redistribution of grey matter in the hippocampus itself can be explained by attributing different functions to the two regions: it is now accepted that posterior hippocampus is involved when previously learned spatial information is used, whereas anterior hippocampus is responsible for learning new spatial information.
ATL skills: Research
If you have difficulty interpreting the expression \( r = 0.6, p < 0.05 \), it is time to go back and review “Correlational studies” in Unit 1, “Research methodology”.

Practical applications
All the evidence discussed so far challenges the idea that the brain is “fixed” and that certain psychological functions are hard-wired in certain parts of the brain. At least to some extent, the brain is a plastic structure that can adapt itself to the demands of the environment. This raises the question: can we use brain plasticity for practical purposes? Potential applications are countless.
Can we rewire the visual cortex of blind people so that they can “see” using some other senses? For a person with a mechanical prosthetic limb, can we implant electrodes into the motor cortex and train the brain to control the artificial limb?

The idea that other senses may be used to make up for the lost sense is known as sense substitution. One of the first neuroscientists to introduce this idea was Paul Bach-y-Rita (1934–2006). His pioneering work in sensory substitution started with the invention of a chair that allowed congenitally blind people to “see” (Bach-y-Rita et al. 1969). It was a chair with a large camera behind it. The camera scanned the area and converted the image to an electrical signal sent to 400 vibrating stimulators attached to the back of the chair. The blind subject sat in the chair and learned to recognize visual stimuli against his or her back. This way subjects gradually learned to recognize images: objects, shadows and faces. The brain interpreted tactile information and converted it into visual images, which requires neuroplasticity to change the visual cortex and other brain areas. As Bach-y-Rita said: “You see with your brain, not your eyes”.

TOK
Paul Bach-y-Rita’s research was inspired by his personal family history. In 1959 his father had a stroke that yielded massive damage to his brain leaving him mostly paralyzed and resulting in loss of speech. At that time, psychological functions were still believed to be strictly localized and hard-wired into specific areas of the brain, so the paralysis and other impairments were believed to be permanent. However, Paul’s brother George did not give up and devised his own rehabilitation programme for his father, turning daily routines into exercises. Over several years, his father restored most of the motor functions and regained speech, a recovery that was named unprecedented.

From the TOK point of view, this is an example of how personal knowledge can affect shared knowledge within a discipline.
Another application of neuroplasticity is **human echolocation**. Some blind people can acquire the ability to see around them with echoes: they produce clicking sounds with their mouth and analyse echoes as the sounds bounce off from the objects in front of them. Studies have demonstrated that this auditory information in such people is processed in the visual rather than auditory cortex areas (Thaler, Arnot, Goodale 2011).

Yet another promising area is **brain-machine interfaces**. These include artificial sensory organs and bionic limbs that can be controlled by thought.

**See video**

Here are some TED talks worth seeing:

- Michael Merzenich reviews modern evidence of neuroplasticity and talks about a range of potential practical applications in “Growing evidence of brain plasticity” (2004): https://www.ted.com/talks/michael_merzenich_on_the_elastic_brain
- Daniel Kish was blind since the age of 13 months. He taught himself to use echolocation instead of vision. He demonstrates how he does it. "How I use sonar to navigate the world" (2015): https://www.ted.com/talks/daniel_kish_how_i_use_sonar_to_navigate_the_world
- Tan Le demonstrates a brain-computer interface that allows you to control your computer with your mere thoughts in “A headset that reads your brainwaves” (2010): https://www.ted.com/talks/tan_le_a_headset_that_reads_your_brainwaves

**Psychology in real life**

Now that you know the research behind neuroplasticity, how could you use it to “make humans better” in Humanborough? Does it give you any more ideas?

Think about one concrete idea that uses the ideas of neuroplasticity to solve a practical problem. This could be something similar to sense substitution, human echolocation or brain-machine interfaces, but you might also think of something completely different. Do not worry about how realistic it is. Present your idea in class and discuss with your classmates.
Inquiry questions

Some aspects of human behaviour can be explained by chemistry—but can we use chemistry to control people’s behaviour in constructive ways? Here are a couple of very specific examples. Do you think we can invent a drug that will make people fall in love with each other? Can we invent a drug that will make people tell the truth?

What you will learn in this section

- Nervous system processes
  - The structure of a neuron
  - The nature of information transmission in the nervous system is partly electrical and partly chemical
  - Electrical processes: threshold of excitation, action potential
  - Chemical processes: neurotransmitters and how they function
  - Excitatory and inhibitory neurotransmitters
  - Agonists and antagonists
- Limitations in neurotransmitter research (increasing levels of neurotransmitter X results in a change in behaviour Z)
  - X may function as an agonist or antagonist for Y, which affects Z (indirect effect)
  - X may serve as a trigger for a long-lasting process of change (postponed effect)
  - X is usually not the only factor affecting Z (multi-determination)
  - X is never the only factor that changes (side effects)
- Effect of serotonin on prosocial behaviour
  - Serotonin reduces acceptability of personal harm and in this way promotes prosocial behaviour
- Crockett et al (2010)—participants solved moral dilemmas after receiving a dose of either citalopram (an SSRI) or placebo.
- Effect of dopamine on romantic love
  - Fisher, Aron and Brown (2005): looking at pictures of loved ones is associated with higher activity in the dopaminergic pathway—a system that generates and transmits dopamine and increases dopamine-related activity in the brain
- The role of dopamine in Parkinson’s disease
  - Freed et al (2001): transplantation of dopamine-producing neurons in the putamen of patients with severe Parkinson’s disease results in some clinical benefit in younger but not older patients
- The role of serotonin in depression
  - The serotonin hypothesis: low levels of serotonin in the brain play a causal role in developing depression
  - Typical study: if a drug that is known to affect serotonin (for example, an SSRI) leads to a reduction of symptoms in the experimental group, it is concluded that the level of serotonin is the cause of depression
The nervous system processes

The first principle of biological approach to behaviour stated: behaviour is the product of physiology (the structure and function of the nervous and endocrine systems). So far, we have considered the structure of the nervous system—we have looked at localization of function and neuroplasticity as structural phenomena. Now let’s consider the function of the nervous system: how does it work? What processes are used to transfer information in the nervous system? What are the behavioural correlates of these processes?

The nervous system is a system of neurons, the nervous cells. A neuron consists of three parts: the body (soma), dendrites and axon. Dendrites and axon are filaments that extrude from the soma: typically multiple dendrites but always a single axon. The function of dendrites (and soma) is to receive signals from other neurons, while the function of axon is to transmit signals further. Where the axon of one neuron approaches a dendrite or soma of another neuron, a synapse is formed. This means that a synapse (or a synaptic gap) is a structure that connects two neurons: the word “synapse” comes from the Greek synapsis meaning “conjunction”. Each neuron on average has about 15,000 connections with other neurons, so it is a very elaborate network.

The nature of information transmission in the nervous system is partly electrical and partly chemical. Every neuron has a certain threshold of excitation received from the other neurons, and if the sum excitation exceeds this threshold, the neuron “fires”—generates a brief pulse called action potential that travels along the axon to other neurons, passing the excitation further. Note that the action potential is all-or-none: it either fires or not, and there is no such thing as “firing weakly” or “firing strongly”. This might remind you of the coding languages used in computers: essentially, they all boil down to a sequence of 0s and 1s.

The pulse reaches the end of the axon and there, at the synaptic gap, the mechanism of transmission becomes chemical. This happens as follows. When the action potential reaches the end of the axon, a neurotransmitter is released from the axon terminal into the synaptic gap. Neurotransmitters are chemical messengers. They are constantly synthesized in the neuron and moved to the axon terminal to be stored there. A released neurotransmitter is available in the synaptic gap for a short period during which it may be destroyed (metabolized), pulled back into the pre-synaptic axon terminal through reuptake, or reach the post-synaptic membrane and bind to one of the receptors on its surface.
If the neurotransmitter binds to a receptor in the post-synaptic membrane, this process changes the membrane potential and so contributes to activating an electric pulse in the post-synaptic neuron. Here the chemical mechanism becomes electrical again.

**ATL skills: Thinking and self-management**

This partially electrical and partially chemical process seems quite complicated. What do you think is the evolutionary rationale behind it? Why did it evolve in this way?

Try to visualize the whole process. If you cannot, view the example video given below or search for other “neurotransmission” videos online.

There are many different neurotransmitters. Their exact number is unknown but more than 100 have been identified. All neurotransmitters are broadly divided into two groups: excitatory and inhibitory. **Excitatory neurotransmitters** allow the impulse to cross the synapse. They produce stimulating effects on the brain. **Inhibitory neurotransmitters** stop the impulse, preventing it from crossing the synapse. They produce calming effects on the brain.

These neurotransmitters are always in a state of dynamic balance. When excitatory or inhibitory neurotransmitters are out of their optimal ranges in the brain, this may cause various behavioural malfunctions such as mental disorders.

Neurotransmitters themselves are affected by agonists and antagonists. **Agonists** are chemicals that enhance the action of a neurotransmitter. **Antagonists** are chemicals that counteract a neurotransmitter and so prevent a signal from being passed further.

Many drugs function as agonists or antagonists. For example, a class of drugs known as SSRIs (selective serotonin reuptake inhibitors) do exactly what their name suggests: selectively inhibit (block) the reuptake of the neurotransmitter serotonin from the synaptic gap. This increases the concentration of serotonin in the synapse. SSRIs have been shown to be effective against depression.

As you see, neurotransmission is a complex process determined simultaneously by multiple factors. Imagine we have artificially increased the level of neurotransmitter X in the brain and this resulted in a change of behaviour Z (for example, elevated mood). Can we say that neurotransmitter X influences elevated mood? Yes, to a certain extent, but with a lot of limitations to be kept in mind.

- X may function as an agonist for neurotransmitter Y, which in turn may affect behaviour Z. In other words, the effects of neurotransmitters may be indirect, sometimes with many links between the “cause” and the “effect”.
- X may serve as a trigger for a long-lasting process of change in a system of interconnected variables. In other words, the effects of X on Z may be postponed.
- X is usually not the only factor affecting Z.
- X is never the only factor that changes. As you artificially increase the level of X, this may result in various side effects.

Research into the influence of neurotransmission on behaviour will therefore always be reductionist in the sense that we need to manipulate one variable (X) and assume that it is the only variable that changes. Nevertheless, with all the limitations in mind, research into the influence of neurotransmission on behaviour is highly intriguing, partly because it gives us a key to influencing human behaviour through intake of chemicals.
Areas that have been shown to be affected by neurotransmitters have included mood, memory, sexual arousal and mental illness, among many others. Let’s look at some of the effects of specific neurotransmitters.

**Effect of serotonin on prosocial behaviour**

Crockett et al (2010) investigated the effect of serotonin on prosocial behaviour. Serotonin is an inhibitory neurotransmitter that is involved in sustaining stable mood and regulating sleep cycles, for instance. This example has been selected intentionally to demonstrate that researchers attempt to establish links between seemingly distant variables. Effects of neurotransmission on something such as mood or fatigue are believable because these are rooted in biological processes, but prosocial behaviour seems to be a person’s own choice or free will. How can a person’s free will be affected by a biological factor?

**TOK**

**Determinism versus free will**

Determinism is the philosophical position that claims everything in the world has a preceding cause. It says that in order to understand the world we need to identify the causes. Modern sciences—especially the natural sciences—are largely deterministic.

An opposing position—known as teleology (from “teleo”, meaning “aim”)—asserts that everything in the world has a purpose. To understand the world we need to understand where it is going. Religion is usually teleological.

When it comes to human beings, some philosophers assume strong determinism, saying that every action is determined by preceding potentially identifiable causes. However, others believe that humans, unlike everything else in the world, are able to choose their actions freely, often despite the factors that influence them.

What do you think?

A sample of volunteers was recruited for the study. It included 30 healthy subjects (mean age 26).

The experiment followed a repeated measures design with two conditions. In condition 1 participants were given a dose of citalopram. This drug is a highly selective serotonin reuptake inhibitor (SSRI): a chemical that blocks reuptake of serotonin from the synapse, in this way boosting its concentration and prolonging its effects. In condition 2 (the control) participants were given a placebo (a harmless substance with no active effect). The design was counter-balanced, and this was a double-blind study.

After taking the drug, participants were given a series of moral dilemmas that involved choosing between a utilitarian outcome (saving five lives) and aversive harmful actions (such as killing an innocent person). Aversive harmful actions in the scenarios were of two types: personal (for example, pushing a man off a bridge to stop a train and prevent it from hitting five people) and impersonal (for example, pressing a lever to divert a train off a track where it will hit five people to a track where it will hit one).
**Results showed that responses in the impersonal version were unaffected by citalopram. However, after receiving a dose of citalopram participants were less likely to push the man off the bridge in the personal scenario than participants in the placebo condition. Would you push a man off the bridge to stop the train and prevent it from hitting five other people? If you were an average participant in this study, you would probably say no, but after receiving a dose of citalopram you would be opposed to the idea even more strongly. Note that your judgment on the impersonal version (pressing a lever) would be unaffected. Researchers concluded that serotonin reduces acceptability of personal harm and in this way promotes prosocial behaviour. It modulates reactions of the brain to emotionally salient situations so that inflicting harm on other people is judged as less acceptable. A limitation of the study that the authors recognize is that citalopram intake induced slight nausea. This might mean that participants could figure out what condition they were in on that trial. However, it is not possible to estimate the extent to which this might have influenced the results.**

---

**Effect of dopamine on romantic love**

**Fisher, Aron and Brown (2005)** conducted a study of the neural mechanisms of romantic love. This study suggested the central role of dopamine in the brain response to loved ones. Dopamine is an excitatory neurotransmitter that is involved in our desire to get things done (motivation), in controlling the brain’s reward and pleasure centres and in regulating emotional responses. Ten men and seven women who were currently “intensely in love” (but not with each other) were recruited for the study by word of mouth as well as through flyers. The mean age was 21 years and the mean reported duration of being in love was 7 months. All participants were placed in a functional magnetic resonance imaging (fMRI) scanner and engaged in a standardized procedure involving looking at photographs while their brains were being scanned. There were four stages.

1. For 30 seconds each participant viewed a photograph of his or her beloved person.
2. Participants were given a 40-second filler activity which was to count back from a given number.
3. For 30 more seconds participants viewed a photograph of an emotionally neutral acquaintance.

4. The final stage was another 20 seconds of counting back from a number.

These four steps were repeated six times, so the total procedure lasted for 720 seconds (12 minutes).

Results showed a specific pattern of activation in the brains of participants in response to the photographs of their loved ones: activation was observed in dopamine-rich neural systems, primarily the ventral tegmental area (VTA) and caudate nucleus. Both these regions are rich in dopamine and form the key part of the so-called dopaminergic pathway—a system that generates and transmits dopamine and increases dopamine-related activity in the brain. It is a reward system because dopaminergic activity is associated with motivation and feelings of pleasure. In this way, dopamine activity in the brain plays a role in romantic love.

**Exercise**

If you are interested to see how the authors themselves justified their conclusion, review the original article:

https://tinyurl.com/mk7hesj

Reading original research articles from time to time is useful to give you a deeper understanding of real-world psychological research.

Helen Fisher’s website is also worth exploring: http://www.helenfisher.com/

It has a number of Helen Fisher’s talks relating to the psychology of love.

---

**ATL skills: Thinking**

You will learn later in this unit that fMRI is based on the principle of measuring blood flow in specific parts of the brain. The assumption is that the more blood flow there is in a brain region, the more active the region is at this particular moment. So what was observed in Fisher, Aron and Brown’s (2005) study is that when you are looking at a picture of a person you love, there is more blood flow in the regions of your brain that are known to produce dopamine.

