**Core study 1: Loftus and Palmer (eyewitness testimony)**

### Qs Experiment 1 (page 35)

1. The IV is the verb used to describe the impact (e.g., ‘smashed’, ‘collided’); the DV is the estimate of speed.
2. Participants’ mean speed estimate is higher if the critical word suggests a more violent impact, or participants’ mean speed estimate is higher if the critical word is ‘smashed’ rather than ‘hit’.
3. This was an independent measures design. One strength of using this design is that, unlike repeated measures, you do not have to worry about order effects (people getting e.g. bored or tired of the task, or getting better at it) as participants only do the task once.
4. The experimental groups were:
   - a. Hit (‘how fast were the cars going when they hit?’)
   - b. Smashed
   - c. Bumped
   - d. Collided
   - e. Contacted
5. Asking 10 questions – 1 target + 9 distractor questions – disguises the true purpose or interest of the study. This means it is less likely that participants will respond to the wording of the questions. This suggests, rather than a material change in memory, that the greater the effect of the leading question on participants (i.e. giving a higher speed estimate), the more likely they were to think there was broken glass.
6. It can be concluded that the leading question actually altered the person’s memory for the event.
7. Experiment 1 was looking at the effect of leading questions; experiment 2 went further to see if the reason that leading questions affect answers was because memory itself was altered rather than just being a response bias at the time of the question.

### Qs Experiment 2 (page 36)

1. A control group means a group which does not get the ‘treatment’ or exposure to any one of the conditions of the IV.
2. In this experiment it is a good idea because Loftus and Palmer need to know what people would normally recall without being given a ‘leading question’, so that then there is a benchmark to compare the results to.
3. One-tailed or a directional hypothesis.
4. There is no difference in the answer to the question about broken glass between the ‘smashed’ group and the ‘hit’ group.
5. The graph shows that the ‘stronger’ the leading question (smashed rather than hit) the more likely participants were to say that they had seen broken glass. Also, the probability table shows that the greater the effect of the leading question on participants (i.e. giving a higher speed estimate), the more likely they were to think there was broken glass.
6. It probably meant that they would pay more attention to the videos than they might do otherwise; e.g. they might, while watching, try to particularly take notice of the details of the cars, surroundings, the order of events and so on, just in case they were asked about them later.
7. This was to make sure there were no particular effects on the DV from watching the films in a particular order. For example, if people got better at estimating speed for each film, it would look as if the last film produced more accurate speed estimates, whereas it was just because the participants got better – this would be a confounding variable.
8. One sampling method that could have been used in this study was an opportunity sample of psychology students.
9. One strength is that the participants would have all wanted to take part and so they would not deliberately try and ruin or falsify the experiment – they would have been engaged and taking it all seriously. One weakness is that the sample is not representative of the public in general. Because they are all students they are all quite young and will have had little driving experience and might be less good at estimating speed than more experienced drivers.
10. One finding from experiment 1 is that the stronger the word in the target question, the higher the mean speed estimate, i.e. ‘smashed’ group estimated 40.8 mph compared to 31.8 for ‘contacted’.
11. One conclusion is that participants’ answers were affected by the wording of the questions. This suggests that eyewitness testimony may not be 100% accurate when leading or suggestive wording is used.
12. Loftus and Palmer wished to investigate whether the reason for the difference in speed estimates was due to an actual change in memory representation or just due to response bias (i.e. just responding in the way the question wording suggests, rather than a material change in memory.)

### Evaluating the study by Loftus and Palmer (page 37)

NB all answers should be contextualised.

### The research method

**Strength:** Variables can be controlled; in this case just the leading questions were varied to show the effect on recall. This means that cause and effect relationships can be demonstrated.

**Limitation:** It is a contrived situation and not like everyday life, where real eyewitnesses could feel anxious and this might affect recall.

### The sample

The US is an individualist culture which may be significant in willingness to be led by leading questions. Students are a group with unique characteristics – more intelligent people with good memories might be less likely to be affected by leading questions.

The students might be less affected than other people, which means the findings have greater validity – we might expect other people to be even more affected by leading questions. However, students would have less experience of driving and estimating car speeds so might be more affected by the leading questions.

They also might be less concerned about being accurate and more willing to try to please the experimenter. It does challenge the issue of generalising from this sample.

### Qualitative or quantitative?

The data was quantitative (estimate of speed, saying whether there was broken glass or not).  

**Strength:** It is easy to analyse such data to see how the leading questions affected speed estimates and recall.  

**Limitation:** The data may not represent what people really think because their choice of answer is so restricted, and this oversimplifies reality. In this study you also could have asked people to describe what they saw, to evaluate whether the questions had had any effect.
Exam-style questions (page 39)

Section A questions

1. The film clips might affect the ecological validity in two ways. Firstly, if participants are watching film clips they are expecting to take note of events; whereas in real life people who eyewitness accidents are not expecting something to happen which they might need to especially take note of. This means that in Loftus and Palmer’s study, the participants might be better at estimating speed than if they saw an accident in real life. Secondly, watching a video does not have the same emotional impact as seeing an accident in real life. In real life, the noise would be much louder and shocking, and the emotional impact – shock, fear, worry for passengers etc. – may affect what details participants can remember.

