Discovering Statistics: Experimental Project

Overview

In this laboratory project you will have to think up an experiment, design the experiment yourself and then collect and analyse your own data. The background for this laboratory project is memory, about which you should be learning in your C8551 Cognitive Psychology module. So, this project is intended to fit in with what you’re currently studying and I hope this enables you to draw upon this information to develop your research ideas.

In the first session you should get into small groups (3-5 people) and discuss some simple ideas based on the ideas in this handout (which are all about memory because they were kindly donated to me by colleagues who know something about memory – you do not have to do a memory experiment if you don’t want to, as long as it addresses a topic covered in C8551 Cognitive Psychology). You should discuss your ideas with your seminar tutor. Use the next two weeks to refine your ideas into an interesting research question, design your experiment, and work out which statistical procedure you will use to analyse your data. You have 4-5 weeks to design a study and collect data before your session in which you analyse the data. You should feel free to discuss ideas and design issues with your seminar tutor or myself in that 4 week period. Some things to remember:

- **Think about how you’re going to analyse your data before you collect it.** This is very important, as it is a fundamental principle of well-conducted research. Also, it’s very soul-destroying to collect lots of data only to realise that they cannot be analysed. Your seminar tutors (and me) are happy to offer advice about analysis before you start collecting — it is much more difficult for us to ‘salvage’ a badly designed experiment.

- **Don’t try to win the Nobel Prize for science:** These are small scale projects so don’t try anything too complex unless you’re very confident about the analysis and execution.

Projects in which you design your own study give you your first taste of doing research (and act as a ‘warm-up’ for your final year project). As such, it’s very important that you embrace the opportunity to practice conducting research. Don’t worry if you find it difficult — it is a very tough transition from carrying out projects to designing your own. As ever, any feedback on this project is welcomed via the module forum. You may also find the Field & Hole book (on your reading list) useful (Chapter 2 suggests how you can refine a research question, and Chapter 3 offers some advice on experimental design).

Ethics

Normally when we do research, we have to ensure that our procedures conform to ethical codes of conduct and also complete an ethics form that is assessed by a university panel to approve the procedures. The procedures for this process at the University of Sussex can be found here: [http://www.sussex.ac.uk/res/1-6-12.html](http://www.sussex.ac.uk/res/1-6-12.html)

To sum up, the university classifies research as being either high or low risk. For these projects, there is not enough time to go through the formal ethical procedures, which is one of the reasons why we get you to base your work on published research, which will have been subject to ethical clearance (in other words, the procedures should be ethical). Instead, for these projects, we have a blanket ethical approval for any project falling within the remit you have (e.g., a cognitive psychology experiment) that the university would deem ‘low risk’, and forbid any project that would be classified as ‘high risk’. Some examples of projects that are forbidden because they would be high risk are:

- Any project involving anyone under 18 years old or other vulnerable groups.
- Any project involving administering food or drugs (e.g. alcohol).
- Any project that induces psychological stress, causes humiliation or
- Any project involving hazardous substances or equipment.
- Any project involving deception of any kind.

In addition, all projects on this module must:

- Obtain informed consent from each participant (example consent forms and information sheets can be found at [http://www.sussex.ac.uk/staff/research/spg/researchgovernance/apply](http://www.sussex.ac.uk/staff/research/spg/researchgovernance/apply))
Memories can be reconstructive. An influential metaphor that Neisser (1967) used was of a paleontologist trying to construct a dinosaur from a few “stored fragments” (i.e., fossilized bones) coupled with theories of paleontology.¹ Memories can

¹ Neisser notes that Hebb used this metaphor for focal attention: “Hebb’s (1949, p. 47) comparison of the perceiver with a palaeontologist, who carefully extracts a few fragments of what might be bones from a mass of irrelevant
be reconstructed from perhaps a few fragments of the filtered perceptions of our cognitive system, from script base knowledge of the event type, and from information about the event that was not part of the initial encoding. Memory distortions can arise from any of these stages.

Memory distortion work has often been linked with applied concerns. In the 1980s this was done mainly in relation to eyewitness testimony and to a lesser extent text processing. In the 1990s the recovered memory debate dominated memory research, with psychology and psychiatry being presented (often in a bad light) on the covers of all the main US news magazines, and with contradictory expert testimony producing multi-million dollar legal settlements and imprisoning innocent people (see Loftus, 1997). Further, there has been a large increase in children testifying in court (Ceci & Bruck, 1993). The applied interests should not take away from the theoretical importance of memory distortion research to our understanding of human memory.