From this the researchers inferred the role of dopamine in romantic love. To what extent do you think it is a substantiated conclusion? Or is it far-fetched?
The role of dopamine in Parkinson’s disease

Freed et al (2001) studied the role of dopamine in Parkinson’s disease. Parkinson’s disease is a degenerative disorder that mainly affects the motor functions of the nervous system. The early symptoms of the disease are shaking, rigidity, and difficulty with movement and walking. Later in the development of the disease, thinking and behavioural problems also occur. Currently there is no cure for Parkinson’s disease and the exact causes are unknown. In the study by Freed et al, the sample consisted of 40 patients who were 34–75 years old and had severe Parkinson’s disease, with the mean duration of 14 years.

The sample was randomly divided into two groups: the experimental group received a transplant of nerve cells and the control group underwent sham surgery. In the transplant group, nerve tissue containing dopamine-producing neurons was taken from embryos aborted 7–8 weeks after conception and transplanted into the patients’ putamen—a structure of the limbic system involved in movement regulation. All surgeries were performed with the patient awake. Local anesthesia was administered to the skin of the forehead and four holes were drilled through the frontal bone, after which the tissue was transplanted through long needles. In the sham surgery group, holes were drilled in the skull but the dura (a thick membrane that surrounds the brain) was not penetrated. Otherwise, the procedure was identical.

The protocol of the study and the consent form describing the risks and potential benefits were approved by the Ethics Committee. A separate written informed consent form was used for the women who donated fetal tissue from abortions.

ATL skills: Thinking

We are dealing with an ethically sensitive study. It involved aborted embryos, nerve tissue transplants and a sham surgery. To what extent are these procedures ethically justified by the anticipated benefits of the study?

Discussion

Would you approve this study if you were a member of the Ethics Committee? State your position in class and discuss it with others.

A number of measures were taken both before and after the surgery. They included clinical observations and interviews, and brain scans—positron emission tomography (PET). All patients were followed up longitudinally for one year.

Results of the study indicated the following.

- Irrespective of the age group, PET scans revealed increased growth of dopamine-producing cells in the putamen.
- A reduction of symptoms by 28% was found in the patients in the transplant group, but only the younger ones (aged 60 or younger). No improvement was registered in the older subgroup of patients (aged over 60).

The overall conclusion was that transplantation of dopamine-producing neurons in the putamen of patients with severe Parkinson’s disease results in some clinical benefit in younger but not older patients. Less response to treatment in the older patients despite successful growth of dopamine neurons may be attributed to lower neuroplasticity of the brain.
The role of serotonin in depression

Serotonin has been shown to be involved in the symptoms of major depressive disorder. If you study Unit 5, “Abnormal psychology”, as an option you will learn a lot more about this topic. The serotonin hypothesis states that low levels of serotonin in the brain play a causal role in developing depression. Studies have mainly involved clinical trials with two groups of patients. The experimental group would be given a drug that affects levels of serotonin in the brain and the control group would be given a placebo (a harmless substance that the patient believes to be a drug), after which the symptoms of depression would be compared. If a drug that is known to affect serotonin (for example, an SSRI) leads to a reduction of symptoms in the experimental group, it is concluded that the level of serotonin is the cause of depression.

However, this logic has a number of limitations. First, drugs affect neurotransmitters within minutes, but the behavioural effects do not manifest immediately. Sometimes it takes weeks. This suggests that the influence may be indirect or there could exist a longer path where changing levels of serotonin is just one stage. For example, one theory suggests that increased stress can damage neurons in the hippocampus and so lead to depression, whereas SSRIs restore the damaged neurons gradually, alleviating the symptoms (Taupin 2006). Second, not all patients benefit from drugs. This suggests that the link between serotonin and depression is not universal (that is, not applicable to all people without exception).

Whether it is universal or not, direct or indirect, the presence of the link between serotonin and depression is not questioned. Recently, depression was also linked to a particular gene—the serotonin transporter gene 5-HTT—and it was shown that this gene determines one’s vulnerability to developing depression in response to stressful life events (Caspi et al 2003).

**Exercise**

Neurotransmission can explain drug addiction. Some scientists say that the mechanism of drug addiction is that a synthetic drug replaces the naturally produced neurotransmitter and the organism starts to depend on the intake of the drug to keep the neurotransmitter at its natural levels.

Review this article on neurotransmission and cocaine, then create a flowchart to visualize the development of drug addiction:
https://tinyurl.com/l9jdghw

**Psychology in real life**

Imagine that Crockett, Fisher and Freed were leaders of three independent research teams that you previously funded in Humanborough. They have reported their findings to you.

What do you make of these findings as a policy-maker? What practical applications can be developed based on this research?

Prioritize these three research programmes based on their practical significance and suggest one concrete practical application for the one at the top of your list. For example, if you chose Crockett et al how can serotonin be used (ethically) to increase rates of prosocial behaviour in society?

Would you continue funding all three of them?
For a long time brain research has been limited to studying victims of stroke or accident and using invasive methods such as autopsy to study their brain. Comparing behavioural deviations observed in these people after the accident and abnormalities in brain structure discovered after their death led to some insights about the functions of certain brain areas. This was how Broca’s and Wernicke’s areas were discovered, for example.

Obviously, it is an advantage to be able to use non-invasive methods, allowing us to study the brain without cutting the skull open. Such methods are widely used today. They are collectively known as brain imaging techniques, or neuroimaging.

Five of the most commonly used brain imaging techniques are computerized axial tomography (CAT), positron emission tomography (PET), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and electroencephalography (EEG).

Computerized axial tomography

Computerized axial tomography (CAT) works on the principle of differential absorption of X-rays. The subject lies on a table that slides inside a cylindrical apparatus, where a moving source of X-rays scans the subject’s head. After passing
through the head the X-ray beam is picked up by a detector and analysed. Bone and hard tissue absorb X-rays better than soft tissue. As multiple X-ray beams go through the head it is possible to reveal the structural features of the brain.

The strength of this technique is that it is a quick non-invasive method of studying brain structure. It has an advantage over standard X-rays because CAT records images of hard and soft tissue as well as blood vessels simultaneously. Unlike some other techniques, CAT scans can be made for people who have implanted medical devices.

The limitation is that CAT scans involve some level of radiation exposure.

The strength of this technique is that it is a quick non-invasive method of studying brain structure. It has an advantage over standard X-rays because CAT records images of hard and soft tissue as well as blood vessels simultaneously. Unlike some other techniques, CAT scans can be made for people who have implanted medical devices.

The limitation is that CAT scans involve some level of radiation exposure.

### Magnetic resonance imaging

Magnetic resonance imaging (MRI) is different from CAT in that it does not involve X-rays. In general, MRI is often compared to CAT because it has the same purpose—to produce a high-resolution three-dimensional image of brain structure. MRI is based on the principle that some atomic nuclei—in particular those of hydrogen atoms—can emit energy when placed in an external magnetic field. When these pulses of energy are detected by the scanner, the relative distribution of hydrogen atoms in the brain can be mapped. Hydrogen atoms exist naturally in the body, but their concentration in different types of tissue is different. For example, the highest concentration of hydrogen atoms is found in water (H\(_2\)O) and fat. Analysing the pattern of emission of energy in response to magnetic fields, we can see inside the brain. After excitation by the magnetic field each tissue returns to its equilibrium state—and the time required to do so differs in different types of tissue. This information is also analysed. This is why it is required to rapidly change the parameters of the magnetic field and switch it on and off repeatedly. The result is the loud noise that is characteristic of any MRI scanner.

The advantages of MRI as compared to CAT include the following.

- It allows non-exposure to radiation and, as a consequence, less risk of radiation-induced cancer.
- MRI has better resolution. This makes it particularly useful for detecting abnormalities in soft tissue—such as the brain.

However, MRI does have disadvantages.

- People with metal in their body, for example cardiac pacemakers or shrapnel, cannot undergo the procedure because metal will attract to the magnetic field (one can only imagine what happens). Several deaths have been reported in patients with undisclosed metallic implants who underwent the procedure.
- An MRI scan can be an issue for claustrophobic people because it requires being placed in a narrow tube. Also, longer scan times are required: in some cases people have to stay inside the tube for as long as 40 minutes. Specially constructed mirror glasses are sometimes used to create the illusion of openness of the space inside the scanner.
- Lying still for a long time may be problematic for young children, especially since the procedure is new and may be frightening (partly because MRI scanners are noisy). For this reason, children having MRI scans are often sedated. Some clinics try to turn MRI scans into a fun adventure, pretending that the MRI scanner is a pirate ship, for example.
- An MRI scan is more expensive than a CAT scan. However, the costs are falling.
- Interestingly, the high resolution and sensitivity of an MRI scan is a risk in itself due to incidental findings. Sometimes the scan will pick up slight abnormalities in the brain structure that are not actually related to the symptoms being investigated. This may create anxiety and cause patients to seek unnecessary treatment.
ATL skills: Self-management

You can continue Table 2.1 by adding more rows to it. For example, you might add “Strengths”. As you read this unit, expand the table and add new details to it.

Table 2.1 Comparison of neuroimaging methods

<table>
<thead>
<tr>
<th>Does it investigate structure or processes?</th>
<th>CAT</th>
<th>MRI</th>
<th>fMRI</th>
<th>PET</th>
<th>EEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>Structure</td>
<td>Structure</td>
<td>Processes</td>
<td>Processes</td>
<td>Processes</td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>Up to 1–2 mm</td>
<td>Up to 1–2 mm</td>
<td>Up to 1–2 mm</td>
<td>4 mm</td>
<td>Very poor</td>
</tr>
<tr>
<td>Temporal resolution</td>
<td>NA</td>
<td>NA</td>
<td>1 second</td>
<td>30–40 seconds</td>
<td>Milliseconds</td>
</tr>
<tr>
<td>Major challenges</td>
<td>Radiation exposure</td>
<td>Up to 40 minutes spent without movement in a narrow noisy tube</td>
<td>Cancelling out random noise, a lot of trials required</td>
<td>Radiation exposure</td>
<td>Cancelling out random noise</td>
</tr>
</tbody>
</table>

Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) is called functional for a reason: the image obtained in the scan is dynamic. While MRI and CAT are only able to reveal the structural features of the brain, fMRI can also show the ongoing brain processes. In a typical fMRI study the subject is required to carry out some task in which periods of activity are alternated with periods of rest. The principle at work is that when a brain region is active during the performance of a task, the flow of oxygenated blood in that region increases. The response of blood to rapidly changing magnetic fields differs depending on the flow and the level of oxygenation. The signal that is analysed by the fMRI scanner to reconstruct brain activity is known as BOLD (blood-oxygen-level dependent) signal. There are other biomarkers as well, but BOLD is the most widely used. The flow of oxygenated blood directly correlates with the energy used by brain cells, and this directly corresponds to the level of activity in a specific brain region.

An fMRI scan, just like any other brain imaging technique, is characterized by spatial resolution and temporal resolution. Spatial resolution is the ability to discriminate between nearby locations: just as with the resolution of your computer screen, the lower it is, the more pixelated the picture and the less detail you can discern. Whereas resolution of your screen is measured in pixels, that of an fMRI scanner is measured in voxels. You can think of them as “volumetric pixels”—a cube of neurons. A voxel is the smallest “brain particle” that we are able to see through a scanner. Typically, the size of a voxel that an fMRI scanner is able to operate with ranges from 1 to 5 millimetres. Small voxels have less blood flow, so the signal is weaker and the required scanning time is longer. A voxel contains several million neurons and several billion synapses. This marks the limit of what can be achieved with brain imaging technology: we can only see a relatively crude picture of brain functioning.
Temporal resolution is the smallest time period in which changes in brain activity can be registered. Think about it as the rate at which snapshots of the brain are taken—“frames per second”. Currently, the temporal resolution achieved in fMRI is about 1 second. This also marks a limitation: fMRI is well suited for studying processes that last at least for several seconds (memory, face recognition, thinking about alternatives of a choice and emotional reactions) but is not suited for studying instantaneous processes such as information travelling from the retina to the visual cortex (which takes milliseconds).

With the ability to capture brain processes comes the challenge of distinguishing the ones that interest you from random noise. Brain activity associated with the task that the subject is performing has to be separated from all sorts of background activity. Potential sources of noise are head movements in the scanner resulting from the subject fidgeting, pressing buttons (as required by the experimental procedure) or even simply breathing. A clear scan requires the subject’s head to be motionless, but this is not realistic. Random thoughts and sensations also result in noise. A lot of noise can be accounted for if the number of trials is sufficient and powerful statistical techniques are used, but some sources of bias are impossible to eliminate.

These are the advantages of fMRI.
- It offers excellent spatial resolution (up to 1–2 mm).
- Unlike structural imaging techniques, it allows us to see brain processes.

There are some disadvantages.
- There is poor temporal resolution (about 1 second) when using fMRI as compared to electromagnetic techniques such as EEG (<1 millisecond).
- All the considerations that were relevant to MRI also apply to fMRI: claustrophobia, cost, lengthy procedure and inability to use it with medical implants.

▲ Figure 2.26  Functional magnetic resonance imaging

ATL skills: Thinking

Every technology, even the most modern, is prone to errors. The more sensitive the technology, the higher the probability of the so-called false positives—results that are not meaningful, but picked up due to the extreme sensitivity of the device. A typical fMRI scans across 130,000 voxels, which inflates the probability that one of them will be shown as “active” accidentally due to random chance.

Review this one-page report of a study that was intentionally designed to demonstrate this problem: Bennett et al [2009] https://tinyurl.com/yb27fq2

The sample of the study was one dead Atlantic salmon that was placed in an fMRI scanner and shown photographs of humans in various social situations. The salmon was "asked" to determine the emotions experienced by people in the photographs. The amazing thing is that comparison of the dead salmon’s brain responses to pictures showing different social situations revealed some statistically significant results.

This study won the IgNobel prize in neuroscience.
Positron emission tomography
Positron emission tomography (PET), like fMRI, uses blood flow as the indicator of brain activity. A radioactive tracer is used that binds itself to molecules naturally used in the brain, such as glucose. This radioactive tracer is administered into the subject’s blood stream. It has a short half-life period (that is, it decays quickly). The scanner then registers radio frequencies emitted by the decaying tracer. Brain areas that are more active require more blood supply, so the distribution of the tracer in the brain will depend on what regions are mostly in use at the time of the scan.

PET has a decent spatial resolution of about 4 mm throughout the brain. However, its temporal resolution is only 30–40 seconds, so quick processes are not easily detected. The biggest advantage of PET scans is their good spatial resolution but they are used less and less these days given the existence of non-invasive alternatives (fMRI) which do not require administration of a radioactive chemical.

PET is useful for detecting tumours and metastases, as well as other diffuse brain diseases, so that it becomes clear what areas are affected by the spreading disease. It is often helpful in diagnosing causes of dementias.

Another advantage of PET is that scanners can be small—so small that a small PET scanner has been constructed that can be worn by a rat on its head like a hat. The device is called RatCAP (Schulz et al 2011). The rat is conscious and fully mobile, it performs various tasks while its brain activity is being measured. This is very useful for research and potentially can lead to a lot of insights into brain functioning.

Electroencephalography
Electroencephalography (EEG) measures electric potentials generated by neural circuits. Neurons communicate with each other by sending electrical impulses along their axons. An impulse fired in an individual neuron is “invisible” to any device outside of the skull because the impulse is too tiny. However, when large groups of neurons fire synchronously, electric potentials generated by these impulses become detectable at the head surface. Electrodes are attached to the scalp in predetermined points and pick up the changes in the electric potential of the scalp areas. This information is used to generate an electroencephalogram.

EEG has a perfect temporal resolution. It is capable of detecting changes in brain activity within milliseconds. In this sense it outperforms other techniques such as fMRI. However, its spatial resolution is a weakness; in practice EEG is not used to establish the origin location of the electrical signal. EEG is good for measuring brain activity “on the whole”.