2. The first explanation was that it could be due to response bias. In other words, at the time of the experiment when participants are asked the key questions, it just encourages them to answer in a particular way (e.g. maybe slightly increase their speed estimate) rather than actually change their memory. The second explanation was that the participants’ memory had actually been altered – the leading question retrospectively changes their perception of the accident.

3. One conclusion is that memory is not a perfect, 100% accurate record of what a person witnessed and so this means that eyewitness testimony should not be used to convict people. Another conclusion is that memory is formed from two sources – information at the time of the event and information from after the event such as leading questions.

4. a. Two verbs were smashed and hit.
   b. Participants who had the word smashed gave a higher speed estimate than the ‘hit’ participants – mean of 40.8 mph compared to 34.0.

5. One procedure was that the candidates had exactly the same set of questions to answer after seeing the videos with the exception of the target question (smashed, hit, contacted, collided etc.). One other standardised procedure was that they all saw exactly the same events from the same stance – this is because the events were shown in video clips.

6. a. In experiment 1, the participants watched the videos and then the target question was ‘how fast were the cars going when they ***-ed?’ (smashed, hit etc.) and the DV was the mean of the speed estimate for each condition. In experiment 2, participants came back a week later after seeing the videos and being asked the question (conditions were smashed, hit and control). Here the DV was the response to the question: ‘did you see any broken glass?’
   b. A control group was shown the same video so that they knew the baseline rate of recalling broken glass (even though there was none) when no leading question had been asked. This means that then they could compare the rate for the ‘hit’ and the ‘smashed’ group and know if there was a significant difference.

7. In the first study, participants were more likely to give higher speed estimates if they had been asked a ‘stronger’ leading question (smashed rather than contacted, for example). Another finding is that when participants had the ‘stronger’ verb, ‘smashed’, they were more likely to recall having seen broken glass, even though there was none. This shows that the leading question had altered the memory of the event.

Section B questions

a. The aim of the Loftus and Palmer study was to find out whether leading questions cannot just influence how people answer – but influence their actual memory. Loftus and Palmer wanted to find out, whether, in these circumstances, people’s memories are totally accurate or whether they can be distorted by information after an event.

b. The sample used were 45 university students in the first study. In the second study, the sample were again students, but this time a new group of 150. Both samples were probably gained opportunistically and were just available and easy to recruit at the time. One limitation of this sample is that they are not like the general population in a number of ways – they are relatively young and similar in age, unlike the general public. Their memories, in some ways, may be better than average because they are younger (memory is said to decline with age) and because they are used to having to revise for exams and retain information.

c. Participants came to the laboratory in experiment 1 and were shown some videos of car crashes. They were then asked a number of questions immediately afterwards. The key question was ‘how fast were the cars going when they ***-ed?’ The verb here was the IV. Different groups of participants had different words – smashed, contacted, bumped, hit, collided. The participants estimated a speed in miles per hour and Loftus and Palmer calculated the mean for each group. For experiment 2, again participants were shown videos of cars and immediately afterwards asked the question ‘how fast were the cars going when they ***-ed?’ In this experiment, there were only two conditions – hit or smashed. There was also a control group who were not asked this question. Loftus and Palmer calculated the mean speed estimate (mph) for each group. But the main DV was measured when, one week later, participants returned to the laboratory and were asked ‘did you see any broken glass?’. They could answer either yes or no. Loftus and Palmer calculated the frequency of ‘Yes’ depending upon the condition of smashed, hit or control group.
Chapter 2 Cognitive psychology

Section C questions

a. One assumption of the cognitive approach is that people's brains process information a bit like a computer – there are inputs and outputs and the memory is like a storage system.

b. The cognitive approach would explain memory as a storage system which retains information which has been received. A bit like a computer memory, information is organised along fairly logical lines. When we 'remember' information or 'recall' information, it is retrieved from memory.

c. One similarity between Loftus and Palmer and Baron-Cohen is that both of them conducted their study in a laboratory under controlled conditions. In Loftus and Palmer, they showed them videos so as to ensure each participant saw exactly the same and so that any differences in the DV was due to the IV (verbs in the leading questions). Similarly, in Baron-Cohen, participants were shown specific pictures of faces to ensure high levels of control. One difference between Loftus and Palmer and Baron-Cohen is that Baron-Cohen used a diverse sample of autistic adults, adults suffering from Tourette syndrome as well as non-clinical (normal) adults. Therefore, they were quite a diverse set of participants. However, Loftus and Palmer used a more homogeneous sample – all students and probably all a very similar age.

