5 The DISCUSSION section

Diagnostic questions on design for Chapter 5

Before reading this chapter, check that you can answer correctly the following diagnostic questions:

1 What is the *internal validity* of an experiment?

2 What is the *external validity* of an experiment?

3 What is a *confounding variable*?

4 What is a *type II error*?

You can find the answers at the rear of the book (p. 275) If you had difficulties answering any of these questions, turn to the relevant sections of Part 2.

Researchers design experiments to test hypotheses to develop theories or to help solve practical problems. However, in psychology at least, the outcomes of experiments are *not* self-evident. That is, they require *interpretation*. It may not be apparent when you first start practical work, but we need to "discover" what we have found out. For instance, in the cheese and nightmare experiment (Section 4.3) we have been compelled by our data to reject the null hypothesis. Now,

however, the real work begins. For we need to find out what *caused* the apparent link we have found between condition and the incidence of nightmares. Was it the IV (eating cheese)? Or are there *confounding* variables that could also account for this effect? The answers to such questions we can only arrive at by looking carefully at our experimental design and assessing the quality of the *control* we exercised. Only once we have done this can we decide (for it *is* a decision) which of the various options is the *likeliest* explanation for our findings. Only once we have done this will we be able to explore the *implications* of our findings for the area under consideration.

Moreover, this remains true of situations in which the data compel us *not* to reject the null hypothesis. Does this mean that the IV is *not* the causal variable after all? Or was there some aspect of the design that made it hard to detect the effect of the IV on the DV? Again, we can only decide this after a close examination of the experiment. Again, we need to sort this out before we can even begin to assess the implications of the findings for the area discussed in the INTRODUCTION.

The point is that our findings are subject to *uncertainty*. We need to interpret our results. This process is undertaken in the DISCUSSION.

There are three definable stages to the process of discussing the findings. These three stages, in sequence, provide the structure of your DISCUSSION. First, you need to say what needs to be explained. So the first task in the DISCUSSION is to state what the results of the experiment are by providing a précis (in words) of the RESULTS and describing how well they fit the original predictions. Once you have done this, you need to try to *explain* these findings. Is the IV responsible for statistically significant differences? If not, what is? Does the lack of statistical significance mean that the IV does not cause changes in the DV? The second task in the DISCUSSION, therefore, is to explore these issues, arriving if you can at a set of reasonable conclusions. Finally, you need to explore the theoretical implications of the findings, together with any practical ones they may have. For instance, if you have concluded that the IV was indeed responsible for the changes detected in the DV, what implications does this have for the material – especially the arguments and theories – that you outlined in the INTRODUCTION? This, of course, is the key issue – the reason why we design and run experiments in the first place.

We can crystallize these three phases of the DISCUSSION around three questions:

1 How well do the findings fit the predictions?
2 What do the findings mean? (What, if anything, can you conclude about the relationship between the IV(s) and DV(s)?)
3 What are their implications (especially with regard to the issues that you outlined in your INTRODUCTION)?

Of course, the balance between these three phases, and the depth to which you will be expected to follow some of the issues raised by your findings, will depend on your experience as a student of practical psychology. As a novice, considerably less will be expected of you here. Your main task will be to state your findings, outline and *assess* the more plausible explanations and, in the light of this, arrive at reasonable conclusions. If you have something sensible to say about how the experiment might have been improved, then consider also including this. In particular, remember that you should be aiming to display clarity of thought and presentation when discussing the outcomes of your experiment. Demonstrate that you understand what you have done and found.

Before we go through the phases themselves, two additional points need to be made. First, if the findings of experiments are subject to uncertainty, you can imagine that findings from studies that are not experimental are even more problematic, as these methods provide us with even less control. Second, science proceeds by argument about evidence. It is important to remember that other people may disagree with your interpretation of your findings. Your task in this section, therefore, is to argue the best case that you can, given your data, and without going beyond the data.

Summary

1 The outcomes of psychology experiments are subject to uncertainty. They require interpretation in order to find out what they have to tell us about whether or not there is a causal relationship between the IV and the DV.