As it makes visible changes in the overall patterns of brain activity (sometimes referred to as brain waves), EEG is commonly used to diagnose such conditions as epilepsy and sleep disorders. These are the advantages of EEG.

- It is a low-cost technique.
- Unlike PET and fMRI, EEG measures neuronal activity directly.
- EEG can be offered as a mobile service because the apparatus can be manually transported. For comparison, the weight of an fMRI scanner is about 1 ton.
- EEG is silent, which is an advantage because responses to auditory stimuli can be studied. This is difficult with noisy fMRI scanners.
- EEG is completely non-invasive in comparison to most other neuroimaging techniques.

Using EEG also has disadvantages.

- EEG offers extremely low spatial resolution, so it only gives a very crude picture in terms of localization.
- EEG is good for measuring electric activity in the cortex, but not so good for detecting activity in subcortical areas. The farther away from the surface of the scalp, the weaker the signal.
It takes considerable experience to interpret an encephalogram correctly because a number of artifacts contribute to noise in the data, and the signal–noise ratio is quite low. Some potential sources of noise are: heartbeat, muscle movements, eye movements and eye blinks, and poor grounding of the apparatus connection.

The use of neuroimaging in research studies

You have encountered some studies using neuroimaging techniques already and there will be many more throughout this book.

These are the examples we have discussed so far.

- Draganski et al (2004) used MRI to determine changes in brain structure in response to learning a simple juggling routine for three months.
- Draganski et al (2006) used MRI to see if brain structure changed as a result of revising for a medical exam.
- Maguire et al (2000) compared MRI scans between London taxi drivers and controls to see if hippocampus played a role in spatial memory.
- Freed et al (2001) used PET scans to study dopamine-producing cells in the brains of Parkinson’s disease patients.
- Fisher, Aron and Brown (2005) used fMRI to study brain processes in response to looking at the picture of a loved person.

Brain imaging technology provided a useful and fruitful alternative to invasive methods such as post-mortem studies or neural stimulation by implanted electrodes.

Knowing relative strengths and limitations of specific brain scanning technologies, as well as major considerations involved in choosing between them, will automatically enable you to add evaluation points to research studies that used technology to scan brain structure or processes. The choice is based on a variety of factors including available technology, cost, ethics and aims of the study.

ATL skills: Self-management

As you continue reading this book, make a rule of noting research studies that used brain imaging technology. Analyse how and why the decision was made to use a particular scanning technique.

Psychology in real life

Look back at the research projects that you have proposed so far for the people of Humanborough and pick your top three. What techniques, if any, would you use in these projects? Would they be invasive or non-invasive? Would they investigate structure or processes? What factors would you need to consider to make the final choice?
What you will learn in this section

- The function of hormones
  - Unlike neurotransmitters, hormones travel with blood, regulate long-term ongoing processes and allow for lesser voluntary control. The nervous system and the endocrine system are interdependent.
  - Hormones do not influence behaviour directly. Instead, they change the probability that a certain behaviour will occur in response to a certain environmental stimulus.

- Oxytocin
  - It is produced in the hypothalamus and released into the blood by the pituitary gland.
  - It plays a role in sexual reproduction, childbirth and social bonding.
  - Oxytocin is also linked to such behaviours as interpersonal trust, fidelity or even intergroup conflict.

- The role of oxytocin in interpersonal trust
  - Experimental task: the game “Investor”
  - Alternative explanations

- Oxytocin reduces risk aversion in general (eliminated because results were not replicated with computers instead of human partners).
- Oxytocin specifically increases trust to other humans (accepted).

- The role of oxytocin in fidelity
  - Scheele et al (2012): by selectively influencing men in a relationship to keep greater distance from women strangers, oxytocin may promote fidelity.
  - Experimental tasks: stop-distance paradigm and approach/avoidance task.
  - Oxytocin selectively inhibits approach to certain stimuli—attractive women—in men who are in a stable relationship, but not in single men.
  - This may play a role in maintaining the stability of monogamous bonds.

- The role of oxytocin in intergroup conflict
  - De Dreu et al (2012): oxytocin promotes defence-motivated non-cooperation between groups.
  - Experimental task: a modified version of the “Prisoner’s dilemma”.
  - Oxytocin-induced non-cooperation was motivated by the desire to protect vulnerable group members (and not so much by the desire to protect oneself).
The function of hormones

In their function hormones are similar to neurotransmitters: essentially both are chemical messengers. However, neural communication (neurotransmission) and hormonal communication differ in a number of ways.

- Hormones are released into the bloodstream and travel with blood to reach their destination. Conversely, neurotransmission is communication along nervous cells. The implication of this is that hormones can reach places that the nervous system does not cover, because the network of blood vessels is more elaborate.
- The nervous system regulates relatively rapid processes (movement, emotion, decisions, and so on), whereas hormones can regulate long-term ongoing processes such as growth, metabolism, digestion or reproduction.
- Generally speaking, the degree of voluntary control over neural regulation is higher than over hormonal regulation. For example, it is possible for you to control your emotions to a certain extent, whereas the degree of control you have over your growth is negligible.
- However, it should be noted that the nervous system and the endocrine system are interdependent. These two systems interact and to some extent can influence each other. Also, some chemicals may be both hormones and neurotransmitters, for example adrenaline.

Hormones are released by endocrine glands: adrenal glands, hypothalamus, pineal gland, pituitary gland, thyroid, parathyroid, thymus, pancreas, testes and ovaries. Together, these form the endocrine system.

Hormones can only influence cells that have receptors for this particular hormone. Such cells are called target cells. When a hormone binds to a receptor it launches a sequence of changes some of which are genomic: gene activation or gene suppression. Essentially, what this means is that hormones do not influence behaviour directly. Instead, they change the probability that a certain behaviour will occur in response to a certain environmental stimulus. This is like buying ice-cream on a hot day: hot weather itself does
not cause you to buy ice-cream, but it certainly increases the probability that you will.

There is a variety of hormones produced in the body and they all have different functions. The most well-known hormones include adrenaline, noradrenaline, cortisol, oxytocin, insulin, testosterone and oestrogen. We will look at evidence that links hormones to behaviour using one specific example—**oxytocin**.

**Oxytocin**

Oxytocin is produced in the hypothalamus and released into the blood by the pituitary gland. It plays a role in sexual reproduction, childbirth and social bonding. It has been referred to as “the love hormone”, “the bonding hormone” and “the cuddle chemical”. For example, oxytocin is released from stimulation of nipples during breastfeeding and this helps to establish a stronger bond between the mother and the child. It is also released with every kiss or hug.

**Romero et al (2014)** demonstrated that oxytocin promotes social bonds in mammals in non-reproductive contexts. In their study 16 dogs were sprayed intra-nasally either with oxytocin or a placebo (in a repeated-measures, double-blind counterbalanced design). They were placed with their owner and another dog in the same room and their behaviour was recorded by four cameras during one hour. The room was empty except for a chair on which the dog owner sat. The owner was instructed to move the chair in pre-designated positions every 10 minutes, but otherwise sit quietly and not actively interact with the dog.

Later the recordings were analysed using a checklist of dog behaviours. Results showed that dogs sprayed with oxytocin showed higher affiliation towards their owner. Affiliation was operationalized as sniffing, licking, gentle touching with the nose or paw, play bouts and body contact. They also spent significantly more time in close proximity to the owner. Similar results were observed for the dog partner (the other dog present in the room): affiliation and approach behaviours were more frequent in the oxytocin condition. Furthermore, the effect of oxytocin was found to be bi-directional: subsequent blood tests showed that the more often the dog interacted with the owner and the dog partner, the higher the levels of endogenous oxytocin it had. So oxytocin “triggers” social interaction, and social interaction affects the release of more oxytocin. The researchers concluded that oxytocin performs the function of maintaining close social bonds in mammals. In the “Acknowledgement” section of their article they did not forget to thank the dogs for participation.

The role of oxytocin in instinctive behavioural sequences, such as attachment to a baby in response to nipple stimulation, is well understood and has been observed in animals as well as in humans. It would be more interesting to find out if there is a link between the hormone and seemingly unrelated behaviours such as interpersonal trust, fidelity or even intergroup conflict.
The role of oxytocin in interpersonal trust

Kosfeld et al (2005) claimed that oxytocin increases trust in humans. Participants were 128 healthy male students (mean age 22 years). Subjects were randomly allocated into either the oxytocin group or placebo group. Substances were administered via an intra-nasal spray.

For the purposes of the experiment the researchers designed a trust game with real monetary stakes. In this game subjects were paired anonymously and played the role of either an investor or a trustee. In game theory this game is known as “Investor”. Each round of the game (each with a new partner) consists of three steps.

As step 1, the experimenter gives both the investor and the trustee an endowment of 12 monetary units. In step 2, the investor needs to decide how much of that to send to the trustee (there are four options: 0, 4, 8, 12). The experimenter triples whatever is sent to the trustee: if the investor sends 4 units, the trustee receives 12, if the investor sends 12 units the trustee receives 36, and so on. Remember that the trustees have their own endowment, which is added to their total. For example, if the investor sends 8, the trustee receives 24 and so has 24 + 12 = 36 monetary units.

In step 3, the trustee decides how much of the now available money to send back to the investor. The idea is that if I (the investor) completely trust you (the trustee), I will send you 12 monetary units, it will turn into 36, and I trust that you will send me back at least 18 and maybe more. Can I trust you, though? We only interact once during this experiment, so you have a temptation of keeping the whole sum with you. In order to trust you, I need to overcome aversion towards this risk.

Participants played the game four times in the same role, each time paired randomly with a new partner. At the end of the experiment the total earned monetary units were exchanged for real money.

Results of the experiment showed that the level of trust in those participants who received a dose of oxytocin was higher than in the control group. The median transfer of investors was 10 in the oxytocin group and 8 in the control group. Forty-five percent of subjects in the oxytocin group showed the maximum trust level (12 monetary units), whereas only 21% in the placebo group showed the maximum trust level.

The authors suggested two alternative explanations for this finding.

- Oxytocin reduces risk aversion in general.
- Oxytocin increases people’s trust in other humans.

In order to clarify, they designed a follow-up study in which an independent group of subjects played a similar trust game, but this time against a random mechanism (software). The algorithm in the software was modelled after decisions of real people (trustees) in the previous experiments, so
the investors faced exactly the same risks as in the “human” experiment, only this time they knew they were playing against machines. No difference was observed in this experiment between the oxytocin and placebo groups. The median transfer was eight monetary units in both conditions. The researchers concluded that oxytocin specifically affects trust in interpersonal interactions.

**ATL skills: Thinking and research**

Write the study in the “A – M – P – R – C – E” format. Be concise. When you evaluate, consider methodological aspects [sampling, generalizability, credibility, bias] and ethical aspects.

Note that the researchers suggested alternative explanations for their findings and then conducted follow-up studies to eliminate one of the alternatives. This shows how a whole research programme rather than a separate study is normally needed to test a hypothesis properly. Alternative explanations and their elimination is also a great exercise in critical thinking.

### The role of oxytocin in fidelity

Scheele *et al* (2012) showed that oxytocin modulates social distance between men and women. The researchers studied 86 heterosexual men. Some of them were single and others were in a stable monogamous relationship. Using a double-blind independent measures design, a researcher administered either oxytocin or a placebo intranasally.

Subjects participated in two independent tasks. In the first task—“stop-distance paradigm”—subjects were positioned at one end of the room with their toes on the mark on the floor, while an attractive female experimenter was positioned on the other side of the room. The subject was then required to move slowly towards the female experimenter and stop at a distance that made him feel slightly uncomfortable (too close). Care was taken to assure that the experimenter maintained the same appearance over all the trials.

Results of the experiment in the first task showed that oxytocin stimulated men in a monogamous relationship, but not single ones, to keep a greater distance between themselves and an attractive woman. It was concluded that oxytocin caused men in a relationship to “stay away from” an attractive woman who was not his partner.

Results of the experiment in the second task showed that the only group of pictures affected by oxytocin and relationship status was the positive social group (pictures of attractive women). Specifically, participants who received oxytocin had slower reaction time (that is, pulled the joystick...
more reluctantly) in response to these pictures, but only if they were in a relationship. It was concluded that oxytocin selectively inhibits approach to certain stimuli—attractive women—in men who are in a stable relationship, but not in single men.

From the results of these trials it is seen that by selectively influencing men in a relationship to keep greater distance from attractive women they do not know, oxytocin may promote fidelity.

**The role of oxytocin in intergroup conflict**

There are less obvious effects of oxytocin, too. For example, do you think it can play a role in prejudice, discrimination or maybe even conflict? It turns out it can, but these come as side effects of increased bonding with your own group. The following two studies will look at these seemingly surprising negative effects of oxytocin.

**De Dreu et al (2012)** looked at the role of oxytocin in intergroup conflict—more specifically, **defence-motivated non-cooperation**.

This was a double-blind experiment using the independent measures design. The sample consisted of 102 males and they self-administered either a dose of oxytocin or placebo through nasal spray. Participants were randomly assigned to three-person groups and told that they would need to compete against another group of three people.

Following this, each participant was paired with a member of another three-person group and played a modified version of the “Prisoner’s dilemma”.

---

**TOK**

“Prisoner’s dilemma” is probably the most commonly mentioned problem of game theory. Game theory is an interesting interdisciplinary field that combines mathematics, decision theory, economics and other areas. It is a theory that models the behaviour of two or more rational agents whose actions and outcomes depend on each other. Things get even more interesting when you compare the rational predictions of the mathematical theory (game theory) to real-life behaviour of participants in these situations (behavioural game theory).

The standard “Prisoner’s dilemma” is as follows. Two members of a criminal gang are arrested. Each prisoner is kept in an isolated chamber and they have no means of communicating with each other. They are interrogated independently and each of them has to choose one of two actions. They are also given the same deal, which is presented in the form of a matrix of outcomes, as shown in Figure 2.32.

![Figure 2.32 Options for players of the “Prisoner’s dilemma”](image)

If A and B each choose to betray the partner and testify against that person, they are both given two years in prison. If A and B both remain silent, they only get one year each. If A chooses to testify against B, but B keeps silent, A will be free and B will get three years in prison (and vice versa).
Figure 2.33 shows an example of the game played by participants in the study.

![Figure 2.33](image)

Just as in the standard “Prisoner’s dilemma”, the two participants are asked to choose, simultaneously and independently, whether they want to cooperate or compete. Although the decision is made independently, both players see the pay-off matrix. They can see the following possible outcomes.

**A** If I (the participant) cooperate and he (the out-group player) also cooperates, I win 7 points, the members of my group win 7, he wins 7 and the members of his group win 7. This is an attractive option: everyone wins quite a lot.

**B** If I compete and he competes, we only win 6—both I and my group members and he and his group members. This is less attractive than outcome A for all of us.

**C** If I cooperate but he competes, he and the members of his group win 8—the largest possible pay-off. At the same time, I only win 5, and the members of my group only win 1. This means that I am vulnerable to my opponent’s non-cooperation, but my group members are even more vulnerable.

**D** If he cooperates but I compete, outcome C is reversed.

What decision would you make if you were playing this game? Outcome A (cooperate-cooperate) seems to be the most rational choice, but the problem is that there always exists a temptation to switch from cooperating to competing because this will increase the pay-off from 7 to 8. You know this, and you can probably decide that 7 is almost as good as 8 and choose to cooperate. However, you also know that your opponent has the same temptation. Will he be reasonable and sacrifice 1 unit of profit for the common good? You hope so—but what if he doesn’t, and goes for outcome C? In this case you will suffer (not too much), but your group will suffer a lot: its pay-off will be 7 times smaller. The main question is: how likely is it that, in order to prevent outcome C from happening, you will strike first and choose to compete rather than cooperate? That’s what we call defence-motivated non-cooperation.

**TOK**

Sometimes a war begins when one country strikes another country pre-emptively, because it is afraid to become the victim if it doesn’t strike first. To what extent do you find this reasoning rational?