d. One strength of the cognitive approach is that it focuses upon the mind and this should be central to the study of psychology. For example, in Loftus and Palmer, they have shown how the mind is not just a straightforward perfect record of events, that our minds are complex. According to Loftus and Palmer our memories are formed not only from the event itself but from information afterwards (such as leading questions) which can alter the original memory. This shows us that because our memories might be a more complex construction of the event, that eyewitness testimony may not be 100% accurate. Another strength of the cognitive approach is that it is one of the more scientific approaches. For example, in the Loftus and Palmer study, they conducted the research in a laboratory and had a high number of controls, e.g., the event was duplicated for each participant because they watched a video. Participants saw seven videos and these were deliberately seen in a different order to prevent any order effects from the sequence of the videos. All participants were asked exactly the same questions (except for the target question). Because the study is highly controlled and reduces the possibility of extraneous variables, we can be more certain the changes in the DV are just due to the IV rather than anything else. One weakness of the cognitive approach is that it ignores emotion. We do not simply guess what is going on in the mind (from measuring “outputs” or responses etc.) – and we may not be right. For example, in the Loftus and Palmer study, because we cannot directly observe memory, we cannot be 100% certain if they were changed or not. Even in experiment 2 when more of the ‘smashed’ group said they had seen glass, it may have just been that they were responding to demand characteristics or response bias from their original question. We cannot be 100% sure that the conclusions are actually correct.

Questions on ‘being sceptical’ (page 41)

No suggested answers are supplied for these questions because your answers will be highly individual. Use the examples in the text to help you formulate your answers, which will help you to understand the subject matter.
Core study 2: Baron-Cohen et al. (autism)

Qs (page 45)

1. Alternate hypotheses: Participants in the autism/Asperger’s group will gain significantly lower scores on the Eyes Task than participants in the other two groups (Tourette and normals).

Null hypothesis: There will be no difference in the scores on the Eyes Task when comparing the autism/Asperger’s groups with the Tourette group or the normal group.

2. ‘Age-matched’ means that the normal participants had the same age range and distribution of ages as the autism/Asperger’s group. This is a control so that any differences between the groups’ performance on the Eyes Task cannot be attributed to the age of the participants.

3. The Tourette participants were a good control group because, in some ways, they have some similarity of experience to the autism/Asperger’s group in that they too have a disorder which was identifiable in childhood and may have led to a disrupted education and problems with relationships with classmates. However, the actual symptoms are very different from autism. Therefore, in Baron-Cohen et al.’s study, if there is a difference in the Eyes Task between the autistic group and the Tourette group, it shows that this is something specific to autism itself, and not just a side effect of having had any developmental disorder or having had a tricky time in school. In other words, reading emotion from the eyes is problematic only for autism/Asperger’s people.

4. It is vital to use a control group in this study as a benchmark – so that it is possible to know what ‘normal’ performance on the Eyes Task is, so that the autism/Asperger’s group can be compared to the benchmark to see if there is a significant difference.

5. Control tasks were important so that Baron-Cohen et al. could be sure that the differences between the groups in performance on the Eyes Task were specifically to do with reading the emotion (i.e. ToM) and not anything else. For example, the control task of Gender Recognition of the eyes was there to make sure that any problems in the main Eyes Task wasn’t due to a more general difficulty of face perception.

6. The Gender Recognition task was to make sure that if participants did not do well on the main Eyes Task, it was not because they had a more general problem with face perception but because of problems specifically with ToM. The Basic Emotion Recognition task, in the same way, is to rule out difficulties with basic emotion recognition of the whole face.

7. Group 1 only did the control tasks because Baron-Cohen et al. were most interested in the autism/Asperger’s group and they were (correctly) predicted to do worse on the Eyes Task. It is important that explanations for this poor performance can focus specifically on impaired ToM, and rule out more general explanations. The other groups (normals and Tourette) were not the focus of the study and were predicted to do relatively well on the Eyes Task so there was no need to compare them to a control.

8. This was to counterbalance any order effects – for example, if one task such as the Eyes Task was always taken by participants last of all, they may do worse on it just because it is last and they are tired. Therefore, it is important that the order is balanced out among all the participants to control for such order effects.

9. You will see 25 pictures of the eye area of people’s faces for 3 seconds each. With each picture, you will be given a choice of two emotions, e.g. happy or sad. Identify which emotion matches the expression in the picture.

10. To ensure that it was not any easier or more difficult to correctly identify the emotion in any particular picture.

11. In Happé’s Strange Stories, participants read a story and are then asked questions checking their comprehension of the story, the key question being about a character’s emotion. This is a way of validating the results on the Eyes Task – checking whether the Eyes Task is measuring the same thing as the Strange Stories task. The scores from the two tasks are correlated and if there is a strong positive correlation, it means that they are probably testing the same thing. This is a way of checking the ‘concurrent validity’ of the Eyes Task.