2 This process of assessment and interpretation takes place in three phases in the DISCUSSION. These phases revolve around three questions: (1) What are the findings of this study? (2) What do they mean? (3) What are their implications?

5.1 How well do the findings fit the predictions?

Open the DISCUSSION by describing what you have found and how well the data fit the predictions. Doing this enables you to be clear from the outset about what needs to be explained in the DISCUSSION. All of you, therefore, regardless of your level of experience, will need to do this adequately. Where you have lots of findings to report, prioritize them in the same way as you did in the RESULTS (Section 4.3).

5.2 What do the findings mean?

Once you have decided what you found, you can set about the *process* of drawing the conclusions from the experiment. If you have *rejected* the null hypothesis, then you need to assess whether this was indeed due to the manipulation of the IV or whether something else is responsible. If you have *failed* to reject the null hypothesis, then how confidently can you conclude that this indicates the absence of a causal relationship between the IV and DV?

In order to answer these questions as well as you can, you need to sit down and re-examine the experiment. Have you become aware of a confounding variable that you failed to spot before running the experiment? How likely is this variable to account for the findings? In retrospect, was the design likely to have been able to allow you to reject the null hypothesis in the first place? It is only when you have satisfied yourself that there are no confounding variables or that your design had sufficient power (see below) that you can draw reasonable and firm conclusions.

So, how long this next bit is and how firm are your conclusions will depend on how well designed and executed your experiment was. Before starting on the DISCUSSION, therefore, take a long and hard look at your experiment, identifying its weaknesses but recognizing also its strengths. Once you have listed (to yourself) any weaknesses, then *assess* each of these. Is it a feature that undermines your capacity to attribute changes in the DV to the manipulation of the IV (*internal* validity) or does it instead compromise the extent to which you can generalize the findings (*external* validity)? Do NOT confuse these (see Sections 10.8 and 10.9); at this stage of the DISCUSSION we are concerned with things that affect internal validity. Remember that

a *confounding* variable is not any old variable you failed to control by holding constant, but one that systematically co-varies with the levels of the IV. So, is the variable that you are concerned about really a confound? If not, it relates to external rather than internal validity.

Once you have done this, decide which (if any) of these features are sufficiently important to require raising and discussing explicitly in this phase of the DISCUSSION. (Your ability to judge this as well as your ability to spot important design problems is being assessed here.) Be honest, but also be sensible – novices especially. Obviously, it is important not to blind yourself to flaws and weaknesses in the design that undermine the conclusions that you can draw. You will not gain marks for pretending that all is well when it is not. However, the solution is not to assume automatically that there are problems and merely catalogue every problem you can dream up, regardless of importance. It can be tedious to read a mindless list of flaws and problems, doubly so when the experiment is in fact pretty sound and could sustain reasonable conclusions. Really watch out for this, *especially where you have failed to reject the null hypothesis*. Your marker wants to see evidence of critical engagement; you need to strike a balance between spotting and discussing important flaws and drawing reasonable conclusions from what you did and found. Look at journal articles to get a sense of how the professionals do this. A good journal discussion will contain an account of the study's limitations alongside sound (justifiable) conclusions.

With findings that are not statistically significant, are there features of the experiment that undermine your capacity to detect the influence of the IV on the DV? Failing to attain statistical significance may tell you more about the *power* of your experiment than about the existence or otherwise of an effect of the IV on the DV. If you are a novice, or otherwise not yet technically able to discuss power, you should at least consider whether you had enough participants – especially on unrelated samples IVs. Please note, however, that this is definitely *not* a licence to mindlessly trot out the old cliché that "we would have obtained significant results if we had had more participants". This is inevitably true – any effect, however trivial, will become statistically significant if you have enough participants. The question to address is whether you might reasonably have expected to obtain statistical significance given the number of participants in your study. You can find advice on this issue in Sections 12.3 and Section 13.1.1.