In general, can negative social events be the result of positive, rational reasoning?

Participants were asked to make a choice between cooperation and non-cooperation five times. The numeric rewards were varied on these five trials. The pay-offs from the game...
were converted to real money and given to the participants.

Results of the study showed that a player's defence-motivated non-cooperation:

- is more likely if vulnerability of the player's group is high (for example, suppose we change 1 to 0 in the pay-off matrix)
- does not depend so much on the player's own vulnerability
- is more likely in the oxytocin condition.

The researchers concluded that oxytocin-induced non-cooperation is motivated by the desire to protect vulnerable group members (and not so much by the desire to protect oneself). This reinforces the role of oxytocin in creating bonds with the members of your in-group, but also shows how oxytocin has a reverse, negative side—defensiveness and non-cooperation with others.

The role of oxytocin in human ethnocentrism

De Dreu et al (2011) found that oxytocin promotes human ethnocentrism, a type of intergroup bias where one's own ethnic group is perceived as more important than or superior to others. When exaggerated, ethnocentrism may lead to xenophobia. (This is not something that oxytocin—"the love hormone", "the bonding hormone", "the cuddle chemical"—has usually been associated with.) De Dreu et al (2011) conducted a series of experiments all of which used double-blind placebo-control independent measures designs. Participants in the studies were indigenous Dutch males. They self-administered either oxytocin or a placebo intra-nasally. Experiments involved exposing subjects to images of people belonging either to their in-group (Dutch males) or out-group (immigrants from the Middle East and German citizens).

Experiments used “moral-choice dilemma” tasks, such as the famous trolley problem we have already discussed (see “Neurotransmitters and behaviour”). Participants in the oxytocin and placebo groups were given a series of moral-choice dilemmas where decision had to be made as to whether one person should be killed in order to save five other people (for example, hitting the switch that will divert the trolley to another track, killing one individual and saving five). In some of these tasks the target person (the one who had to be killed) was a member of participants’ in-group, and in other tasks a member of their out-group. This was achieved by manipulating the name of the target person: either a typically Dutch name (Dirk or Peter, for example) or a typically Arab name (such as Ahmed or Youssef) or a typically German name (such as Markus or Helmut). The other five individuals were unnamed and so their identity or ethnic background was not indicated. The question was, in a task like this, would a Dutch person prefer to sacrifice a non-Dutch person, and how does this depend on oxytocin?

Results showed that under oxytocin males were more likely to sacrifice an out-group target than an in-group target, while under the placebo there was no significant difference. There are two alternative explanations for this finding:

- Oxytocin promotes in-group favouritism.
- Oxytocin promotes out-group derogation.

Further analysis of the data revealed that, compared with males in the placebo (control) condition, males under oxytocin were less likely to sacrifice a member of their in-group, but were...
not more likely to sacrifice a member of the out-group (as shown in Table 2.2). Based on this, the first explanation has to be the preferred one.

|                                | Oxytocin | Control     |
|---------------------------------------------------------------|
| Frequency of sacrificing out-group targets                   | Oxytocin = control |
| Frequency of sacrificing in-group targets                    | Oxytocin < control |
| Are out-group targets sacrificed more frequently than in-group targets? | Yes | No |

Table 2.2

Psychology in real life

Now that you know about various effects that oxytocin has on human behaviour, how can you use this knowledge to inform policy-making in Humanborough?

Think about one concrete project that you would like to implement. List its potential risks and benefits.

Oxytocin creates intergroup bias by increasing in-group favouritism. It does not promote out-group derogation. This is good news... or is it?
The Athena Institute, founded by Dr Winnifred Cutler, markets two brands of Athena pheromone: one for men and one for women. The one for women is a fragrance additive. You are supposed to add it to your regular alcohol-based perfume and “dab the mixture above upper lip, behind ears and elsewhere at least every other day”. It is claimed that the product enhances your “sex appeal” and promotes your “sexual attractiveness”. This is one of the customers’ feedback published on the website:

Suzie: “Men are absolutely freaking out. I work as a bartender and men are flocking around me. I have told my girlfriends it must be the LOVE POTION. My girlfriends in the bar say they can almost feel the energy of the men’s attraction to me”.

Source: https://www.athenainstitute.com/1013.html

How do you evaluate the credibility of such products?

What evidence should be used to back up the claims made about such products?

Would you buy the Athena pheromone? (Note: at the time of this book’s publication, the product costs around US$100 for a vial.)

● Pheromones
  - Chemical communication among members of the same species
  - Although the role of pheromones in animal behaviour is not doubted, the effects of pheromones on human behaviour have been the subject of much debate

● Localization of processing pheromonal information in the brain
  - The vomeronasal organ (VNO) and accessory olfactory bulb
  - Human fetuses have the accessory olfactory bulb, but it regresses and disappears after birth. The VNO in people appears to be non-functional: there is no connection to the central nervous system.

● Search for a human sex pheromone: laboratory experiments
  - Lundstrom and Olsson (2005): androstadienone increases women’s mood in the presence of a male experimenter and has no effect when the experimenter is female
  - If a chemical does not perform gender-signalling function, the chances are that it will not produce any other gender-related effects and so it is not a sex pheromone. Hare et al (2017): androstadienone and astrotetraenol do not act as signals of gender or of attractiveness.
  - Conclusion: some chemicals influence women’s mood, but these same chemicals do not signal gender or affect mate perception, so we cannot classify them as a sex pheromone

● Search for a human sex pheromone: field experiments
  - Attempt to solve two problems: look at a broader range of behaviours and overcome artificiality of experimental tasks
  - Cutler, Friedmann and McCoy (1998): (male participants) a synthetic human pheromone did not just increase libido but increased the attractiveness of men to women
McCoy and Pitino (2002): (female participants) a synthesized pheromone increased sexual attraction of women to men

- Criticism of research into human pheromones
  - Limitations of a typical experiment (Verhaeghe, Gheysen, Enzlin 2013)
    - Population validity—most studies used self-selected samples
    - Participant bias (demand characteristics)—there are hints that may lead participants to guess the true aims of the study
    - Ecological validity—a concentration solution of the pheromone much higher than what is found in natural sweat is used
    - Internal validity—other smells act as confounding variables
    - Experimenter bias—the gender, the looks and the behaviour of the experimenter might be significant

- Construct validity—even if the influence of a chemical substance or a scent on the behaviour of human subjects is demonstrated, this does not mean that the chemical substance is a pheromone
- Ethics
  - Further limitations
    - Publication bias—researchers who conduct human pheromone studies are often commercially interested in the results
    - Replicability—this has been an issue in the research with human pheromones; research remains inconclusive

This section also links to:
- hormones and behaviour
- localization of function
- Unit 1 “Research methodology”
- Unit 7 “Psychology of human relationships”: formation of personal relationships, attraction

Pheromones

The word pheromone is derived from the Greek phero (I carry) and hormon (stimulating), so pheromones are chemicals that “carry stimulation”. The term appeared when scientists observed termites and noticed that a chemical substance released by one termite affects the behaviour of other termites. It was suggested as an umbrella term for various forms of chemical communication among members of the same species. For some social insects (for example, termites) chemical communication is the main form of communication. Pheromones have also been shown to play a role in the behaviour of mammals, though, mainly in mating behaviour. For example, if he cannot sense pheromones signalling fertility, a male rhesus monkey will ignore the romantic attention of a female (Herz 2009).

Such findings are inspiring and of course they raise the question: do pheromones play a role in human attraction? The effects of pheromones on human behaviour have been the subject of much debate.

Localization of processing pheromonal information in the brain

Although many pheromones have a smell, pheromonal information in the brains of animals is not processed in the same brain regions as ordinary smells. The region of the brain responsible for processing smell is called the main olfactory bulb. However, pheromones are processed differently from regular smells. Mammals have a separate structure called the vomeronasal organ (VNO) which is located in the anterior nasal cavity. Nerves from the VNO in animal brains connect to a special region called the accessory olfactory bulb. This region is adjacent to, but separate from, the main olfactory bulb (Herz 2009).

A major difficulty with extrapolating animal research to human research to human behaviour is linked to the fact
that humans do not have either the VNO or the accessory olfactory bulb. On this point, though, we need to be accurate: human fetuses do have the accessory olfactory bulb, but it regresses and disappears after birth. Some people do have the VNO while some don’t. Even in those who have the VNO, it appears to be non-functional: there is no connection to the central nervous system. If pheromonal information is indeed processed in the human brain, it must be processed somewhere else.

Search for a human sex pheromone: laboratory experiments

Evidence for the influence of pheromones on human behaviour has been inconclusive. There are many intriguing findings, but there is always an alternative explanation or lack of clarity that prevents us from claiming pheromonal effects with certainty.

For example, Lundstrom and Olsson (2005) studied the effects of androstadienone—a derivative of testosterone and one of the chemical components of sweat. The study involved studying a woman’s mood after being exposed to: either androstadienone or control solution; and in the presence of either a male or a female experimenter. Results showed that androstadienone increased women’s mood in the presence of a male experimenter and had no effect when the experimenter was female. It is tempting to conclude that androstadiene as a pheromone intensifies women’s reactions to men, but the study has important limitations.

- The concentration of androstadiene used in the study was much higher than the normal amount found in male sweat—which is a common feature of pheromone studies.
- It is hard to separate the effects of the pheromone from the effects of the experimenter. For example, what if the male experimenter was simply particularly handsome? Additional research has to be done in this area.

ATL skills: Thinking and research

As a researcher, how would you overcome the two problems listed above? Can you decrease the pheromone concentration? Can you think of a way to quantify and control attractiveness of the male experimenter?

It is widely recognized that a basic function of a sex pheromone, whatever other additional functions it performs, is to signal gender. If a chemical does not perform gender-signalling function, the chances are that it will not produce any other gender-related effects and so it is not a sex pheromone. Therefore, to prove that androstadienone or any other chemical is a sex pheromone one needs to show that it signals gender.

Following this logic, Hare et al (2017) investigated whether androstadienone (AND) and estratetraenol (EST)—the best-known candidates for human sex pheromones—signal gender and affect mate perception. The experiment used a repeated measures design. Heterosexual participants completed two computer-based tasks twice on two consecutive days. While completing the task, on one of the days they were exposed to the putative pheromone (AND or EST) masked with clove oil, and on the other days they were exposed to a control scent (clove oil only). Substances were administered by a cotton ball taped under the nose throughout the task. The design was counterbalanced (some participants had the pheromone on the first day and the control substance on the second day, some vice versa). The first computer-based task involved showing the participants five “gender-neutral facial morphs”, and participants had to indicate the gender (male or female). In the second task participants were shown opposite-sex photographs and asked to rate them for attractiveness on a scale from 1 to 10. The study was double-blinded. There were two experimenters—a male and a female—and they alternated for different sessions.
Results of the first task revealed no difference in gender assigned to the morphed faces in the pheromone versus control condition. Similarly, results of the second task revealed no difference in the average attractiveness ratings of opposite-sex photos. The authors concluded that AND and EST do not act as signals of gender or of attractiveness, which means that they do not qualify as sex pheromones. Incidentally, the gender of the experimenter had no effect on the results.

So far, we have two contradictory studies—but they are not as contradictory as it seems. If we assume that the methodological quality of both studies was high and the findings accurately reflect reality, we may admit that some chemicals such as androstadienone influence women’s mood. However, these chemicals do not signal gender and they do not affect mate perception, so we cannot classify them as a sex pheromone. Of course there is always an alternative explanation, which is that one of the studies was biased in some way.

ATL skills: Communication

If you were to assume that the contradictory findings in these two studies were in fact due to the bias present in one of them, which one would you pick as the more biased study? Write a short statement explaining your point of view and giving reasons.

Maybe the problem is that both the studies concentrated on a very specific behaviour and in order to clarify the functions of a putative pheromone we need to look at a wider range of behaviours. Or maybe the problem is in the artificiality of the experimental tasks.

The following two research studies attempted to address both the problems—they were conducted in real-life settings and accounted for a wider range of behaviours.

ATL skills: Social

Before you go on, brainstorm some ideas for field experiments with putative human pheromones. If you had a substance that you believed was a candidate for the human sex pheromone and you had to conduct a field experiment to test this substance, what would you do?

Search for a human sex pheromone: field experiments

Cutler, Friedmann and McCoy (1998) investigated whether synthesized male pheromones increase sociosexual behaviour of men. Participants (38 men) were recruited through local press releases that invited volunteers to participate in an experiment with the aim to “test whether a male pheromone added to aftershave lotion would increase the romance in their lives” (Cutler, Friedmann, McCoy 1998: 4). There were a number of selection criteria: male, heterosexual, 25–42 years old, in good health, not taking any medication, with regular appearance (“neither unusually handsome nor unattractive”), shaving regularly and having adequate social skills with women. To ensure that participants fitted the selection criteria they were screened with personality questionnaires and anyone with personality traits too strongly deviating from the average was excluded.
Participants were randomly divided into two groups (in a double-blind manner). Each participant brought his aftershave lotion (which was examined by the researchers) and was asked to use it after every shave and at least three times a week throughout the study period. Participants were also given a behavioural calendar which they had to fill out daily indicating the incidence of six behaviours on that day. The behaviours were:

- petting, affection and/or kissing
- sleeping next to a romantic partner
- sexual intercourse
- informal dates (that is, dates not arranged before that day)
- formal dates (that is, dates that were pre-arranged)
- masturbation.

After a baseline period of two weeks, subjects returned to the laboratory and the technician (who was blinded to the conditions) added either ethanol or pheromone with ethanol to their aftershave lotion. The pheromone was a synthesized version of a pheromone naturally secreted by men. Participants went on to use their aftershave lotion for a six-week trial period (so the total study time was eight weeks).

Results showed that there were significantly more men in the pheromone group (as compared to the placebo group) who had an increase over the baseline in the first four behaviours (petting, affection and/or kissing; sleeping next to a partner; sexual intercourse; and informal dates). For example, 47% of men in the pheromone condition reported an increase in the frequency of sexual intercourse, as compared to 9.5% in the placebo group. Differences were not observed for the last two behaviours (formal dates and masturbation). Why? The researchers concluded that applying the synthetic pheromone resulted in an increase of sociosexual behaviours “in which the willingness of a female partner plays the major role” (Cutler, Friedman, McCoy 1998: 10). In contrast, behaviours like masturbation did not increase. The researchers took it as evidence that the synthetic human pheromone did not just increase libido but actually increased the attractiveness of men to women.

McCoy and Pitino (2002) conducted a similar study with female subjects. Participants were 36 regularly menstruating women (mean age 28). Either the synthesized pheromone or a placebo was added to their perfume. Seven sociosexual behaviours were recorded weekly across three menstrual cycles—the same six as Cutler, Friedman and McCoy (1998) used plus an additional category of “male approaches”. Similar to the previous study, a significant increase over the baseline was found in the pheromone group (as compared to the control group) in such behaviours as sexual intercourse; sleeping next to a partner; formal dates; and petting, affection and/or kissing. There was no increase in the other three behaviours (male approaches, informal dates and masturbation). The authors concluded that the synthesized pheromone increased sexual attraction of women to men.

These findings seem promising. However, the discovery of human pheromones has not been widely recognized by the scientific community.
The reason is the existence of multiple counter-arguments and important limitations in all typical research studies in this area.

**Criticism of research into human pheromones**

A typical experiment with putative human pheromones suffers from a number of methodological limitations (Verhaeghe, Gheysen, Enzlin 2013).

- **Population validity.** The fact that most of the studies used self-selected samples (that is, volunteers who respond to posters or advertisements). This means that the majority of studies are performed with young, relatively educated participants.