There are several different procedures to examine memory distortion. The following is a partial list with a few thoughts about the research approaches and some references. I have not limited the list to ones where ANOVA-type designs are usually used, because you may design a study that requires a different analysis and this is perfectly OK (you do not have to use ANOVA).

A good introduction to memory can be found here:

http://www.exploratorium.edu/memory/index.html

1. Post-event information (PEI)

This refers to the classic Loftus-like studies where people see an event, in some form, then are given a post-event narrative which, for some people, presents misleading information, and then participants are tested. This procedure has been used more than other procedures and there is much research on differences in characteristics of the event, the PEI and testing.

→ Loftus, Miller & Burns (1978).

2. Deese-Roediger-McDermott Procedure

In North America, over the last ten years, it seems that to qualify as a memory researcher you have to have used the DRM procedure. This involves showing people lists of semantically related words. For each list, there is a critical lure that is missing. When asked to recall the lists, people often include these critical lures. Lists are included at the end of the Roediger and McDermott (1995) paper. The Roediger and McDermott (1995) paper and subsequent studies have used words. It would be nice to have studies designed, which use more ecologically valid stimuli (such as pictures or sounds). Deese (1959) first used this technique, but the paper was not well known when it first came out.


3. Biasing questions

Loftus and Palmer (1974) showed people several cars crashes and then asked them the velocity of the cars. They used different verbs to describe the speed (for example, “hit” and “smashed”) and found people gave faster estimates the higher implied speed of the verb. Burt and Pople (1996) have more recently shown this works also with duration estimates. It is important to choose a continuous variable as the response variable.

rubble and ‘reconstructs’ the dinosaur that will eventually stand in the Museum of Natural History” (1967, p. 94). Hebb’s own description, “a drawing or a report of what is seen tachistoscopically is not unlike a palaeontologist’s reconstruction of early man from a tooth and a rib,” is less ambitious than Neisser’s elegant reconstruction. It is worth stating that Neisser still believes that memory is constructed, he just believes we no longer even have the few fossilized remains to aid the reconstruction, but that memories are complete constructions.
It might be worthwhile to look at the recent recommendations for avoiding biasing questions when interviewing eyewitness (http://www.ncjrs.gov/pdffiles1/nij/178240.pdf).

4. Memory implanting

There have been some impressive studies, with adults and children, showing memories for events that did not occur. The most impressive demonstrations are therapists getting clients to believe that they were abducted by aliens or experienced biologically impossible events. The laboratory analogs involve getting people to remember some childhood events that happened (parents are usually contacted) and one that did not happen. This is probably impossible for the current assignment, because it takes time to contact parents. Still, some of you may come up with good ideas on how to do this within the constraints of the assignment.


5. Retrieval induced forgetting

Anderson, Bjork and Bjork (1994) showed participants several sets of words. The sets were, for example, a list of fruits. Later, participants were re-presented with half the items for half of the sets. Later, there was a recall test. The items that were repeated were recalled the best. The items from sets not presented were the next best recalled. The worst recalled items were those from the sets from which some items were recalled, but not those particular items. Bjork and Bjork describe how the retrieving of related items inhibited memory for the other items in the set. These sets are included in an appendix to Anderson et al. (1994).


6. Imagination inflation and dream interpretation

People are given a list of experiences and asked if they experienced them. They are then required to imagine some of these experiences. Later, they are asked the likelihood that the experienced really happened. Those experiences imagined are rated as more likely that those not imagined. In the Mazzoni et al. paper, they use dream interpretation to increase event likelihood.

→ Mazzoni et al. (1999).

7. Memory Conformity

In eyewitness situations, if several people see the same crime and then discuss the event then often their memories will be different. When talking, what one person says can influence what the other person reports on later tests. Schneider and Watkins (1996) showed that this happened in a memory test for words. Wright, Self and Justice (2000) showed that this works also with memory of events, and in fact the effects look strong.