If you understand about power, then use any relevant confidence intervals you reported in the RESULTS to reflect on the likely power of your study where you have failed to reject the null hypothesis. Comment also on any relevant statistics of effect size that you have reported there. (Remember, it is possible to estimate effect size whether or not the results are statistically significant.) You can find advice on this in Section 12.3.1.

If your findings are unexpected, especially if they contradict established findings or ideas in the area, be careful. Under these circumstances, many of you lapse into what I call the "chemistry experiment" mentality. That is, when the results do not come out as predicted, you assume that the experiment has not "worked" and search for where you went "wrong" – just like we used to when our test-tubes of noxious substances failed to behave as anticipated in chemistry classes (which, as I recall, was most of the time). This is *not* the way to proceed. It may well be that a flaw in your experimental design has produced "anomalous" results. However, you should not automatically assume that this is the case. As a general rule, *believe* your findings until you discover a feature of your design or procedure that casts doubt upon their validity.

On the other hand, the fact that your findings fit your hypotheses should not blind you to problems. Confounding variables can be lurking anywhere. Check your design for alternative explanations for your findings, even when the explanation that you favour seems so obvious to you as to be undeniably true. (Such as when your findings fit your predictions.) Watch out for this – it is a very common and easy mistake to make.

So, it is a question of balance. Do not jump to conclusions, but do not dwell on every flaw.

If your experiment has been reasonably well designed, at the end of this process you should be in a position to say how likely it is that any difference between your conditions was caused by your manipulation of the IV, or that the absence of such a difference suggests that there is no causal relationship between the IV and DV. This is, of course, what the whole enterprise is about. This is why you spend so much time in the design phase working on such things as your controls for confounding variables. So, if at the end of this phase of your DISCUSSION you are unable to decide about the impact of your IV on your DV with *reasonable* confidence, there are obviously design lessons to be learnt. So learn them! (If at this stage your findings remain equivocal, one question that you might address is whether you adequately pilot-tested your experiment; see Section 13.9.3.)

5.3 What are the implications of these findings?

The first two phases of the DISCUSSION really are preliminaries for what is to follow. Having decided what you have to explain, and having drawn sensible and balanced conclusions about how best to explain it, you can now get to the nitty-gritty – assessing the implications of the experiment for the work outlined in the INTRODUCTION. What, if anything, have we learned about the IV from the experiment? Does this advance our ideas in any way, or at least qualify them? To what extent can the findings be reconciled with the theoretical ideas discussed in the INTRODUCTION? Indeed, do they enable you to decide between any competing theories?

A good DISCUSSION, therefore, depends upon an adequate INTRODUCTION. Essentially, your task now is to return to the material that you addressed in your INTRODUCTION with the benefit of the findings of your study. So, what you are able to say at this crucial stage of your report depends critically on how well you prepared the ground. Indeed, unless something particularly unforeseeable has occurred, you should find yourself addressing the same themes here as in your INTRODUCTION, albeit with additional knowledge. In general, there should be no need to introduce new evidence from the psychological literature to this section.

Once you have done this, you should think about the direction that future work might take. At this stage you might suggest problems that need to be addressed next and, if possible, ways in which this might be done. However, be sensible about this. Think about further work that you might reasonably do yourself if you had the time – sensible next steps in the process of exploring the causal relationships that you have been investigating. Make positive and constructive suggestions. Never write "there is a need for further research" without indicating something about what form this further research should take.

This is all very well, of course, if your findings are relatively unambiguous. However, what if it has not been possible to say much at all about the relationship between the IV and DV? If this has occurred because of design flaws, then you should go some way towards improving your reputation as a designer of studies here by indicating ways in which future experiments on the same topic might overcome the difficulties that you encountered.

5.4 What to do when you have been unable to analyse your data properly

Sometimes, however, these problems will have arisen because you experienced considerable difficulties with your study, difficulties (such as very few participants) that render your data effectively unanalysable. This, of course, will make writing the DISCUSSION that much harder. Whenever possible, avoid such circumstances *before* running the study. However, such things can occasionally happen even when the study has been well thought out in advance. In such instances, when it comes to writing the report, the important thing is to *be seen to have made an effort*.