- **Participant bias (demand characteristics).** In most studies, researchers try not to disclose the true nature of the study to participants. Mild deception is used and subjects are told that the study looks at effects of “odours”. However, there are hints that may lead participants to guess the true aims of the study.
  - Many volunteers participate in more than one psychological experiment and they may know that researchers use deception when it comes to revealing the aims of the study.
  - Participants are aware of the exclusion criteria (for example, women using contraceptive pills are not included in the sample).
  - Study surveys or interviews include questions about participants’ sexual orientation.

- **Ecological validity.** Studies typically use a concentration solution of the pheromone much higher than what is found in natural sweat. As a result, some participants can identify the smell and report that the applied solution smells like “sweat”, “urine” and “clothes”. This artificially high concentration can distort participants’ behaviour in ways that do not occur naturally. Researchers take effort to mask the smell by adding a masking agent both to the pheromone and the control solution. This may partially solve the problem of demand characteristics, but not that of ecological validity.

- **Internal validity.** Other smells act as confounding variables, so it is important to control subjects’ odourlessness, which is difficult.

- **Experimenter bias.** Since the study of pheromones focuses on participants’ responses to other people, there are important sources of bias that are more crucial in pheromone research than anywhere else: the gender, the looks and the behaviour of the experimenter or the research assistant conducting the study. This is difficult to control or keep constant in all the groups.

- **Construct validity.** Even if the influence of a chemical substance or a scent on the behaviour of human subjects is demonstrated, this does not mean that the chemical substance is a pheromone. There are many smells and substances (such as those resulting from industrial pollution or naturally found in the environment) that can have an effect on human behaviour. To be a pheromone, the substance must perform the function of communication between two individuals.

- **Ethics.** There may be some ethical issues involved. For example, in one study women were required to wipe pads containing armpit sweat obtained from donors under their noses each day for three months.

**ATL skills: Thinking and research**

Review the studies discussed in this section and apply the seven listed limitations to them. Is it likely that these limitations were inherent in the studies? Which of the studies were most vulnerable to which limitations?

Apart from the methodological quality of a typical study, researchers who conduct human pheromone studies are often commercially interested in the results. So it is likely that publication bias will occur—with researchers publishing only supporting evidence and failing to publish “unsuccessful” research.
Dr Winnifred Cutler (the author of a study discussed earlier) is the founder of a company that produces and sells a synthesized human pheromone. McCoy and Pitino are her colleagues, and their study used a pheromone produced by the same company. While this in itself does not imply that publication bias has occurred, more independent researchers without any commercial interest in the findings are needed to increase credibility.

The most straightforward way to establish credibility of an experiment is to replicate it. Replicability has been an issue in the research with human pheromones. Despite some promising findings, experiments discussed above are countered by other studies that fail to show an effect of pheromones on human behaviour. So, the effect is elusive and the research inconclusive (a statement that can apply to many studies in psychology).

See video
In his TED talk “The smelly mystery of the human pheromone” Tristram Wyatt explains the fundamental flaws in current pheromone research.
Watch the talk and write down:
- arguments that are new
- arguments that have been discussed in this section.
https://www.ted.com/talks/tristram_wyatt_the_smelly_mystery_of_the_human_pheromone

Psychology in real life
Now that you are familiar with more arguments in the quest for human sex pheromones, would you reconsider your opinions about Dr Cutler’s commercial products (and hundreds of similar alternatives that exist on the market)?
You might want to visit the website again and look at it through the lens of ideas discussed in this section.
Here is a website of another company marketing pheromones, for comparison: http://pheromones.com/
Inquiry questions

To what extent is our behaviour determined by genetic inheritance?
Is intelligence genetically pre-determined?
Can genetic and environmental influence interact?

In what ways can genetic inheritance be modulated by environmental influences?
How can we estimate heritability of a trait or behaviour?

What you will learn in this section

- Genotype and phenotype
  - DNA, chromosomes, genes, base pairs, alleles: sentence, lines, words, letters, spelling
  - gene—a unit of heredity
- Nature-nurture debate
  - The debate needs to be reformulated
    - What are the relative contributions of biological and environmental factors to a specific trait or behaviour?
    - How do biological and environmental influences interact?
- Methods of research
  - Methods based on the principle of genetic similarity
    - twin studies
    - family studies
    - adoption studies
  - Molecular genetics
- Genetic heritability: the Falconer model
  - The Falconer model assumes that phenotype comprises three types of influences: genetics, shared environment and individual environment. \( I = A + C + E \)
  - Genetic heritability (A) can be estimated by directly measuring the similarity (correlation) between monozygotic twins (rMZ) and dizygotic twins (rDZ)
- The influence of genetics on the environment: niche-picking
  - Genetic predisposition causes people to select certain environments which, in turn, start to affect their behaviour
  - This may explain why heritability coefficients change during life, typically becoming larger
- Heritability of intelligence: twin studies
  - Bouchard and McGue (1981): a review of 111 studies on IQ correlations between relatives. Results show that intelligence is to a large extent genetically inherited.
  - One needs to keep in mind the typical limitations inherent in any twin study (Joseph 2015)
    - Many twin pairs were not separated immediately after birth
    - Many separated twins grew up in similar environments as regards culture and socioeconomic status (SES)
    - Twins share a common prenatal environment
Findings might not be generalizable to a wider population.

The similar physical features might elicit similar responses from the environment.

Tests of intelligence have certain issues with validity and reliability.

Heritability of intelligence: adoption studies.

Most of the existing studies support the idea that IQ is increased by adoption into more prosperous families. At the same time, the same studies demonstrate that adopted child–biological parent correlations are always higher than adopted child–adoptive parent correlations.

Kendler et al (2015): study of sibling pairs in which one of the siblings was home-reared and the other one was adopted away.

Cognitive ability is environmentally malleable: there was a 5-point IQ increase on average by age 18.

On the other hand, results also suggest heritability of intelligence: there was a correlation between the cognitive ability of adopted children and the educational level of biological parents.

Explanation: additive influence of environment and genetics.

Scarr and Weinberg (1983): The Transracial Adoption Study.

Additive influence of environment and genetics (in line with previous studies).

Average intelligence increases: black children placed in white families increased their IQ scores substantially as compared to black children reared in their own homes.

Correlation with biological parents is higher: biological parent–adopted child correlations were higher than adoptive parent–adopted child correlations (0.43 versus 0.29).

Contradictory finding.

Young siblings were very similar to each other, whether they were genetically related or not (correlations 0.42–0.44). This finding can only be explained by the predominant influence of the rearing environment on the development of IQ.

This contradiction is reconciled in Scarr and Weinberg’s (1983) second study: the adolescent adoption study.

IQ correlation of adopted children who were reared together for 18 years was zero.

Explanation: niche-picking; biologically related children select similar environments, and genetic heritability of intelligence becomes higher with age.

The influence of environment on genetics: regulation of gene expression.

Gene expression: transcription and translation.

Regulation of gene expression: epigenetic changes.

Having a gene does not automatically mean that this gene will be manifested in the phenotype.

Methylation: the process when chemicals are added to the DNA molecule and repress gene transcription.

Behavioural epigenetics: regulating response to stress.

Weaver et al (2004): rats raised by mothers who were less nurturing were more sensitive to stress when they became adults. This was linked to the suppression of the glucocorticoid receptor gene, meaning a smaller number of glucocorticoid receptors in the brain and so increased production of stress hormones.
Genotype and phenotype

All cells in the human body that have a nucleus contain a set of chromosomes (from the Greek chroma meaning “colour” and soma meaning “body”, which is due to their strong staining by certain dyes). A chromosome is a thread-like structure that contains a DNA molecule. The long DNA molecule is tightly coiled many times around supporting proteins, so a chromosome is a “package” that contains folded DNA.

DNA (deoxyribonucleic acid) stores information. It is a code made up of a long sequence of four chemical bases (A = adenine, G = guanine, C = cytosine, T = thymine). The bases are paired up, making a sequence of base pairs. The DNA has a characteristic structure of the double helix which looks a bit like a ladder where base pairs are the ladder’s rungs. Information is coded in this sequence of bases like letters in a sentence (change the order of letters and you get a different sentence). This is

Genotype and phenotype

Research with humans is limited because brain tissue has to be obtained (through post-mortem examination). Also, research is contradictory.

Miller et al (2009): studying gene expression in people raised in poverty versus wealthy environments, the researchers expected to find suppression of the glucocorticoid receptor genes, but did not. However, they analysed blood cells and not brain cells.

McGowan et al (2009): conducted post-mortem examinations of brains of 24 individuals who had committed suicide. People who had been abused as a child had more chemicals in their brain cells suppressing the expression of the glucocorticoid receptor gene.

Behavioural epigenetics: personality traits

Studies suggest that measurable environmental differences cannot explain all the discordance in identical twins’ phenotypes. Epigenetics can be a factor that explains this.

Kaminsky et al (2008): in a case study of a pair of identical twins the researchers ran epigenetic tests of DNA extracted from blood cells

One of the twins was more risk-averse and had a tendency to overreact to minor problems with a high degree of anxiety.

The DLX1 gene of this twin was significantly more methylated. This gene is involved in the production of neurons that form a part of the stress centre of the brain

This section also links to:

- Unit 1 “Research methodology”
- Unit 5 “Abnormal psychology”: biological explanations for depression
- Caspi et al (2003), Chiao and Blizinsky (2010): molecular genetics
- Gene–environment correlation (rGE)

Exam tip

This section is longer than the others because it embraces two topics from the syllabus, which are:

- genes and behaviour
- genetic similarities.

Genes and behaviour is a more general topic so whatever is covered in this section is relevant to it, with a focus on the nature-nurture debate. Genetic similarities is a more specific topic as it relates to the research methods (and studies) based on the idea of similarity between relatives and non-relatives: twin studies, family studies and adoption studies. The second topic is “embedded” in the first topic.
an incredibly long sentence, though: human DNA consists of about 3 billion bases.

As a long sentence will break into lines, this long sequence of chemical bases is broken up into 23 chromosomes, so each chromosome contains a part of the sequence. Each chromosome is present twice in each cell (except for sex cells). Humans have 23 pairs of chromosomes. One of the chromosomes in each pair is from your mother and the other one from your father. Both the chromosomes in the pair have a code for identical characteristics (height, eye colour, and so on), but the chromosomes themselves might not be identical.

If DNA is one extremely long sentence, and base pairs are letters, then genes are probably words. A **gene** is a unit of heredity, a region of DNA that encodes a specific trait or function. For example, there is a gene for eye colour, a gene for height, and so on. The total number of genes in the human organism is currently estimated to be around 20,000.

To summarize our metaphor, human DNA is a sentence that consists of 23 lines, 20,000 words and 3 billion letters. Each word is spelled twice, once by the father and once by the mother—and it may be spelled a little differently by each. The combination of these two spellings determines the trait or a function.

What we have referred to as “spelling” are components known as **allele**. Alleles are different forms of the gene. They can be dominant or recessive. The trait controlled by the recessive
allele only develops if the allele is present in both chromosomes in the pair, whereas the trait controlled by the dominant allele will develop if at least one of the chromosomes in the pair contains it. For example, in the gene that codes for eye colour the allele for brown eyes is dominant and the allele for blue eyes is recessive. So you will have blue eyes only if both the alleles in your chromosome pair are recessive. In the other three combinations your eyes will be brown.

The set of traits as coded in an individual’s DNA is called **genotype**. The set of traits that actually manifest in an individual’s body, appearance or behaviour is called **phenotype**. Phenotype comprises observable characteristics (eye colour, height, and so on) and unobservable characteristics (blood type, immune system, and so on), as well as behaviour. Genotype is the “plan” and phenotype is its implementation.

**The nature-nurture debate**

Nature-nurture is the long-lasting debate in psychology and philosophy that attempts to establish whether human behaviour is determined primarily by biological factors such as genetics and brain structure (that is, nature) or environmental factors such as education and friends (that is, nurture).

Today it is widely recognized that the debate needs to be reformulated. There is little doubt that human behaviour is influenced by both nature and nurture. What is more interesting is to answer the following questions.

- What are the relative contributions of biological and environmental factors into a specific trait or behaviour? For example, what is the relative importance of biological factors as compared to the environment in developing intelligence? Can we quantify these relative contributions?
- How do biological and environmental influences interact? For example, can biological factors influence environmental factors and only then influence behaviour? Can environment influence biological factors such as genetics?

**Methods of research**

The main methods used to study the influence of genotype on behaviour are as follows.

- **Twin studies.** The main principle is estimating the similarity between identical (monozygotic—MZ) twins and comparing it to the similarity between fraternal (dizygotic—DZ) twins. MZ twins develop from the same egg and share 100% of genotype. DZ twins develop from different eggs and share 50% of genotype, just like regular siblings. If identical twins are more similar to each other than fraternal twins, we can attribute it to genetic influences.
- **Family studies.** This method also uses the principle of genetic relatedness, but compares relatives on a broader scale and across generations: for example comparing children to parents, grandparents, siblings, cousins, and uncles and aunts.
- **Adoption studies.** These compare adopted children to their adoptive parents, biological parents, adoptive siblings and biological siblings. We can infer genetic influences if adopted children are more similar to their biological parents than to their adoptive parents.
- **Molecular genetics.** Studies of molecular genetics are based on using modern technology for genetic mapping and identifying the alleles of particular genes in a particular individual. Genetic variants are then correlated with observed behaviour. These methods are usually used to identify specific genes responsible for specific behaviour—the “gene of depression”, “gene of aggression”, and so on.

The first three methods involve the use of **genetic similarity** as the principle of research.

**Genetic heritability: the Falconer model**

**Genetic heritability** is the quantitative measure of the relative contribution of genetic factors into a trait or behaviour. Estimation of genetic heritability is performed in **twin studies** and is based on the so-called **Falconer model**, which assumes that phenotype is comprised of three types of influence. These are:

- genetics
- shared environment
- individual environment.

Shared environment is the part of environmental influences that is common to the two twins (such as similar schooling, and the same exposure to books and technology). Individual environment comprises environmental influences that are
unique to each of the twins (different friends at school, different hobbies, and so on). This idea can be written in the following form:

\[ I = A + C + E \]

(where \( A \) = genetic inheritance, \( C \) = shared or common environment, \( E \) = individual environment).

In this formula, \( I \) means that the combination of these three influences theoretically can explain 100\% of observed variation in phenotype. In other words, there exist no other factors that influence a certain trait or behaviour apart from these three. Heritability in this model equals \( A \).

The influence of genetics on the environment: niche-picking

Genes and environment are not completely independent: in many instances genes influence environment too. So we need to look at how the interaction between these two factors develops dynamically. One form of this dynamic development is **niche-picking**: the phenomenon when genetic predisposition causes individuals to select environments that, in turn, start to affect their behaviour. For example, a child predisposed to depression may intentionally seek out high-demanding environments where it is hard to succeed.

Niche-picking may explain one interesting property of heritability coefficients: they change during life, typically becoming larger. This means that if you use a sample of adolescent twins and the Falconer model to arrive at an estimate of heritability (\( A \)), this estimate will typically be smaller than if you use a sample of older twins. As you grow up,
your genetic programme “unfolds” causing you to choose certain “niches” in the environment. In this way, in terms of their behaviour, MZ twins become more and more similar with age. This phenomenon cannot be explained by the Falconer model.

**Exercise**

Focus on intelligence: what are the relative contributions of nature and nurture in a person’s IQ? Do you think your intellectual abilities are mostly due to the genes you inherited from your parents, the environment around you or your own efforts? Give your reasons.

---

**Heritability of intelligence: twin studies**

Bouchard and McGue (1981) conducted a meta-analysis of 111 studies on IQ correlations between relatives.

The median correlations they obtained are shown in Table 2.3. Let’s take a close look at the table. MZ twins share 100% of their genes because they develop from the same egg. DZ twins share 50% of genes, just like siblings and just like parents with their biological children; (children take roughly half of their genes from each of the parents). There is no genetic similarity between adopting parents and their adopted children.