8. “You got ‘em” studies

Wells and Bradfield (1999) showed people an event, and then had them try to choose the culprit from what’s known as a ‘target absent’, or ‘blank’, identification parade (i.e. an identification parade in which all six of the people were not the culprit). Even though the people hadn’t seen the culprit, if they were told that they had chosen the correct person then not only was their estimated confidence higher, but so were characteristics like how long they saw the culprits face.


Available: www2.fiu.edu/~dwright/research.htm

9. Levels of Processing

How well you remember something depends on how you processed it (how you thought about it), with ‘deeper’ processing leading to better memory. Deep processing means attending to the meaning of the stimulus (its ‘semantics’); shallow processing means attending to its physical attributes. (See, for example, Eysenck & Keane, *Cognitive Psychology: A student’s handbook*, or any other textbook on memory).

Hyde and Jenkins (1973) showed participants a list of words. How participants processed the words was determined by which of three orienting tasks they were given:

1. One group were asked to detect an “e” or a “g” in the words.
2. One group had to decide the part of speech of the word (e.g. noun, verb, etc).
3. One group had to rate the pleasantness of the word.

The rationale was that to rate the pleasantness of a word you have to think about the meaning of that word in depth. Therefore, people in the group who made these ratings should remember more. In fact, they found that pleasantness rating led to double the recall compared to only looking at physical attributes, like which letters are in the word.

To minimize participants trying to process the words in other ways, for half the participants, learning was incidental (i.e. they were not told there would be a subsequent memory test). For the other half of the participants, learning was intentional (they were told of the subsequent memory test). Hyde and Jenkins found no difference between participants who learnt intentionally and incidentally: The intention to remember in itself did not help people remember.

For your experiment you could look at what sort of tasks might improve people’s memory compared to other tasks? You can remain close to the Hyde and Jenkins experiment, but it would be more interesting to use your experiment to provide an answer to a question to which you would actually like to know the answer.
How Similar Can My Experiment Be To Past Work?

You can, if you like, simply replicate one of the studies that I have just described. However, better marks will usually go to people who have done something slightly different. You could replicate one of the studies but using a different modality (say, replicate an experiment using visual stimuli, but use auditory stimuli instead), or you could improve the methodology (if you feel there were problems with the original study). Some people do completely new things but be realistic: don’t design something with 58 different groups and then spend every waking hour trying to collect the data.

Things to think about When Designing Your Experiment

Data Analysis

You should use the method of analysis that is most appropriate for the data that you have collected. This method could be anything that you have been taught in both first and second year. Contrary to a pervasive myth, you do not have to analyze your data with an ANOVA …. Unless, of course, an ANOVA is the most appropriate method for analyzing your particular data set 😊

References

We have just put one or two references for each suggested idea. It will be worth searching for more. The Web of Knowledge (http://wok.mimas.ac.uk/) will be useful for this.

Format of your Data

In memory research often the key response variable is whether somebody remembers an item. This is not appropriate for parametric statistical tests based on the normal distribution (i.e. t-tests, ANOVA etc.). A common way to turn this into a more continuous variable is to ask people how confident they are in their memories. Pretend you use a five-point scale, and the person either remembers or does not remember. You need to calculate a new variable. If the person remembers the event, then the new variable is equal to the confidence score. If the person does not, then it is the negative of their confidence score.

Remembers:

<table>
<thead>
<tr>
<th>Yes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

New Variable: -1 -2 -3 -4 -5 0 1 2 3 4 5

Another useful way around this is to have several targets (several things that you want people to remember) and use the number (or percentage) that they do remember as your dependent variable.

You should still check the distributional assumptions.

Random assignment of participants to groups

Participants should be randomly assigned to your experimental groups (you can add the constraint that each group has an equal number of subjects). This is for a couple of reasons.

First, the way you run the experiment will subtly change as you run it. You may be more smooth or more or less motivating in subtle ways. Second, the sort of participant you can recruit as you run the experiment will change as you run the experiment. First you will ask your friends. Pretty soon you have used up your friends and you are on to acquaintances. By the 6th participant you need to run complete strangers. Your friends may have certain common characteristics – e.g. you all like to study hard or you are all alcoholics. Also the friends you run at first will be the ones that agreed quickly; the friends you run last will be the ones you had to twist the arm of to get to do the experiment. So if you finish group 1 first then group 2 then finally group 3, there could well be a different type of person in each group. If there was a difference between the groups’ recall you would not know if it was the experimental task that made a difference or the type of person you’d tested.