Do not, therefore, jump at the opportunity (as you see it) of only having to write a brief and dismissive DISCUSSION, but write as thoughtfully as you would have done in the presence of suitably analysed data. Indeed, one of the most effective ways around this problem is not only to examine *carefully* the reasons for the inadequacies in the data (together with the ways in which such occurrences might be avoided in future work), but also to explore actively the sorts of implications that your data would have held had the results turned out (a) as predicted, and (b) contrary to prediction. At least in doing this you will be able to demonstrate your reasoning skills to your reader – as well as being able to practise them.

You should regard this, however, as a last, somewhat desperate, exercise in damage limitation. It is not an alternative to spending time designing sensible studies and doing adequate pilot work. The best way of dealing with this problem is by making sure that it does not arise in the first place.

Finally, one of the aspects of the implications of your findings that you should bear in mind concerns their external validity or generalizability.

5.5 External validity: the generalizability of findings

All studies have limits on their **generalizability** – that is, on the extent to which we can extrapolate the findings to situations other than those directly assessed in the study itself (Section 10.8). In some cases these limitations are severe. In others, they are not. Yet these

limitations are often among the first things to be forgotten when studies are discussed in general terms.

Many factors can affect the generalizability of a study's findings: the equipment or participants involved, the procedure used, the wording of the instructions, and so on. One particularly potent source of such limitations can come from the level at which *controlled* values were held constant. For example, if you undertake an experiment in which you control for sex by using only women, then perhaps the data are applicable only to women. Or, if you run an experiment in which you test the effects of alcohol on driving performance using participants who are only occasional drinkers, then perhaps the findings will not apply to those who drink alcohol more often.

The generalizability of findings is one of the more neglected issues in experimental psychology. It seems likely that the findings of many studies have been *over*generalized. As students, this is one issue that it will be important to think about in *all* aspects of your course, not just in your practical work. Nevertheless, in your practical work, consider this issue when assessing the studies you include in your INTRODUCTION and, of course, bear it in mind when evaluating your own findings in the DISCUSSION. However, again be sensible and balanced when raising this issue. It is easy to drone on and on about the failure to employ a random sample of the general population. Yet this is typically to miss the point. Examine whether there are likely to be any important limitations on the generalizability of your findings and whether any simple and reasonable steps could have been taken to improve the situation. Do not waffle on in a tone of self-righteous indignation about the failure to sample adequately every sentient human being within a 100-mile radius of the experimental setting.

? SAQ 21 What is the purpose of the DISCUSSION?

Summary of Sections 5.1–5.5

1 You should open the DISCUSSION by summarizing the main features of the RESULTS, so that it is clear both to you and your reader what you have to explain in the DISCUSSION. Relate your findings to the predictions.

2 Once you have done this, search for the best available explanation of the results, examining and assessing the likely candidates in

order to arrive at an overall assessment of what your findings have to say about the existence or otherwise of a causal relationship between the IV(s) and the DV(s).

3 These two stages are the necessary preliminaries for the final phase of the DISCUSSION, in which you assess the implications of your findings for the area as outlined in the INTRODUCTION. At the same time, you should think about the direction and form that future work might take.

4 Where your findings have been too ambiguous to do this, then you should examine both why this has been the case and ways in which the problems that you encountered might be avoided in future work.

5 End with a paragraph summarizing your main conclusions.

5.6 Six tips to help you to avoid some common failings in the DISCUSSION

1 Do not repeat material that you covered – or *should* have covered – in the INTRODUCTION. In this section you can assume that your reader knows the relevant literature – after all, it was you who described it. Where there are gaps in their knowledge that make it difficult to conduct your argument, and where these are the result of unforeseen rather than unforeseeable factors influencing your findings, then the gaps are of your own making and reflect omissions from your INTRODUCTION.

2 Include a final paragraph summarizing your main conclusions. Be careful, however, to distinguish between conclusions and mere restatements of findings.