<table>
<thead>
<tr>
<th>Expected similarity</th>
<th>IQ correlation between</th>
<th>% of shared genes</th>
<th>Median correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MZ twins reared together</td>
<td>100</td>
<td>0.85</td>
</tr>
<tr>
<td>2</td>
<td>MZ twins reared apart</td>
<td>100</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>DZ twins reared together</td>
<td>50</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>Siblings reared together</td>
<td>50</td>
<td>0.45</td>
</tr>
<tr>
<td>3</td>
<td>Parent and offspring reared together</td>
<td>50</td>
<td>0.39</td>
</tr>
<tr>
<td>4</td>
<td>Siblings reared apart</td>
<td>50</td>
<td>0.24</td>
</tr>
<tr>
<td>4</td>
<td>Parent and offspring reared apart</td>
<td>50</td>
<td>0.22</td>
</tr>
<tr>
<td>5</td>
<td>Adopting parent and offspring</td>
<td>0</td>
<td>0.18</td>
</tr>
</tbody>
</table>

▲ Table 2.3

Imagine genes were the only thing causing IQ differences between people. In this case we should observe the following pattern: MZ twins have a perfect correlation of 1 (irrespective of whether they are reared together or apart), DZ twins, siblings and parents with their biological children have the second largest correlation, and adopting parents and offspring a correlation of zero.

However, environment also contributes to the variability of IQ, so MZ twins reared together are expected to have a higher correlation than MZ twins reared apart due to the exposure to a common environment. Taking this into account, the expected degree of similarity is given in the first column of Table 2.3. As you see, the median correlations obtained from the summary of 111 studies follow this predicted pattern.

Note the following points.

- At the same time, a correlation with an effect size 0.85 is large. (It would be a good idea to go back to Unit 1, “Research methodology”, and review the boundaries of correlation coefficients.)
- If you put two of the values—correlation between MZ twins reared together and correlation between DZ twins reared together—into Falconer’s formula (see above), you obtain the estimate of heritability of IQ: $2 \times (0.85 - 0.58) = 2 \times 0.27 = 54\%$
  
In other words, intelligence (based on the results of this review and the Falconer model) is 54% inherited.

All in all, results of the study demonstrate that intelligence is to a large extent (54%) genetically inherited. However, one needs to keep in mind the typical limitations inherent in any twin study (Joseph 2015).
The assumption that similarity between MZ twins reared apart is solely due to genotype is limited, for these reasons.

- Many twin pairs were not separated immediately after birth, so they experienced some formative months or years together.
- Many twin pairs, even when separated, grew up in similar cultural and SES environments. They were not “randomly allocated” into different environments.
- Twins share a common prenatal environment. Moreover, prenatal environment of MZ twins is more similar than that of DZ twins.

- Twin studies are usually small in sample size and rare due to the uniqueness of their target group. This implies fewer opportunities for replication.
- Twins might not be as representative of the general population as we would like them to be, so twin study findings might not be generalizable to a wider population.
- The similar physical features of the twins might elicit similar responses from the environment (for example, it is known that attractive people are treated better than average-looking people).

**Heritability of intelligence: adoption studies**

Adoption studies provide a direct test of environmental malleability of cognitive abilities. There are two aspects of adoption studies that may provide slightly different angles on the nature-nurture problem. These aspects are:

- computing the correlation between cognitive abilities of the adopted child and the adoptive parents and comparing it to the correlation between cognitive abilities of the adopted child and the biological parents
- comparing cognitive abilities of adopted children to those of their siblings who were not adopted but raised by their biological parents.

Interestingly, these two approaches yield contradictory results. In general, most of the existing studies support the idea that IQ is increased by adoption into more prosperous families. This is demonstrated by comparing the average IQ of children adopted into higher-SES families and the average IQ of their biological home-reared siblings. At the same time the same studies demonstrate that adopted child–biological parent correlations are always higher than adopted child–adoptive parent correlations, suggesting that the genetic component in cognitive abilities is strong. Together these two effects suggest the **additive influence** of genetics and environment on the development of intelligence: adopting into a higher-SES family results in an increase in IQ, but this increase will be higher or lower depending on the genetic inheritance of the child.

An example of a study that demonstrated this additive influence is **Kendler et al (2015)**. The researchers conducted a rigorously designed adoption study of a sample of sibling pairs in which one of the siblings was home-reared and the other one was adopted away.

The complete national Swedish register of male-male siblings was searched, initially identifying 436 male sibling sets where one of the members was reared by adoptive parents. IQ scores were taken from the Military Conscription Register (which includes cognitive assessment data for all 18-year-old men in Sweden). Available data also included the educational attainment of both biological and adoptive parents.

Demand for child adoption in Sweden was considerably larger than the number of children available for adoption, so potential adoptive parents were carefully screened. The mean educational level was significantly higher in the group of adoptive parents as compared to biological parents. There was a modest correlation ($r = 0.18$) between the educational levels of biological and adoptive parents, which may suggest some effects of **selective placement**.
### ATL skills: Research

Selective placement is the main limitation of adoption studies. It occurs because adoption agencies take special care to place children in environments that are similar to the biological parents’ environment.

Why do you think they do that?

To what extent do you think selective placement might have compromised the results of this study?

Results of the study are summarized in Table 2.4.

<table>
<thead>
<tr>
<th>Mean IQ at age 18</th>
<th>Correlation with Education of biological parents</th>
<th>Correlation with Education of adoptive parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adopted siblings</td>
<td>96.9</td>
<td>0.20</td>
</tr>
<tr>
<td>Home-reared siblings</td>
<td>92.0</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Table 2.4

Interpretation of these findings suggests that cognitive ability is environmentally malleable: there was a 5-point IQ increase on average by age 18. The fact that there is a correlation between cognitive ability of adopted children and educational levels of adoptive parents supports this conclusion. On the other hand, results also suggest heritability of intelligence: this is evident from the correlation between cognitive ability of adopted children and the educational level of biological parents. Results seem to suggest an additive influence of environment and genetics: the largest IQ scores were observed in adopted children from well-educated biological families adopted into well-educated families.

**Scarr and Weinberg (1983)** reported on the results of two longitudinal studies launched in 1974, both of which investigated malleability of intelligence. One of the studies—**the Transracial Adoption Study**—was designed to see if black children reared by white families performed on tests of IQ and school achievement as well as other adoptees. The other study—**the Adolescent Adoption Study**—looked at how differences in cognitive ability accumulate over years till adolescence. Collectively these two studies are known as the **Minnesota Adoption Studies**.

**The Transracial Adoption Study** sampled 101 adoptive families who had their biological children but who also adopted transracially. Some of the adopted children were black and some white, also some children were adopted in the first year of life and some after 12 months of age. All children were assessed on IQ and school achievement tests.

### ATL skills: Thinking

Why would the authors of the study assume that adoption of black children into white families would result in cognitive benefits?

To understand the rationale behind the study, note the following background details.

- Back in 1974, random samples of black and white participants did not perform equally well on tests of intelligence and school achievement (the average IQ for white participants was around 100 while the average IQ for black participants was around 85–90). This could mostly be attributed to a lower SES status of black families, as well as stronger exposure of the white population to the culture of achievement, schools and tests.
- The black population of Minnesota was small in 1974 (around 1% of the total population). There were many black children available for adoption and few black families to adopt them.
- Typically, adoption agencies place children in well-off families with higher-than-average SES and non-abusive environments. All adoptive families that participated in the study could be classified as highly educated and above average in income.
- So a black child adopted into a white family at that time would typically end up in a richer environment. If intelligence is malleable, we can expect IQ scores of adopted children to be higher than the population baseline.
Table 2.5 summarizes the main results of the study. Overall, results of the study support the idea of additive influence of genetics and the environment to the development of IQ. In this sense it corroborated the findings of Kendler et al (2015).

On the one hand, adoption increased cognitive abilities of adopted children.

- Black children placed in white families increased their IQ scores substantially as compared to black children reared in their own homes (an increase of 16 IQ points).
- Early adoption resulted in higher IQ scores than late adoption. Black children adopted early scored 110 IQ points on average—almost as high as the average score of adopted white children.

On the other hand, just as in Kendler et al’s (2015) study, the correlation between the IQ of adoptive parents and adopted children was lower than the correlation between adopted children and their biological parents (0.29 and 0.43 respectively). From these data researchers estimated that 40–70% of IQ variance in the sample was due to genetic differences among the children (Scarr and Weinberg 1983: 262–63).

The additive influence suggests that an increase in IQ occurs due to environmental factors, but also suggests that how responsive a particular child would be to environmental influences depends on the child’s genetics. Children with “good” genetics placed in “good” environments get double advantage.

A contradictory finding from the same study, however, is that young siblings were very similar to each other, whether they were genetically related or not. The IQ correlations of adopted siblings were very nearly as high as those of the biological siblings reared together (0.44 and 0.42 respectively): there is the same amount of similarity between two adopted siblings as there is between two natural siblings. This finding can only be explained by the predominant influence of the rearing environment on the development of IQ.

This contradiction is reconciled in the second study—the Adolescent Adoption Study. In this study participants were adolescents who had been adopted early in the first year of life and spent an average of 18 years in their new families. Their adoptive parents, their biological parents and biological children of their adoptive parents also participated in the study. The IQ correlations of the biologically related siblings were similar to those in the Transracial Adoption Study: 0.35. However, the IQ correlation of adopted children who were reared together for 18 years was zero.

Researchers concluded that observed results may be due to niche-picking. Young children reared in the same family are similar to each other, no matter whether they are genetically related or not, because they share a similar rearing environment. On the other hand, older adolescents are similar only if they share genes. This may mean that they have escaped the influences of the family and are now free to select their own environment.

The fact that biologically related children select similar environments (which may explain the observed correlation of 0.35) is an example of niche-picking.

In this way, genetically related people become more and more similar with age as the genetic
programme “unfolds” and the child begins to pick his or her “niches” in the environment. This process can either strengthen or weaken the effects of the rearing environment.

Exercise

The Minnesota twin projects are ongoing. The Minnesota Center for Twin and Family Research (MCTFR) is currently heading five different projects, including two that added modern technology to twin research: the GEDI project has added the study of DNA and the MRI project has added MRI parameters to the list of characteristics that are compared in twins. You can learn more about the projects here: http://mctfr.psych.umn.edu/aboutus/index.html

ATL skills: Self-management

Write down a summary of the findings of these three studies:
- Bouchard and McGue (1981)

For each study highlight the findings that are most essential in the context of the nature-nurture debate. Compare the findings.

The influence of environment on genetics: regulation of gene expression

ATL skills: Thinking

We have seen how genetics can influence a trait either directly or indirectly through niche-picking. Do you think the environment can influence genetics? How?

Biologically, genotype becomes manifested as phenotype through a process called gene expression. Each gene contains instructions for the synthesis of a functional product—in most cases a protein—a molecule which will then influence the chemical composition of the cells that determine the trait (eye colour, for example). For simplicity, from now on we will call all functional products proteins. Proteins usually are a chain of amino acids.

The process of constructing a protein based on the plan encoded in the DNA involves two major steps: transcription and translation. In transcription, the sequence of the gene is copied to make an RNA (ribonucleic acid) molecule. In translation, the RNA molecule is decoded into a sequence of amino acids in a protein. So transcription uses the same “language” of base pairs while translation uses a different language of amino acids in a protein. Transcription is like photocopying and translation is like reading aloud from the photocopy.

In humans transcription takes place in the nucleus of a cell, and translation takes place in cell structures known as ribosomes. A ribosome latches onto the RNA molecule, finds its starting point (indicated by a special chemical) and rapidly moves along the RNA, synthesizing one amino acid at a time, building a protein that mirrors the RNA. Once the protein is finished, it is transported to its destination (either within the cell or not) and performs its job.

A wide range of sophisticated mechanisms can be used by the body to increase or decrease the production of proteins based on the genetic code. Collectively these mechanisms are known as regulation of gene expression. The important implication here is that having a gene does not automatically mean that this gene will be manifested in the phenotype. Any step of gene expression can be modulated, from the DNA-RNA transcription to modification of a protein after
translation. Some genes can be suppressed completely. The process when chemicals are added to the DNA molecule, and so repress gene transcription, is known as **methylation**. Imagine your photocopying machine started to print certain words indistinctly, as if trying to refuse to print these words. That’s what methylation can probably be compared to.

Regulation of gene expression results in **epigenetic changes** (from the Greek *epi* meaning “over” or “outside of”)—deviation of phenotypes from the genetic code in the DNA sequence. Epigenetic changes can be attributed to environmental influences, and in this sense it is a study of how nurture influences nature.

Epigenetic changes can influence the growth of neurons in the developing brain of a child and influence brain activity in adults. Both processes result in a change of behaviour.

**Behavioural epigenetics: regulating response to stress**

Behavioural epigenetics was demonstrated in the pioneering research of **Weaver et al (2004)**. They found that the type of nurturing rats receive from their mothers in the young ones’ early life affects the way their brain responds to stress later in life. More specifically, rats raised by mothers that were less nurturing (for example, licked and groomed their young less often) were more sensitive to stress when they became adults. For example, when their movements were restricted (by placing them in a narrow tube), their adrenal glands produced more stress hormones. This increased production of stress hormones was linked to a fewer number of receptors for these hormones in the brain (specifically **glucocorticoid receptors**). In its turn, the smaller number of glucocorticoid receptors in the brain was linked to the suppression of the **glucocorticoid receptor gene**. The gene itself did not differ in the groups of rats receiving different nurturing, but rats raised by less-nurturing mothers had more chemicals that inhibited transcription of the glucocorticoid receptor gene. As a result, fewer receptors were produced; more stress hormones were released; and the organism suffered more consequences of stress.

To confirm their findings, the researchers conducted studies where they gave rats substances that reversed the effects of transcription suppression for that particular gene, and this treatment normalized responses to stress even in less-nurtured rats (Miller 2010).

Such research studies have far-reaching implications. If similar mechanisms are demonstrated in humans, then we can pinpoint specific causes for certain behavioural changes on the level of proteins. For example, there are well-researched and documented effects of poverty in early childhood on health and behavioural patterns later in life (see Unit 8. “Developmental psychology”). Children who experienced poverty early in life undergo changes in cognitive abilities, social behaviour and other areas. The question is: can the mechanism of these changes be behavioural epigenetics? Imagine we found that the effects of extreme early poverty on cognitive development of children can be traced back to specific chemicals suppressing the...
expression of certain genes—then can we invent a drug that would suppress the effects of these chemicals and so reverse the influence of poverty?

While this area of research is intriguing, there is a major issue that slows down the development of our knowledge in this field. To study gene expression, human brain tissue needs to be obtained—and the only way to do it (ethically) is through post-mortem examination.

Evidence in this area is inconclusive. Miller et al (2009) studied gene expression in people raised in poverty versus wealthy environments. The researchers expected to find increased concentrations of chemicals that suppress the glucocorticoid receptor genes, as predicted by Weaver et al’s research with rodents, but they didn’t. However, they measured these chemicals in white blood cells, and arguably epigenetic changes in brain and blood cells might not be the same. McGowan et al (2009) conducted post-mortem examinations of the brains of 24 individuals who had committed suicide. Half of these people had been abused in childhood. Examination revealed epigenetic changes in brain cells similar to those in the rodent study: people who had been abused as children had more chemicals in their brain cells suppressing the expression of the glucocorticoid receptor gene.

The work is still ongoing, and it is a burning topic, so any wide generalizations would at this point be premature (Miller 2010).

**Behavioural epigenetics: personality traits**

MZ (identical) twins are 100% similar in terms of their DNA sequence. However, certain phenotypical differences between MZ twins are observed. Traditionally these differences have been attributed to individual environments. However, some recent studies suggest that measurable environmental differences cannot explain all the discordance in identical twins’ phenotypes. There may exist other factors over and above what was suggested in the Falconer model. One such possible factor is epigenetics.