12. This is a natural experiment (or quasi-experiment). It is an experiment because there is an IV with different conditions (autism/normal/Tourette) and a DV. It is a natural experiment because these conditions are not directly manipulated by the researcher (you cannot make someone suffer from Tourette for a day!), but occur naturally.

Evaluating the study by Baron-Cohen et al. (page 47)

NB all answers should be contextualised.

The research method

Strength: High degree of control, e.g. over the eyes shown, means the findings are more likely to be valid.

Limitation: The fact that the IV was not deliberately manipulated means that you can’t claim that autism causes the lack of theory of mind; there may be intervening factors.

The sample

There were only 16 people in the autism/Asperger’s sample which is not very many and the smaller the sample, the greater risk that they might not be representative of the target population as a whole. However, there were more males than females (this is true of the autistic population generally) so it is probably a reasonable sample.

Individual differences

It is not the case that all the autistic adults scored worse than all the normals, which you might expect. For example, the score range for the autism/Asperger’s group was 13–23, while the score range for the normal group was 16–25. This means there was quite a lot of overlap in scores between the two groups. It means that not all autistic adults have a deficit in their ability to read emotion from faces … or … It might mean that some normal people have some problems in reading emotion from faces.

Qualitative and quantitative

The data is mostly quantitative – this is because the data collected was the participants’ scores for the Eyes Task. The strength of this is that you can easily compare the performance of each of the groups of the task and see if there are differences in the ability to read emotions. One limitation is that it may not give a full picture of why there are differences. Without collecting qualitative data, we cannot be absolutely sure of the reason why the autism/Asperger’s group did not do so well overall. If participants had been asked about why they had given one answer rather than another, it may be possible that in some cases it was not due to misreading the emotion, but to problems with the vocabulary used.

In terms of qualitative data, many of the participants with autism/Asperger’s reported that they found it very hard to comprehend what was going on in a movie – they couldn’t work out who knows what and who doesn’t, and why people laugh at particular points. To extend this part of the study and make it more experimental, participants could be shown two movies: an action movie (where the actors’ emotions and motivations are probably not so crucial to the plot development) and a romantic movie/comedy (where the actors’ emotions, thoughts and motivations are probably crucial to the plot development).
In this study, Baron-Cohen found that there was no difference between the groups. 1 and 3 (autistic versus Tourette) on the Gender Recognition task.

One similarity was that the autism/Asperger's participants performed the same as the control groups on the gender and emotional control tasks. One difference was that the autism/Asperger's had a lower mean score on the Eyes Task than the normal group (16.3 versus 20.3).

In the Eyes Task, participants were shown 25 pictures of the eye area of faces. Each set of eyes was displayed for just 3 seconds and then participants were asked to choose an emotion from two possibilities (e.g. happy versus sad) that were presented.

Another way to test Theory of Mind would be the Sally- Anne Task. There are two dolls, Sally and Anne and they are made to move around etc. by the researcher. Sally puts a ball in her basket and walks away. Anne gets the ball from the basket and puts it in a box. Sally comes back into the room. Participants are then asked ‘Where will Sally look for her ball?’ Autistic children generally say ‘In the box’ as they cannot appreciate (due to lack of ToM) that Sally will not ‘know’ that Anne has moved the ball.

‘Ceiling effects’ occur because a test is designed for a certain group or ‘ability’ of people; when people well above that ability take the test, the test will not discriminate (show individual differences) between them. The measuring tool does not ‘go high enough’.

‘Ecological validity’ refers to whether the procedures in the study will mean the participant has a normal, everyday experience and therefore the results can be generalised to everyday. A study with low ecological validity would be one where the participant had unrealistic scenarios or tasks etc.

This study could be said to lack ecological validity because usually we are only motivated to read the emotions or mental states of people who we can glimpse for much longer than 3 seconds, and generally we can use their whole face to read emotion and not just the eyes.

Finally, in everyday life, we do not just decide someone's emotion from a possible list of two, but have to think about all possible emotions – probably hundreds.
Section C questions

See exam questions and answers in the book on page 68.

Qs (page 51)

1. One core study that has both qualitative and quantitative data is Dement and Kleitman. For quantitative data, they collected data on, for example, the length of the REM period, the duration of sleep, the number of wakings, the number of times participants said they were dreaming during REM and NREM, and the number of words used to describe the dream. For qualitative data, they collected the descriptions of the dreams that the participants described into a microphone.

2. The strength of the quantitative approach in this study is that Dement and Kleitman could easily make comparisons to help test their hypotheses, so they could compare whether there were more times when dreaming was reported in REM than in NREM and whether the length of REM matched the reported duration of dreaming etc. This meant that the hypotheses in this study could be accepted because the differences were statistically significant. One limitation of the qualitative data is that it may hide or mask important events or the reasons behind a number. For example, in this study, with only the quantitative data it would not be possible to know whether the dreams reported in NREM were of the same ‘quality’ or ‘type’ as those reported in REM.