3 Remember: Do not confuse *statistical significance* (see Section 11.2) with *meaningfulness*. We can draw useful conclusions from data that are not statistically significant. On the other hand, statistically significant effects can be psychologically trivial. Nor should you confuse statistical significance with *proof:* statistically significant results do not *prove* that the theoretical underpinnings of your study are sound, nor do findings that are not statistically significant necessarily *disprove* your arguments. Life would be much simpler if they did, but unfortunately this is not the case; in either instance, you will have to justify your point of view in the DISCUSSION.

4 Do not simply reformulate and repeat points that you have already made. This is **waffle**. Markers are not *that* stupid. We know when

you are waffling, so do not waste your time. Each statement should add something to the picture.

5 Do not see the DISCUSSION as an opportunity to indulge in fanciful speculation. Make conjectures with caution and keep it brief.

6 Once you know how, consider the likely effect sizes of the IVs on the DVs when discussing the implications of your findings. Remember that highly statistically significant findings do not necessarily indicate large effect sizes (see Chapter 12).

5.7 Two example DISCUSSION sections

Below are two sample DISCUSSIONS, one for the cheese and nightmare experiment, and one for the mnemonic experiment. These are only suggestions as to how you might go about writing discussions for these sections, and in some respects they provide only outlines of what might be argued. (For those unfamiliar with the term, **demand characteristics** are features of an experimental setting that provide the participant with cues about the experimental hypothesis – such as asking people to eat cheese before going to bed and then questioning them about whether or not they experienced nightmares. In a culture where eating cheese is believed to cause nightmares participants may well guess what the study is about and this may influence what they report.)

5.7.1 The cheese and nightmare experiment

Discussion

The results are consistent with the experimental hypothesis: those who ate cheese 3 hours before going to bed reported a significantly greater tendency to experience nightmares than did those who did not eat cheese at this time. Participants were randomly allocated to conditions, thus lessening the likelihood of there being differences between the groups in tendency to experience nightmares to begin with. (1)

Perhaps the most obvious explanation for this finding is that eating cheese before going to bed leads to the experience of nightmares. It may be that one or more of the ingredients in cheese affects the human nervous system and induces nightmares. However, an alternative explanation is that participants were aware of the aims of the experiment and this influenced their responses. Although the instructions concealed the purpose of the experiment, those in the

cheese condition ate a measured quantity of cheese at a specified time and (among other things) subsequently recorded the number of nightmares they experienced. This may have led them by suggestion either to *experience* an increase in the number of nightmares or to *report* such an increase.

It is hard to control for this possibility. However, the questionnaire completed each morning contained three pages of questions, many of which were "filler" items designed to draw attention away from the questions on nightmares. Careful post-experimental interviewing by an investigator blind to condition revealed few participants able to state the hypothesis of the experiment, even when strongly encouraged to do so. This suggests that the participants had little or no awareness of the experimental hypothesis.

On balance, given this apparent inability of participants to state the hypothesis, it seems reasonable to assume that there may be something to the common-sense idea of a link between cheese and nightmares. Further work should attempt to replicate the current findings, controlling as much as possible for demand characteristics. These studies should move away from a complete reliance on self-reports; researchers should include less subjective measures, such as measures of rapid eye movement (REM) sleep. Although it is not possible to obtain objective measures of dream content, measures might reveal differences in the amount of REM sleep in the two conditions. In all experiments, investigators should make every attempt to conceal their hypotheses.

Once further research has established that there is a link between cheese and nightmares, work can begin to determine which of the ingredients of cheese causes the problem. Isolating such an ingredient might help to reduce such unpleasant experiences (particularly if the ingredient is common to a number of foodstuffs) and may also provide useful insights into brain chemistry.

Thus the data from this experiment raise the possibility of a link between eating cheese and experiencing nightmares. However, it is not possible to rule out an explanation of the findings in terms of demand characteristics. The next step is to replicate this study controlling as much as possible for demand characteristics.

5.7.2 The mnemonic experiment

Discussion

Participants using the mnemonic recalled significantly more of the easily imaged words than did those not using the mnemonic. However, the groups recalled similar numbers of the hard-to-image

words. Thus, as predicted, the mnemonic specifically enhanced recall of the more easily imaged words.