Kaminsky et al (2008) conducted an extensive case study of a pair of identical twins using cognitive and personality tests as well as genetic and epigenetic tests. Epigenetic testing was performed on DNA extracted from blood cells. Participants were two 49-year-old female MZ twins, one of them a war journalist and the other an office manager in a law firm. When they were young they were very close to each other and their parents tried to raise them in the same way and ensure they were undistinguishable (for example, the parents dressed the girls the same). At age 17 the “war twin” left home, travelled a lot and ended up choosing the career of a war journalist working in multiple war zones in Africa, the Middle East and the Balkans. She was exposed to atrocities of war, saw people killed and lost close colleagues. She married in her forties and never had children. Occasionally she drank alcohol in excess. Her sister’s life turned out to be quite different. She settled down early with a career in law, married young and soon had two children. She drank alcohol occasionally, but never in excess. Despite living far from each other, the “war twin” and the “law twin” remained emotionally close and met as often as they could.

Personality questionnaires showed a difference in that whereas the war twin’s profile appeared normal, the law twin had a tendency to overreact to minor problems with a high degree of anxiety and tension. According to tests, the law twin also was more risk-aversive than the war twin.

Variations in gene expression were examined by comparing methylation pairwise in 12,192 DNA regions (genes). Results showed that one particular gene was differentially methylated in the war twin and in the law twin. This was the DLX1 gene. This gene is known to be involved in the production of neurons that form a part of the stress centre of the brain. The authors suggest that this discordance in DLX1 gene methylation can explain the reduced overall level of anxiety in the war twin as compared to the law twin.

Note that a cause–effect relationship in the study should be inferred with caution because differences in DLX1 methylation may themselves have been caused by the environment. The researchers also recognize that one twin pair is not enough to make definitive statements about the role of DLX1 methylation in the development of stress responses in MZ twins. Another important limitation of the study is that DNA was obtained from blood cells and, as you know, gene expression in blood and brain cells may be different.
In this section we have looked at various aspects of genetic influences on behaviour. These are the key points.

- Different traits are influenced by genetic inheritance to different degrees.
- The traditional way to estimate this influence is through studies based on genetic similarity (twin, family and adoption studies).
- Using the example of intelligence, we have seen that these studies demonstrate additive influence of genetics and environment. On the one hand, adoption into more enriched environments results in an increase in average IQ scores. On the other hand, IQ of adopted children correlates more strongly with that of their biological parents. This suggests that biological influences add to environmental influences.
- Another finding is that genetic heritability increases with age. Similarity between related individuals increases as they grow older. This suggests niche-picking: biological factors can influence environmental factors.
- Environmental factors can influence biological factors through the process of regulation of gene expression. Genes can be switched on and off in response to environmental influences.
- We have therefore seen that there is a complex dynamic interaction between genetic and environmental factors, which makes the nature-nurture debate in its original form outdated.

Review the section and find the supporting arguments and evidence for each of these key points.

To what extent would it be possible—and acceptable—to use genetics in Humanborough in the following projects?

- Using knowledge of heritability of certain traits and genetic mapping, you set out to establish a dating agency that would match people based on the predicted qualities of their offspring.
- Your aim is to invent a drug that would reverse the effects of methylation of the glucocorticoid gene and market it as a drug that “reverses the effects of bad parenting”.
- You want to enhance requirements for selective placement in adoption agencies.
Evolutionary explanations for behaviour

What you will learn in this section

- The theory of evolution
  - The need to survive and reproduce
  - Differential fitness
  - Survival of the fittest
  - Natural selection
- A range of evolutionary explanations in psychology
  - Chiao and Blizinsky (2010): gene-culture co-evolution theory
  - LeDoux (1996): brain pathways of processing fear stimuli
  - Call and Tomasello (2008): theory of mind, animal research
  - Harlow (1958), Shaver and Hazan (1988): theory of attachment (from animal to human research)
- Evolutionary explanation for disgust
  - Curtis, Aunger and Rabie (2004): disgust evolved as a protection from risk of disease
  - Five specific hypotheses. A strong support for the evolutionary explanation is only expected if all the five hypotheses are backed up by evidence
- Criticism of evolutionary explanations in psychology
  - Massive modularity versus neuroplasticity
  - Speculations about the environment
  - Testability
  - Assumptions about linearity of development
  - Cultural variation
  - Adaptation versus other evolutionary mechanisms

This section also links to:

- Unit 3 “Cognitive approach to behaviour”: emotional processing
- Unit 5 “Abnormal psychology”: biological explanations for depression
- Unit 7 “Psychology of human relationships”: prosocial behaviour, altruism, formation of personal relationships
- Unit 8 “Developmental psychology”: theory of mind
- genetics and animal studies

Inquiry questions

Is the theory of evolution useful for explaining human behaviour?

What is the explanatory power of evolutionary explanations for behaviour? Can they be used for all possible behaviours?

If we are genetically similar to animals and genes affect behaviour, does it mean we are behaviourally similar to animals too?
One major conclusion that can be drawn from research into genetic foundations of behaviour is that genes can code for behaviour as well as physical traits.

We know that physical traits are subject to evolutionary pressures. If genes code for behaviour as well as physical traits, does it mean that behaviour is subject to evolutionary pressures too? This seems to be a reasonable idea. This is why a variety of behaviours (more accurately, the origin of these behaviours) have been explained using evolutionary reasoning.

First we will briefly review the ideas of evolution, then review some major examples of evolutionary explanations in psychology, including the ones you will come across in other units, and then we will focus on one example—the emotion of disgust—to demonstrate typical reasoning of evolutionary psychologists. We will also discuss major flaws and limitations of evolutionary explanations for behaviour.

The theory of evolution

Evolution is the process by which organisms change from generation to generation as a result of a change in heritable characteristics. It is not just any random change; as suggested by Charles Darwin, there is deep logic to this process.

The modern theory of evolution (a combination of Darwin’s theory with the discoveries of genetics) is based on the following premises.

- Biological organisms are driven by the need to survive and reproduce.
- There is considerable variation in the traits of individual organisms from the same population. Organisms having different traits are adapted to their environment to varying degrees—some better, some worse. This is called “differential fitness”.
- Those organisms that are well adapted to the environment have higher chances of surviving and producing offspring. Organisms that are less adapted die out or are unable to produce offspring. This is called “survival of the fittest”.
- Gradually as those organisms that are less adapted do not pass on their genes, those genes disappear from the population gene pool. More adapted organisms produce more offspring, so their genes in the gene pool get stronger. This is called “natural selection”.

As the environment changes, organisms need to adapt to this change. Scarce resources (such as food and mates) make organisms fight for survival.

The theory of evolution has a great explanatory power. It explains the variety of species and their modifications that we observe in the world by placing all these species into a developmental historical perspective. It also traces back common ancestors for all organisms, including humans.

Modern research has shown that in terms of their DNA humans are 99.5% similar to each other. As a species, we share 98% of genes with chimpanzees, 90% with cats, 69% with rats, and 60% with chickens and fruit flies. This hints at the possibility of using animals to get an insight into human behaviour and links to the principles of a biological approach to behaviour.

TOK

Can you name examples from various areas of knowledge (history, natural and human sciences, mathematics, arts, religious knowledge systems, indigenous knowledge systems, ethics) where evolutionary ideas have been influential?

ATL skills: Thinking

Review the principles of a biological approach to behaviour from the start of this unit. How are they linked to the idea that behaviour can have an evolutionary basis?

A range of evolutionary explanations in psychology

Evolutionary psychology attempts to explain psychological traits or behaviours as adaptations. Evolutionary explanations in psychology have been proposed for a wide range of phenomena. Here is just a brief overview of some popular explanations, some of which are discussed in greater detail elsewhere in this book.
<table>
<thead>
<tr>
<th>Topic</th>
<th>Evolutionary explanation</th>
<th>Notes</th>
</tr>
</thead>
</table>
| Mental disorders | Chiao and Blizinsky (2010) proposed **gene-culture co-evolution theory** to explain the higher prevalence of genetic susceptibility to depression in collectivistic societies (where the prevalence of depression itself is lower). In their conclusions they synthesized three findings.  
- Short alleles of 5-HTT, the serotonin transporter gene, make people more vulnerable to stressful life events.  
- There is a higher frequency of short alleles of 5-HTT in people in countries with collectivistic values.  
- In collectivistic societies people report depression more rarely. The researchers suggested that collectivistic values have evolved in these societies as a buffer against increased genetic susceptibility to depression. | See Unit 5, “Abnormal psychology” |
| Altruism (prosocial behaviour) | **Kin selection theory** explains altruistic behaviour observed both in animals and humans. Altruism does not fit well into the concept of natural selection because one organism helps another with no reward and even at some cost to itself. Kin selection theory suggests that the evolutionary meaning of altruism is the increase of survival of one’s genes rather than an individual. For this reason it predicts that altruistic acts will be more frequent in close relatives than in distant relatives. The theory was proposed by Hamilton (1964) and tested in a number of experiments including with human subjects (such as Madsen et al 2007). | See Unit 7 “Psychology of human relationships” |
| The influence of emotion on cognition and behaviour | LeDoux (1996) described two physiological pathways of emotions; namely, processing fear stimuli. The fast pathway goes through thalamus and amygdala, whereas the slow pathway also involves the structures of the hippocampus and neocortex. The evolutionary meaning of this is that our brain is hard-wired to produce a quick automatic reaction to fear stimuli, but this reaction can also be “overridden” in exceptional circumstances. This may be linked to principles of natural selection because in limited time and dangerous situations it adds survival value to react quickly rather than deeply assess the situation. The same explanation can be extended to stereotypes and cognitive biases. | See Unit 3 “Cognitive approach to behaviour” |
| Attachment | Research of attachment is closely linked with animal studies and animal models of behaviour. From the survival perspective, an organism maximizes its survival fitness if there is a balance between:  
- staying close to the attachment figures in unfamiliar potentially dangerous situations  
- venturing outside and exploring the world to develop necessary life skills.  
**Harlow’s** (1958) studies with monkeys demonstrate the existence of attachment styles and the importance of such behaviours as clinging and grooming. **Shaver and Hazan’s** (1988) research shows parallels between attachment styles in childhood and later adult relationship patterns in humans. | See Unit 8 “Developmental psychology” |
Evolutionary explanation for disgust

To demonstrate typical reasoning of evolutionary psychologists, we will look at one specific example.

Curtis, Aunger and Rabie (2004) published a study suggesting that disgust evolved as a protection from risk of disease. Researchers reasoned that if this was true, then the following conditions have to be fulfilled.

- Disgust should be felt more strongly when faced with a disease-salient stimulus as opposed to a similar stimulus with less salience.
- Disgust should operate in a similar way across cultures.
- Disgust should be more pronounced in females since they have to protect their babies in addition to themselves.
- Disgust should become weaker as the individual’s reproductive potential declines with age (there is less responsibility to care about offspring).
- Disgust should be stronger in contact with strangers than with close relatives because strangers potentially can carry novel pathogens.

Note that if any of these conditions were not fulfilled, this would present a challenge for the idea that disgust is a product of evolution. For example, if disgust was felt equally in response to disease-salient and less salient stimuli, then the proposed evolutionary explanation would fall apart: disgust would not be connected to risk of disease. Cross-cultural differences are also hardly compatible with the explanation: how can there be cross-cultural differences in something that is the product of evolution of humans as a species?

So, a strong support for the evolutionary explanation is only expected if all the five hypotheses are backed up by evidence. Evolutionary explanation in this case is a model, and we are trying to fit it into observational data. If it fits well, the evolutionary explanation is accepted (with caution), because it seems plausible. If it does not fit well, we change the model. The more observations we have that are consistent with the predictions of the model, the higher is our trust in the model itself.

ATL skills: Thinking

Can you name other examples where models are used in psychological research?

To test their hypotheses, Curtis, Aunger and Rabie (2004) used a survey placed on the BBC Science website. The survey was advertised in a BBC documentary. It was completed by over 77,000 people from 165 countries. However, after data cleaning, the final sample size was slightly less than 40,000. For example, all participants who had watched the BBC documentary were excluded because they could have been exposed to the hypothesis of the study.

First, respondents were asked a set of demographic questions on their age, sex, country, and so on. Then they were asked to rate 20 photographs (appearing one by one on separate screens) for disgust on a scale from 1 (not disgusting) to
5 (very disgusting). Of these photographs, 14 comprised 7 pairs of disease-salient versus less salient stimuli. For example, one photograph depicted a white towel with a blue stain on it, and the paired photograph showed the same towel with the stain depicted in reddish-yellow resembling blood and bodily secretions.

Therefore, all five tests supported the evolutionary explanation of disgust as a response that reduces risk of disease.

TOK

Now that many humans have constructed an artificial environment and no longer need to focus on physical survival to the same extent as before, do you think evolution is losing its power as an explanatory principle? Many people live in the world of business partners and competitors rather than physical dangers. While your economic status does affect your offspring, physical survival (in the developed countries) is no longer at stake. Does it mean that natural selection is not working anymore? If so, what was it replaced by?

Criticism of evolutionary explanations in psychology

Evolutionary psychology has a great explanatory power. In many instances evolutionary explanations fit nicely into our observations, tying them all together. However, some of the limitations that are commonly mentioned by critics are as follows.

• **Massive modularity versus neuroplasticity.** Researchers have attempted to expand the field and propose that mind on the whole is a product of evolutionary processes. If you suggest that mind is a product of evolution, though, you must also make one major assumption—modularity of mind (also known as “massive modularity”). It states that mind consists of modules that have evolved to perform certain fitness-related functions (Samuels 1998). These modules must have a neurological basis. However, what we know about neuroplasticity contradicts this assumption. If the brain can change itself dramatically during the course of life, boundaries between modules are erased. This raises this question: how much neuroplasticity would be enough for us to challenge the assumption that the “modules” even existed in the first place? There have been advances that demonstrated that certain modules do exist. For example, macaques have been shown to have a specialized group of neurons that function as a snake-detection brain module. These neurons

Results showed support for all five hypotheses.

• First, disease-salient stimuli were rated as more disgusting than less salient ones. For example, the plate of organic-looking fluid was rated as 61% more disgusting than the plate of blue fluid that looked chemical (ratings were 1.6 versus 2.6). For the towel pictures, the organic-looking substance produced much higher ratings of disgust than the blue chemical (1.6 versus 3.9).

• Second, the results were consistent across cultures.

• Third, females rated the disease-salient pictures as more disgusting than men. This was true for all the disease-salient pictures used in the study.

• Fourth, as predicted, there was an age-based decline in the sensitivity to disease-salient stimuli.

• Finally, there was one question in the survey that asked participants to choose with whom they would be less likely to share a toothbrush. The average responses were ranged in the following order: postman (least likely), the boss, the weatherman, a sibling, a best friend, the spouse or partner. This shows that disgust is felt more strongly in contact with strangers than with relatives.

\[ \text{Figure 2.41 Examples of photographs used in the study} \]

Source: Curtis, Aunger, Rabie (2004: 131)
respond very quickly to images of snakes even if the macaque has never seen a snake before (Le et al 2013). So, highly specialized modules such as snake detection cause little doubt, but it is massive modularity on the whole that is questionable.

- **Speculations about the environment.** Evolutionary adaptation is always adaptation to a certain environment. So to suggest an evolutionary explanation of a trait, you need to have knowledge about the environment in which this trait evolved. However, our knowledge of the environments in which Homo sapiens evolved as a species is scarce. A lot of reasoning in this field, arguably, is speculative.