3. The strength of the qualitative approach in this study is that you can understand the data in more depth. For example, Dement and Kleitman found that the dreams reported in NREM seemed to be more vague than those reported in REM – which led them to believe that the NREM dreams were really just memories of REM dreams. One limitation of the qualitative data is that it is difficult to analyse objectively and interpret. For example, when Dement and Kleitman looked at whether the dreams matched the pattern of eye movements, they may have been interpreting the description of the dreams to match their own expectations.
Suggested Answers

Chapter 2 Cognitive psychology

Study 3: Savage-Rumbaugh et al. (animal language)

Qs (page 55)

1. The main difference is between Kanzi, Mulika and Matata versus Sherman and Austin. Kanzi, Mulika and Matata were pygmy chimps, while Sherman and Austin were common chimps. The main focus of the study is on Kanzi and Mulika; while Sherman and Austin were really part of an earlier study and just for comparison purposes. Sherman and Austin were overtly trained how to use lexigrams, while the others were not. Instead, Kanzi and Mulika had lexigram use ‘modelled’ by humans and their mother Matata. The lexigrams were matched by what the ‘teachers’ said. At the end of the study, there were differences between Kanzi and Mulika on the one hand and Sherman and Austin on the other – Kanzi and Mulika acquired the language with greater ease; they understood spoken English, they used lexigrams in a more specific way (e.g. could use them to differentiate between coke and juice).

2. They used ‘lexigrams’, colourful arbitrary symbols which were presented either by computer or on a board. The indicated their choice of lexigram by pointing to it.

3. The researchers communicated with the chimpanzees in two ways at the same time, (1) they would say it out loud and (2) point to the relevant symbols on the lexigram too.

4. They used structured questions (rather than just open questions or open ‘conversation’) to see if the chimps could give a specific answer; the questions were ‘structured insofar as the researchers knew the answer they were seeking and were asking the question to see if the chimpanzee knew the answer.

5. ‘Behavioural verification’ indicates when a chimp communicated something in lexigrams and then verified that they really meant this (and had not just pointed randomly) by actually going and doing it. ‘Concordance score’ is whether the chimp’s behaviour matched what he or she had ‘said’ on the lexigram.

6. Inter-rater reliability was checked by comparing two sets of ratings from a period of 4.5 hours of the same behaviour – one set was made during real-time, the other was made of the same behaviour but which had been video recorded.

7. There was 100% agreement on which lexigrams were used and whether or not they were correct. There was just one disagreement about whether it was spontaneous or not, indicating high reliability.

8. The formal tests were necessary because Savage-Rumbaugh et al. wanted to make sure that the chimps’ performance was not due to inadvertent cues or glances (as was the case with Clever Hans described on page 53 of the book). The formal tests were of two types. One type involved the chimps being show photographs of familiar things and being asked to select the lexigram which corresponded with the photo. This was a test of their productive capacities. The other test involved listening to a word said out loud, either by a human or a synthesised voice. Again, to show that they had understood, they were asked to indicate the right photograph or lexigram. This was a test of receptive capacity. The reason for these formal controlled tests was to ensure that the chimps really did know and understand the vocabulary and that weren’t (like Clever Hans) just responding to context or inadvertent clues.

Evaluating the study by Savage-Rumbaugh et al. (page 57)

The research method

As a case study, we learn about the chimps, Kanzi in particular, in a wealth of detail. This means that we get a really thorough understanding of how Kanzi acquired language and the sort of language he used and in what contexts. One limitation of the study is that it does not necessarily help us understand whether other pygmy chimps (or other chimps of other species) could acquire language to the same degree with the same sort of attention from researchers.

One advantage of an observation study is that Kanzi’s actual behaviour was recorded as it happened in realistic and natural everyday (for him) occasions. This means that, on the whole, he was not put under pressure or only tested artificially – mainly how he normally communicated in his everyday life. One disadvantage is that such open ended observation must have produced a huge amount of data over the course of the study which is difficult to analyse and make sense of.

The sample

The size of the sample means that we should be cautious about what conclusions we can draw. It may not be correct to infer that all pygmy chimps could acquire sign usage to the same degree; or that they are better than common chimps. A much larger sample of each would have to be tested.

Representativeness

It may be that Kanzi is a particularly communicative and bright chimp who is switched on to language – just as there are differences in the rate of acquisition of vocabulary in human infants. We do not know whether Kanzi is average, above average, or even below average!