The pre-test data suggest that, overall, the two groups had similar ability to recall lists of words prior to the experimental group being taught the mnemonic. It thus seems reasonable to conclude that the randomization to conditions was successful and that it is unlikely that the findings arose from individual differences between the groups in their abilities to recall lists of words.

The manipulation check confirmed that the participants rated the words classified as easily imaged as being easier to image than those classified as hard to image. This corroborates the classification of Clark (1971) and the manipulation of word type imageability used in the current experiment.

The findings of this study are thus consistent with those of Clark (1971). The data suggest that the method of loci does make a difference to recall and does so by enhancing the recall of easily imaged words. Earnshaw et al.'s (2007) failure to find an improvement in recall among those using the mnemonic appears to have resulted from not allowing the participants enough time to practise using the mnemonic.

The current findings have limited theoretical implications. There is some suggestion in the data that the participants recalled easily imaged words more readily than they did hard-to-image ones. (The 95% confidence intervals for the means for the easily imaged and hard-to-image words for the no-mnemonic group overlap but only by a small amount.) Of course, despite being matched for length and frequency, there may have been other ways in which the lists differed over and above their imageability that could account for any differences in recall. However, the data in both Clark (1971, Table 2) and Earnshaw et al. (2007, Table 1) show differences in the same direction and of similar size. Earnshaw et al. used different words. In combination, these data raise the possibility that easily imaged words may be intrinsically easier to recall even without the use of a mnemonic (perhaps because they typically denote more familiar, concrete objects). Given that this difference would be theoretically interesting, future research might test whether people recall easily imaged words more easily and investigate reasons for the difference if it is shown to exist.

The practical implications of the current findings are clear. Researchers must ensure that their participants have had sufficient practice in the use of the mnemonic before the start of any experiment in this area.

In conclusion, the findings suggest that the method of loci does make a difference to recall and does so by enhancing the recall of

easily imaged words. Earnshaw et al.'s (2007) failure to find an improvement in recall among those using the mnemonic appears to have resulted from not giving the participants enough time to practise.

5.8 Writing a DISCUSSION when your study is not an experiment

When your study is non-experimental, the only real difference in your DISCUSSION is that you will not be attempting to interpret the causal relationship between an IV and a DV. In all other respects, however, your task will be the same and much of the advice I have offered in the rest of the chapter will also be relevant to you. You should therefore still find yourself addressing the three questions on p. 74: (1) What are the findings of this study? (2) What do they mean? (3) What are their implications?

If your data are correlational, in relation to question 1 you will be assessing the evidence that there are associations between your key variables and the magnitude of these relationships; for question 2, you should explore the extent to which the presence of third variables or other features of the data may account for your findings (see Section 9.2), and for question 3, you should consider whether your findings justify undertaking relevant experimental studies and, if possible, say something about how this might be done. You must be careful at all times not to draw any causal inferences from your data – remember, the relationship could always be the other way around. (See Section 13.8 for more on this.) This can be especially a problem when you have analysed your data using linear regression (see Section 4.6.12).

If your data are descriptive, in relation to question 1, you should summarize the key elements of your findings so that your reader knows which aspects you think are the most important. When addressing question 2, you should examine what your data have to say in relation to the goals of your study – are the findings consistent with your expectations and if not why not?; if your findings are consistent with your expectations, are there competing explanations as to why? As your research may well be descriptive because it is exploratory, question 3 is very important. What do your findings have to tell us about what research we should do next? Are they promising or discouraging? What kind of research is needed to build on your findings? In

particular, can you say something about what questions this research should address and what the next study or studies should look like? What relationships between variables should be explored in future correlational research? Is there scope for any experimental research and, if so, what variables should we try to manipulate (and do you have any ideas about how?) and what should be the dependent variables (and do you have any thoughts about how these should be measured)?

If you do not understand the distinction between experimental, correlational and descriptive research, see Chapter 9.