Random assignment literally means random, it does not mean that you just did not have any specific method you can put your finger on. If participants are randomly assigned to groups, any significant result you get CANNOT be
explained by individual differences that you did not control. If you get a significant result, do not say in your discussion “But we didn’t measure IQ, this must influence how well people can remember words and perhaps this explains the significant result”. Random assignment means that any characteristic that may differ from one person to another should be roughly equally distributed across the experimental groups. Of course, random assignment will sometimes produce a large excess of e.g. clever people in one group, but that’s exactly what your p values measure: How likely is it that you’d get the data that you have if chance produced the results. Chance is always at work, and that’s what you are controlling by setting your p values.

**Experimenter effects**

You will have more than one experimenter (because you’re working in small groups). Therefore, each experimenter should run the same number of people in each group as every other experimenter. Do not say “Ok, I’ll run the people in group 1, you run the people in group 2 and so on”. The reason why you should do this is because of experimenter effects. Who the experimenter is, their expectations and character can strongly affect how participants perform. Even when all instructions are written down, it has been shown that different experimenters can produce different levels of performance. Even rats perform differently when different experimenters with different expectations are administering the task.

**Write down your instructions**

Give participants written instructions. Writing the instructions will help you think clearly about how best to instruct people. It will also help you give a more standard delivery throughout the experiment.

**Keep everything constant except for the cover task**

In your experiment you’ll have different groups of people doing different tasks (perhaps with different cover stories). Do not use different sets of stimuli (most of you will be using words or pictures) for different experimental tasks, otherwise you won’t know if any difference between the groups is because the tasks varied or because the words varied. For example, different words are more or less easy to remember – for example, concrete words are easier to remember than abstract words. If you are interested in the effect of a task, use the same stimuli for each task. (Of course, you might be interested in different types of words – then you would use the same experimental task for different sets of words.)

For example, if one cover task is to state the colour of the words or the letter case they are in, you will need words of different colours and in upper and lower case. But then use the same stimuli for all groups. So the group that does a semantic task still sees the words in different colours. By keeping constant as many things as possible you can say more precisely what could have caused any differences between the groups.

You should also control the exposure time of each stimulus. If you just give people the list of words all at once, people might take 2 minutes to state the length of each word, but 3 minutes to say what room in a house it is most likely to be found. If recall is different between the groups, is that because of the type of processing they engaged in, or just because they looked at the words longer in one case rather than the other? You could show each word on a card for a set amount of time, e.g. five seconds, to get round this problem.

**Recency effect**

Just after a participant has seen a set of stimuli, the last few stimuli will be remembered almost perfectly, regardless of the experimental task. This is known as the recency effect. Therefore, the last few stimuli will not be especially useful in distinguishing the groups if there is a recency effect. To get around this, you could give participants a distracter task after the study phase (e.g. solve some maths problems for a minute), or make sure that you use some distracter stimuli at the end of the study phase, so that the stimuli you’re actually interested in do not appear at the end of this phase.

**Congruency effect**

For experimental tasks that involves a yes/no answer, participants remember those words best that they gave a “yes” answer to rather than a “no” answer. If participants are classifying the stimuli into different categories, the number of categories can also influence recall. So try to have the same number of response types for each experimental task, and the same number of stimuli for each response type. For example, if one task is “Are the words in upper case?” you
could have half the words in upper and half in lower case. If another question is “Is it an animal?”, have half the words being an animal and half not.

**How is memory assessed?**

You could use a free recall test ("In the next two minutes, try to recall all the words you saw"), or a recognition test ("Here are 40 words, half were the ones you just saw and the other half are new. Say which is which.").

**How Many people should I test?**

You should try to test at least 10-15 people in each experimental group.

**Pilot experiment**

Before you run the experiment proper, it is a good idea to run a pilot. Just run a couple of people through the procedure. Check that the instructions were clear, the timings are appropriate, and the tasks are neither too easy nor too difficult. If all goes smoothly you are ready to run; otherwise, alter the procedure and try it on another pilot participant.

**References**


*This handout has evolved over many years with contributions from Zoltan Dienes and Dan Wright.*