Qualitative or quantitative

Both types of data were used in this study. The advantage of the quantitative data is that you can chart exactly how Kanzi acquired the lexigrams and vocabulary, the scores on the formal tests and compare these with Austin and Sherman. This all makes the data more objective and helps to summarise the findings so they are easy to understand. The advantage of the qualitative data is that you get a rich understanding of Kanzi’s abilities, descriptions of the contexts in which he would use a word (e.g. ‘strawberries’ was used only in the mushroom site at first, but then could be used in any place) and so we can understand more about the process and complexity of acquiring language in chimps. The qualitative data gives us much more insight.

Ethical issues

Is it ethical for great apes to learn a language? The same ethical issues do not apply to non-human animals. So, issues such as informed consent, deception, right to withdraw and so on are not an issue according to ethical guidelines and codes. And, great apes have the same ethical principals applied to them as do rats and mice. This seems a bit odd, as from our understanding of Kanzi, as well as some of the abilities he acquired, it seems possible that he could be ‘deceived’ and might get upset about it. It might be possible to ask his permission or consent to do something. Therefore, it starts to feel uncomfortable that really he is just a very large and intelligent ‘guinea pig’ in one long experiment ... which he did not ask or volunteer to take part in. So what are the advantages of this kind of study? Perhaps Kanzi ‘feels’ more fulfilled being able to communicate with humans (there is evidence he was always interested in humans and human interaction); also, from a scientific point of view, we learn that, maybe, we as humans might not be quite as special as we thought – human language is not just the preserve of humans – other creatures have sufficient intelligence to acquire human language. The costs are that Kanzi is now not a normal chimpanzee and can never lead a normal chimp life again and will be alienated from other pygmy chimps. Given that chimps live until around the age of 50, that is a long time to ensure that Kanzi has a decent quality of life.
Nature or nurture?

It is more than likely that, as humans, we have a strong ‘language instinct’ – that we are predisposed to acquire language, and, providing that we are exposed to language, we will do so. We are highly related to pygmy chimps who are social animals, and, given that we probably share some evolutionary history with the great apes, it is not altogether surprising that at least one pygmy chimp seems to show some ability with language acquisition. Therefore this study underlines one aspect of similarity to the great apes.

Applications/usefulness

This is not really a useful study. We cannot improve the lives of humans – or chimps – through knowing they can acquire language. However, it might help humans to reflect upon what rights great apes should have and underlines how important it is that we do not rob them of their habitat etc. It perhaps also raises issues about whether great apes should ever be involved in medical experiments.

What next?

On a much larger scale, it would be good if a much larger number of pygmy chimps, e.g. at least 20, could have the same ‘training’ in the use of lexigrams and exposure to spoken human language. This would help determine how generalisable this study would be to other pygmy chimps and to our understanding of human/great ape differences.

Exam-style questions (page 59)

Section A questions

1. One method was that when the chimpanzees were inside, the computer would automatically record the lexigrams used. The other was when the chimpanzees were outside, the researchers had to make a note of which lexigrams were used.

2. One difference was that Kanzi and Mulika acquired an understanding of spoken English, which had not been seen in chimps before. Although they could not produce it themselves, they could comprehend it and respond appropriately. Austin and Sherman did not acquire this. Another difference was the specificity of use of language. Austin and Sherman often used words quite generally, e.g. juice might be used interchangeably with coke, and not used to refer to specifically different things. Even when they were taught to differentiate, they were inclined to drift back into using them generally. Kanzi and Mulika, on the other hand, could use words more specifically.

3. a. Kanzi and Mulika were exposed in a more ‘natural’ fashion – like human babies, they were exposed through observation and natural communication, rather than being explicitly trained. So the ‘teachers’ would say something out loud and then translate that utterance into lexigrams. Kanzi seemed to acquire a new term in an associative context first (i.e. a particular environment) and then later be able to use it in a context free situation. This stages process is similar to human infants language acquisition (Lock’s two steps – see page 54 in the book).

b. It is difficult to tell, but it might make Kanzi and Mulika more inclined to communicate naturally, rather than just communicating for treats or tickles.

c. Syntax is the rules of grammar which govern word order and sentence structure so that the meaning is clear.

4. a. Kanzi and Mulika were exposed in a more ‘natural’ fashion – like human babies, they were exposed through observation and natural communication, rather than being explicitly trained. So the ‘teachers’ would say something out loud and then translate that utterance into lexigrams. Kanzi seemed to acquire a new term in an associative context first (i.e. a particular environment) and then later be able to use it in a context free situation. This two stage process is similar to human infants language acquisition (Lock’s two steps – see page 54 in the book).

b. Formal tests were necessary to make sure that the chimps were not just appearing to use language with meaning when really they were responding to inadvertent cues.

c. One formal test involved being shown photographs of objects and being asked to select the right lexigram – so the test was context free.

7. True comprehension was checked by ‘behaviour verification’. If Kanzi said something using a lexigram, the researchers would note whether this then matched the following behaviour (e.g. select a specific object from a group when asked to) and were also able to do it without any specific training – just by a process similar to that of natural language acquisition in human infants.