- **Testability.** Evolutionary explanations for behaviour are difficult—and in many cases impossible—to test. Critics of evolutionary psychology say that these explanations rest on a logical fallacy known as ad hoc reasoning. Ad hoc fallacy, also known as “a just-so story”, takes a phenomenon as it exists and “cooks” an unverifiable story about how it came to be. Critics argue that such believable stories may be made up for just about anything: for example, you may explain altruism by survival of genes of relatives and egoism by survival of your own genes. Evolutionary psychologists respond that it is not exactly like that. For example, as you have seen, Curtis, Aunger and Rabie (2004) formulated a set of five predictions all of which, logically, had to be true if the evolutionary explanation was true. Then they tested these predictions. This is not exactly like “cooking a story”.

- **Assumptions about linearity of development.** This is a related argument. Evolutionary explanations have no other choice but to assume that a trait has been gradually evolving to perform a certain function. However, it is possible that at some point in the past it actually evolved to perform some other function, different from the one it is (presumably) performing now. An example of this is exaptation—the situation when a trait evolves to perform one function but later starts performing a different function. Bird feathers are an exaptation: they originally evolved for temperature regulation but later re-specialized on flight.

- **Cultural variation.** Just as neuroplasticity is not entirely compatible with the idea of massive modularity of mind, existing cultural variations in traits are not entirely compatible with the idea that these traits developed as a universal adaptation to the universal challenges that humans faced as a species. Of course, one can always claim that different geographical groups of people faced different environmental challenges, but it weakens the evolutionary argument. This is why evolutionary psychologists prefer to study universal traits. For example, Curtis, Aunger and Rabie (2004) chose to study the emotion of disgust—one of the few universal, basic emotions experienced and interpreted similarly across cultures.

- **Adaptation versus other evolutionary mechanisms.** It may be hard to distinguish between genuine adaptation and other, more neutral processes. Examples of these include genetic drift (random variations in genotype that occur naturally) and spandrels. A spandrel is a by-product of evolution, a trait that developed as a result of the evolution of some other characteristic (Gould, Lewontin 1979). An example of a spandrel is the midget arms of tyrannosaurus rex. A common explanation would be to assume that these arms were developed to serve a purpose—such as raising the animal up after sleep—but Gould and Lewontin (1979) dismiss such explanations as ridiculous. Instead they suggest that the animal’s tiny arms are simply a by-product of the rest of the body getting bigger and bigger. In other words, the arms did not change, the body did.

All limitations notwithstanding, evolutionary psychology remains a very promising field, mostly because it provides a theoretical framework that brings together multiple different observations and pieces of knowledge about human behaviour. No other theoretical framework with the same integrating potential exists so far. Modern psychology has been broken down into many research areas, but it lacks an overarching theory. Evolutionary psychology makes a good candidate for filling this gap.
Psychology in real life

To what extent can evolutionary explanations of behaviour be used in applied projects in Humanborough? Do they have any practical value at all?

For example, if you believe that disgust is an adaptation, you know what categories of people are more “resistant” to disgust (according to Curtis, Aunger and Rabie’s study, it was elderly males). Can this knowledge be used for practical purposes?

Review evolutionary explanations briefly discussed in this section other than those relating to disgust. Try to come up with one concrete practical project that is based on ideas of evolutionary psychology.

See video

Robert Wright in his talk “The evolution of compassion” (2009) looks at evolutionary roots of prosocial behaviour:
https://www.ted.com/talks/robert_wright_the_evolution_of_compassion

Lisa Nip in her talk “How humans could evolve to survive in space” (2015) claims that humans have reached a stage when they can start controlling how evolution further develops, and so they can modify their own bodies in the desirable direction:
https://www.ted.com/talks/lisa_nip_how_humans_could_evolve_to_survive_in_space
The role of animal research in understanding human behaviour (HL only)

Inquiry questions

Can animal studies provide an insight into human behaviour?
Is psychological experimentation with animals ethical?

What you will learn in this section

- The value of animal models in psychology research
  - Purposes of animal research
  - An animal model is a concept that refers to using animal research to test a certain cause–effect hypothesis about a certain human behaviour
  - Types of experimental manipulation used in animal models
  - Using animal research to inform our understanding of human behaviour relies on the assumption that animal and human brains are similar. MacLean (1990): the theory of triune brain.
  - Comparative neurobiology has discovered microscopic differences in certain brain areas. This suggests that the evolution of the brain might have been more complex than simply building newer structures upon older structures.
  - Therefore, in addition to comparing brain structure, we need to compare psychological functions

- Premack (2007): in order to prevent confusing similarities with equivalence, we need to focus on the areas of difference. Examples are teaching and short-term memory.
- Summary of pros and cons of working with animal models
- Examples of animal research
- Overview of the animal studies in this chapter
  - Lashley’s experiments with rats; Merzenich et al (1984): brain and behaviour
  - Weaver et al (2004): genetics and behaviour

- Ethical considerations in animal research
- Overview of ethical guidelines from APA Code of Ethics

This section also links to: any animal studies elsewhere in the book (see, for example, Unit 8, “Developmental psychology”, theory of mind in humans and animals).

Throughout this unit we have discussed biological foundations of behaviour: the nervous system and the endocrine system, genetic inheritance, evolutionary considerations. We have looked at several animal studies and discussed the findings in the context of human behaviour. The third principle of biological approach to behaviour (see “Introduction”) stated: “Animal research may inform our understanding of human behaviour”.

In some sense this principle is a consequence of the first two principles: “Behaviour is the product of physiology” and “Behaviour can be genetically inherited”. Human physiology and genetic set-up is similar to that of animals, which naturally suggests that animal research is to some extent generalizable to humans. The question is, what is this extent exactly?
This section summarizes some arguments related to the value of animal models in research in psychology. We will refer once again to animal studies covered in this unit, as these can be used to support the arguments. Finally, we will discuss ethical considerations in animal research.

### The value of animal models in psychology research

The number of animals used in psychological research annually in the USA alone has been estimated to be 1.25–2.5 million. About 7.5% of psychological research is animal-based (Shapiro 1998). The most popular species to be used in psychological research are rats, mice, pigeons, cats, rabbits, hamsters, dogs, chimpanzees and baboons.

Research differs in terms of purposes for which animals are used. Researchers in the field of comparative psychology are interested in animal research as an end in itself. They either focus on a particular species or compare this species to humans. In other cases, animals are studied as models of human beings and the expectation is that the findings will be universal and generalizable. A third group of researchers use animal studies to understand particular human conditions such as diseases. **An animal model** is a concept that refers to using animal research to test a certain cause–effect hypothesis about a certain human behaviour. So an animal model is not just broadly “using animals to understand human behaviour”. It is a specific model. For example, there are several animal models to explain depression: stress models (which explain the onset of depression by higher exposure to stressful situations), separation models (which explain depression by being separated from attachment figures), medical models (which explain depression by chemical imbalances in the brain), and more.

There are four major types of experimental manipulation used in animal models (Shapiro 1998). These four types are:

- genetic manipulation (when animals are bred in a certain way)
- invasive manipulations with the nervous system (parts of the brain are stimulated with electrodes, lesioned or removed)
- invasive manipulations with other body parts (parts may be stimulated by substances or damaged)
- behavioural and environmental manipulations (such as electric shocks for rats depending on their performance in a maze-learning task).

Using animal research to inform our understanding of human behaviour relies on the assumption that animal and human brains are similar. Comparison of animal and human brains seemingly conveys a very consistent story of evolution: as species evolved, new structures were built on top of older structures; so the deeper we go into the brain, the more “primitive” structures we will find. There is a popular theory of **triune brain** proposed by MacLean (1990). This theory divides the human brain into three parts: reptilian complex, paleomammalian complex (the limbic system), and neocortex. The idea is that the deeper brain structures can be found in animals as well; and the further down you go inside the brain, the further down you see in evolution. For example, the reptilian complex that you have in your brain should resemble the full brain of a reptile.

![Triune brain](image)

However, some recent developments in the field of comparative neurobiology have led to the discovery of microscopic differences in certain brain areas. For example, some brain areas that both humans and primates have in common were found to be different in terms of how neurons are structured within them. This suggests that the evolution of the brain might have been more complex than simply building newer structures upon older structures.

These discoveries led some scientists to argue that comparison of brain differences might not give us full understanding of how animals are psychologically similar to humans, and that we need to compare **psychological functions** as well.
(that is, we also need to look at the problem from the cognitive and the social perspective).

Premack (2007) carried out such a comparison. Premack argued that in order to prevent confusing similarities with equivalence, we need to focus on the important areas of difference between humans and animals relevant to psychological research. Every time we find a similarity, we need to ask ourselves: what is the dissimilarity? Let’s look at two examples.

One example is teaching. Some animals teach their young. For example, adult cats injure mice and bring them to their kittens so that the kittens can practise stalking and killing their victims. However, although this looks like some of the basic forms of teaching found in humans, human teaching is certainly more complex than that. For example, animals predominantly teach one thing—eating, while in humans the targets of teaching are very diverse (Premack 2007).

Another example is short-term memory. A chimpanzee has the same limit for the number of units it can remember without rehearsing as a human being—about seven units. If you are told a sequence of seven numbers (for example the numbers 8, 3, 6, 9, 1, 4, 2) and asked to repeat them (without rehearsing) several seconds later, you will be able to do so because your short-term memory can keep that much information. A chimpanzee would also be able to do this. There is a temptation to see a deep similarity between the species in terms of their short-term memory, but a closer examination shows that humans, unlike primates, are capable of chunking. For us the following sequence—54, 12, 47, 89, 71—is a sequence of 5 units, whereas for a chimpanzee it is a sequence of 10. It works even better if the units are letters and the “chunks” are words. It can be said that short-term memory in humans and chimpanzees is similar on some level, but not equivalent (Premack 2007).

Here is a summary of some advantages and disadvantages of working with animal models.

These are some of the advantages.

- Humans and animals are identical in many ways, both in terms of brain structure and genetically.
- Studies with animal models do produce results: useful models of human behaviour and life-saving treatments have been developed based on animal experimentation. For example, insulin was discovered in an experiment where dogs had their pancreas removed.
- Animal studies allow researchers to embrace the full lifespan. While human subjects often outlive researchers themselves, laboratory mice live 2–3 years and this presents an opportunity to see their behaviour across their lifespan and even across generations. This is especially helpful in genetic research.
- Animal research may be highly controlled. For example, the “knockout” technique has been developed to selectively switch off one of the genes in the DNA sequence. All other things being equal, this technique provides great insight into the function of individual genes. The ability to better control confounding variables means higher internal validity of experiments.
- Animal subjects are relatively inexpensive and easily accessible, easy to handle and manage.

Some of the disadvantages are as follows.

- Animals and humans are never exactly the same, and we can never know the extent of the difference. This means that animal research, if successful, still needs to be replicated with humans in order to be sure that findings are generalizable.
- Even if humans and animals are similar in some aspect biologically, they can still differ psychologically (see Premack 2007).
- When scientists develop new biomedical treatments for mental disorders, they usually first test them with mouse models. However, results from mouse models are never directly applied to humans. Even if
mouse models yield successful results, the drug needs to be tested on larger animals first. It is like a pyramid of generalization where mice are at the bottom and humans are at the top.

- Animals are tested in strictly controlled laboratory environments, so arguably they may be under stress. As a result, their reactions to experimental manipulations may not be quite the same as in their natural environments: there may be an issue with ecological validity.

- Although humans and animals are similar in many ways, they are still essentially different. For example, over 85 vaccines for HIV worked well in primates but all of them have failed in humans. On the contrary, some results that are negative in animals can actually turn out to be positive in humans. For example, aspirin proved dangerous for animals but it is now one of the most widely used drugs for humans.

Table: Examples of animal research

<table>
<thead>
<tr>
<th>Topic</th>
<th>Study</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain and behaviour (localization)</td>
<td>Lashley’s experiments with rats: removing varying portions of the cortex to see if memory of the maze disappears</td>
<td>Performance deterioration depends on the percentage of cortex destroyed but not on the location of the destroyed cells. This challenges the idea of localization of function for memory.</td>
</tr>
<tr>
<td>Brain and behaviour (neuroplasticity)</td>
<td>Merzenich et al (1984): cortical representations of the hand in adult owl monkeys</td>
<td>There was re-specialization of brain matter responsible for one digit so that it became responsible for other, adjacent digits.</td>
</tr>
<tr>
<td>Hormones and behaviour</td>
<td>Romero et al (2014): the role of oxytocin in promoting social bonds in mammals in non-reproductive contexts</td>
<td>Dogs were placed in a room with their owner and another dog. Dogs sprayed with oxytocin showed higher affiliation towards their owner. Similar results were observed for the other dog in the room: affiliation and approach behaviours were more frequent in the oxytocin condition.</td>
</tr>
<tr>
<td>Genetics and behaviour (epigenetics)</td>
<td>Weaver et al (2004): epigenetic research on how the type of nurturing rats receive from their mothers in early life affects the way their brain responds to stress in later life</td>
<td>Less nurturing in early life was linked to the suppression of the glucocorticoid receptor gene. The genetic sequences themselves did not differ. This study demonstrates the effects of gene suppression on behaviour.</td>
</tr>
</tbody>
</table>

There are also some useful examples that you will learn in other parts of the course. One of them is the extensive research of theory of mind in humans and animals (see Unit 8, “Developmental psychology”, “Theory of mind”).

Every time you encounter an animal study you should use the context to critically evaluate the value of animal models in that particular study. Use the advantages and disadvantages listed earlier in this section. The value of animal models links
directly to whether or not animal research can provide insight into human behaviour.

**Ethical considerations in animal research**

Either incorporated in their main code of ethics or separately, professional associations of psychologists in most countries publish guidelines for research with animals. For example, the American Psychological Association (APA) has a separate document for this purpose. The document outlines the main considerations that must be addressed at all stages of research involving animals. Here are just the most crucial of them (APA 2012).

- Any animal study should be justified “with a clear scientific purpose”. One of the following justifications may be used. The study will:
  - increase scientific knowledge of behaviour
  - determine replicability or generalizability of prior research
  - increase our understanding of a particular species
  - give results that will benefit humans or other animals.

- If non-human animals are chosen for research, it has to be ensured that the chosen species is the best choice to address the research question, the minimum required number of non-human participants is used, and it should be assumed that whatever procedures cause pain in humans would cause pain in animals too.

- All animal research proposals must be submitted to the Ethics Committee prior to conducting the study.

- Psychologists and their assistants conducting the study must be familiar with the species-specific characteristics of normal behaviour so that they will be able to tell when the animal subject is stressed or unhealthy.

- Laboratory animals must be given humane care.

- Whenever possible, the experimental procedures should be designed in a way that minimizes discomfort of the animal. APA guidelines also advise researchers to first test the painful stimuli to be used with non-human animals on themselves, whenever reasonable.

- If a research animal is observed to be in distress or chronic pain and this is not necessary for the aims of the study, it should be euthanized.

- Animals reared in the laboratory must not be released into the wild.

**Exercise**

To learn more about ethical guidelines in research with non-human animals read the full APA guidelines:
https://tinyurl.com/jw3yca2

**TOK**

Ethics as an area of knowledge involves the use of thought experiments. Here is one possible thought experiment that has been used as an argument to say that animal experiments are ethically justified.

Imagine you see a small van with 500 mice in it rolling slowly towards the edge of a cliff. There is no driver in the van. In its way there’s a stroller with a human baby in it. There are two possible outcomes.

- You push the stroller away and let the van roll slowly off the cliff, killing the mice.
- You do nothing and the stroller prevents the van from going over the cliff, but this kills the human baby.

Defenders of animal experimentation say that everyone would choose the first option because human life is more valuable than animal life. They also say that using animal studies to develop potentially life-saving medicine is equivalent to this thought experiment.

What do you think about this argument?

**Psychology in real life**

Review all the projects you have proposed for the people of Humanborough. What is the role of animal studies in these projects? After reading this section, would you enhance or reduce the animal research component in these projects?

**Exercise**

Summarize all the Humanborough projects on one poster and have a “gallery walk” with your classmates or even present the poster in a session for the whole school.