The sample consisted of five chimps, though not all were studied in equal depth. There were three pygmy chimps, Kanzi, his younger sister Mulika and their mother Matata. Also, there were two common chimps, Sherman and Austin. These two were just used in the report for the purpose of comparison and had been the focus of a previous attempt to teach language. At the beginning of the study, Kanzi (the main object of the study) was about 30 months, while Mulika was about 11 months. They both preferred human company but did spend time with their mother, Matata. Pygmy chimps are rarer than common chimps, and they were probably chosen as the focus of this study because pygmy chimps appear to have more human characteristics than other chimp species – they are more social and have more social skills such as food sharing. This might make them better candidates for learning a language, as language is an essentially social tool (the need to communicate with others!)

The data was collected in many ways. First of all, all utterances using a lexigram were recorded (inside on computer, the data was captured automatically – while outside involved having to take notes). Altogether, this meant there was a complete record of Kanzi’s lexigram usage from 30 to 47 months and for Mulika from 11 to 21 months. All of this data was coded as either correct or incorrect; as spontaneous, imitated (if a companion had included it in any part of a previous utterance) or structured (produced in response to a particular request or question). Therefore, a lot of data was collected about the context of the utterances such as where and in response to what and so on. They also collected data on what the chimps’ behaviour was after an utterance, and whether this matched the utterance. This was called ‘behaviour verification’ and was the way that Savage Rumbaugh et al. could check whether the word was being used with true comprehension or just randomly or coincidentally. They also collected data from formal tests (not natural interaction as above). Here, the chimps would be given a photograph and asked to indicate the correct lexigram. Or, they would hear a synthesised voice say a word (so no cues could be given from intonation etc) and the chimps would have to indicate the correct lexigram.

b. SYNTAX is the rules of grammar which govern word order and sentence structure so that the meaning is clear.
e. The conclusions drawn from the study were that pygmy chimps are more naturally inclined to learn language than common chimps. Chimps are also capable of understanding spoken human language. Another conclusion was that the way in which the pygmy chimps acquired the language was not dissimilar to how children learn by associative learning first (very context dependent) and then representational learning – using the sounds to refer to the objects. Other conclusions were that Kanzi showed that he could combine signs to create novel utterances and that these had syntax – so he could tell the difference between Matata hugged Mulika versus Mulika hugged Matata.

f. One change that could be made to this study would be to raise one pygmy chimp and one common chimp at the same time in the same way and give them the same opportunities to learn language. In the study as it is, Kanzi seemed to get much more natural opportunities to converse than Sherman and Austin. For example, they could both be spoken to in the same way, introduced to the lexigram at the same time; both could be present when conversation between the ‘teachers’ was translated onto the lexigram etc. It is difficult to say what impact this might have on the results. It still may end up being the case that the pygmy chimp acquired language faster and better than the common chimp – however, one impact of this change would be that, because both species would be reared side-by-side, we would be able to be more convinced and certain that differences in language acquisition were due to the species, and not just due to differences in treatment and language exposure.

Another change could be to use a portable touch screen laptop. In the current study, the researchers had to take a board with the lexigrams on and make notes of which lexigrams the chimps used when they went outside into the more natural habitat. This way, the laptop will automatically record e.g. Kanzi’s choice of lexigram, so that the researchers do not have to painstakingly record what was said, which may have got in the way of a natural ‘conversation’. With this change, it is likely that the validity of the recordings will increase as it will be less prone to error from missing things out, mis-remembering what lexigrams were signalled. It might also mean that Kanzi’s language improved as the researcher would have more time to converse and expose Kanzi to language, rather than busily taking notes.

Section C questions

a. Quantitative data is data collected in the form of numbers, while qualitative data is in the form of words and descriptions. In the Savage-Rumbaugh et al. study, quantitative data includes the scores on the formal tests (e.g. how many words a chimp could recognise when spoken and point to the correct lexigram; or produce correct lexigram when shown a picture of the item), the number of words acquired through the course of the study (e.g. Kanzi acquired 46 words), the number of word-combinations produced and how many of these were novel, the percentage of words which were just imitative and which were spontaneous.

Qualitative data involves descriptions of e.g. what sorts of situations the chimps used language in, such as when they only used a word in a particular place; if they used a word specifically or overgeneralised the meaning; also, which actual word combinations were used. Also, there are descriptions of how the chimps responded to the formal test – e.g. Kanzi and Mulika liked them from the start, but Austin and Sherman did not quite get the point of them and thought that if they saw a photo of a banana and got the correct lexigram that they would get a banana. This helps us understand the chimps more.

b. The advantage of the quantitative data is that you can chart exactly how Kanzi acquired the lexigrams and vocabulary, the scores on the formal tests and compare these with Austin and Sherman. This helps to summarise the findings and make comparisons so the findings are easy to understand. Another advantage is that quantitative data is usually more objective. This means that the results are less prone to researcher bias or misinterpretation. For example, the data from the formal tests, where there are clearly right and wrong responses, is very objective and helps us be sure whether or not the chimps had really acquired the signs by counting up the numbers of correct and incorrect responses. These can then be compared between each chimp or even to human children.

One disadvantage is that quantitative data can be a bit reductionist and overly simplistic – so we do not necessarily really understand what is going on. For example, without the qualitative data, we would not understand in what situations or contexts a word might be acquired or that often a word might be produced to refer only to a specific item in a specific context before it was used in a more representational or flexible way. Therefore, it can oversimplify data.

Another disadvantage of quantitative data is that, if researchers are only collecting quantitative data, it gives the researchers and the participants less freedom and constrains the data only to collect what they expect. So, for example, if the researchers had only counted new words and whether they could be recognised in the formal tests, there would be no freedom to collect the more interesting stuff on the word combinations or on when Kanzi started to acquire lexigrams when his mother went away etc. – which has given us a lot of understanding. This means that purely quantitative studies may miss some things and have less opportunity to discover unexpected and interesting facts.

c. One similarity to Thigpen and Cleckley is that they both use a case study and focus on a small number of individuals. Thigpen concentrated on Eve White and her alter-personalities while this study concentrates mainly on Kanzi, but also on Mulika, Austin and Sherman. So both collect rich, highly individual but in-depth information. One difference is that Thigpen and Cleckley only looked at one individual (with Multiple Personality Syndrome) while Savage-Rumbaugh et al. look at more than one. This means that the Thigpen study does not give us any clues about how similar or different Eve’s experiences and the interrelationships of the personalities are when compared to other Multiple Personality sufferers.

Meanwhile, because Savage-Rumbaugh et al. had collected data on two common chimps (Austin and Sherman) and two other pygmy chimps (Mulika and Matata) we can at least begin to understand whether Kanzi is typical or atypical of chimps learning a language.

Qs (page 61)

1. a. Cats associate the can opener with being fed. This means, like Pavlov’s dogs, they associate an unconditioned stimulus (food) with a neutral stimulus (here, the sound of the tin opener), and so after a while of repeated pairings, the tin opener gets the same reaction (conditioned response – here, running towards you) as just the food on its own. Or, food could be seen as a positive reinforcer and so they learn that running when hearing the tin opener is often rewarded with food. Behaviour which is positively reinforced once or on several occasions means that the behaviour is more likely to be repeated in the future.

b. Being shouted at might actually be a positive reinforcer – because at least it is attention. Therefore, a child might repeat naughty behaviour because at least someone takes notice of them and stops ignoring them.

2. Which clothes make you look nice or suit you (get compliments).

b. How to use a simple object for the first time – e.g. switch on a tap – the reinforcer here would be that you get the water you wanted.

b. How to shave with a razor – getting it right is positively reinforcing because it has the desired result of removing hair, but negative reinforcement also comes in here. If you cut yourself shaving you will learn how to avoid this by holding or manipulating the razor.
3. It would be difficult to acquire complex behaviours such as how to write and construct a good essay. There may be some shaping through encouragement, the right sort of feedback from teachers, increasing marks etc. However, the content of an essay, and how it is structured involve different kinds of learning, cognitive schemas, good motivation to get it right.

4. How to form good relationships cannot just be learned through simple rewards and punishments. This is because they are very complex — good relationships take a long time to build up, involve complicated things such as being considerate and understanding. There may be long-term rewards from relationships, but these may not be simple!

5. Conditioning cannot explain all our behaviour. Behaviour is learnt and acquired and produced for many reasons — conditioning being just one way. For example, we learn how to do lots of things just by watching other people (social learning theory) which saves us the time of learning how to do it by trial and error. Also, many behaviours are motivated by other things (and not just reinforcers), e.g. a desire to improve yourself, reach your potential, know more etc. (and this is a humanistic view). Some behaviours are also biologically motivated, e.g. hormones and neurotransmitters have a role in, e.g. making someone behave in a moody fashion, feel sleepy, excited, aggressive and so on.

5. Bandura uses an extension to behaviourism — social learning theory. Griffiths does not ‘use’ behaviourism but the rewards of ‘winning money’ can be seen as having a reinforcing effect for some of the gamblers. Reicher and Haslam’s guards may have tried to use behavioural techniques to get prisoners to behave — punishment etc. In Milgram, one reason the obedient participants continue to administer electric shocks is because they do not want to let down the experiment — this could be seen as a form of negative reinforcement. In Piliavin et al., the model of bystander intervention can be seen to include some aspects of behaviourism. For example, we are motivated to reduce our feelings of stress/anxiety/arousal through acting in some way such as helping the victim or walking away — again this is negative reinforcement. Also, in this model, the cost-reward matrix shows that we are more likely to act in a way which we think will maximise rewards (positive reinforcers) and minimise costs (punishment).