Short guide to writing a report

This version contains some general comments, some specific to our topic in this semester.

What you should aim for

First of all, your most important aim is honesty. If you lie and people believe you, decisions will be based on false information and will more likely be wrong. If you lie and are found out, you reduce trust in science, people will base decisions on worse information, and again decisions are more likely to be wrong. False beliefs leading to bad decisions can have quite serious consequences. For an example, look at the South African policy not to provide antiretroviral drugs to AIDS sufferers because a previous health minister chose to believe that AIDS has nothing to do with the HIV virus. People died as a result of that decision.

Your second aim is clarity. If your reader comes away from your report thinking "I never knew it was so simple!", you have succeeded. If your reader thinks "My word, this is complicated!", you have failed. The standard structure of a scientific paper is standard because it usually helps to keep your presentation clear. If you ever find that it hinders clarity, you certainly have my blessing to abandon it. As far as I am concerned, the standard structure is a means, not an end. If you are writing a report for someone else, whether a lecturer in a course or a journal editor, check whether they have stronger feelings than I do about standard formats.

I recommend you compare my recommendations in this guide to an example paper by Karpicke and Roediger. Please read and analyse that with regard to clarity. If your analysis is good, you may be able to improve on it.

Your third aim is that people should remember your arguments. If you can find a metaphor, a story or a joke that illustrates the point you are trying to make, that is generally a good thing. But a joke or story that is irrelevant merely distracts and should not be there. Worst of all is something memorable that is related to your argument, but is ambiguous or misleading.

On the positive side, it helps to find a surprising link to something that most people know. Try to be creatively mundane.

Introduction

The introduction should start off with a "Why should you care?" bit. To get you in the right frame of mind, imagine one of the following scenarios. You are in the pub and someone asks you what sort of practical work you are doing right now. What would you say if you want to give the impression that you are doing something fascinating? Or imagine you are applying for a research grant. These days, typically only the top 10 – 20% of projects get funded. How can you persuade the committee, which contains a good number of people who know nothing about your specialist subject, that your project is that good? A brief version of the arguments you would use in these situations should be at the beginning of a scientific paper.

In our specific case, why should you care about the relationship between working memory, Boolean categories, and intelligence? What increase in theoretical understanding or practically applicable knowledge might follow from looking into this specific question?

Next comes some more information on the general topic which you are investigating. What is already known about the subject?

You have already seen my approach to this: I start by emphasising how little we know about intelligence. We are a *long* way from understanding intelligence well enough to reverse engineer and reproduce it. The best we can do at present is identifying some limiting factors. Analogous to how we can identify a car's engine power, weight, aerodynamics, and suspension design as limiting factors for speed, we can also identify psychological traits that limit intelligence. In the papers I have presented to you, the suggestions are:

1) Simultaneous storage and processing. That is intuitively appealing, since reasoning involves both. However, this really describes more the tasks that are being used to measure *something*, without really telling you what that something is. Anyway, there are studies showing good correlation of

intelligence and tasks without a memory component. For example,

2) Engle (2002) argues that the ability to focus attention is the mental capacity crucial for intelligence. Then again, it seems that focused attention depends mostly on capacity to inhibit automatic responses, which, according to Friedman et al (2006) has not direct relationship to intelligence. However, if you want to discuss its importance in daily life, you could check whether it is related to self-control. I have seen some papers that indicate a connection.
3) Updating (Friedman et al., 2006).

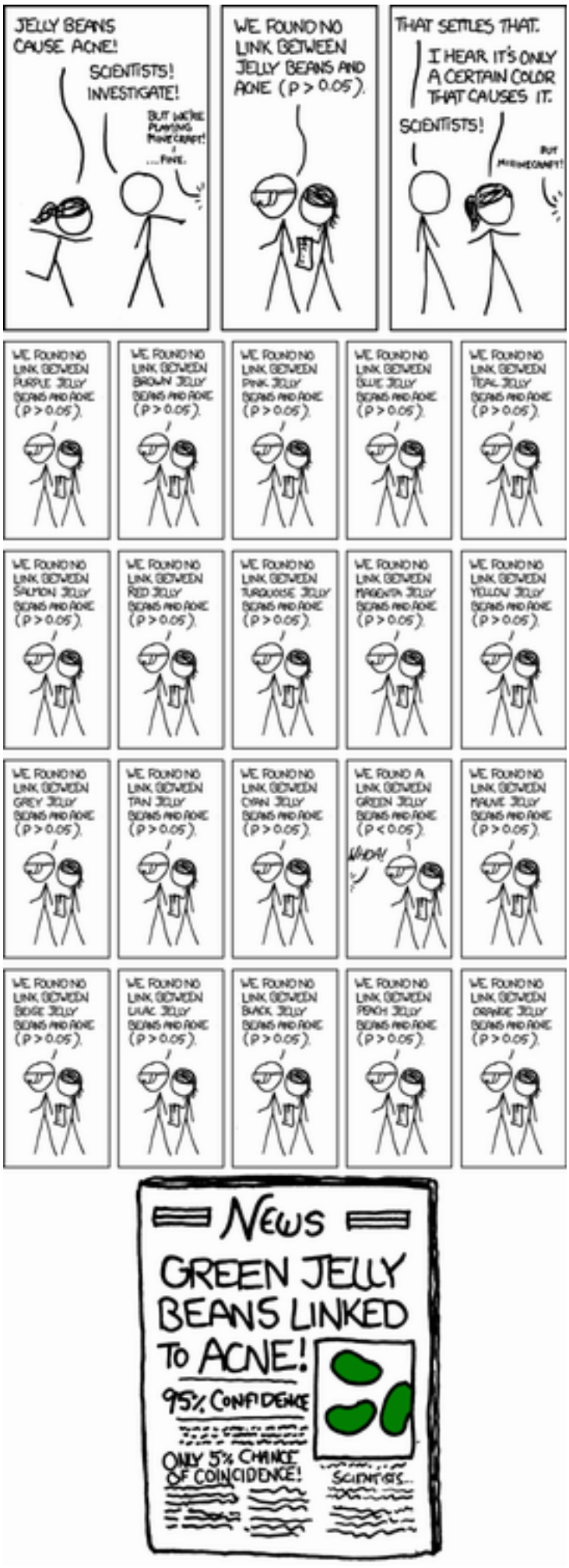
Of course, that work has relied on the assumption that there is one factor to intelligence. Hampshire et al. (2012) make a good case that there are two, memory and reasoning. Those were also two of the intelligence factors that Oberauer et al. (2008) measured when they investigated another possible limiting factor for intelligence, namely

4) Relational integration (Halford, 2005, Oberauer et al., 2008). Oberauer et al. say relational integration is highly correlated with reasoning, and that it explains differences not explained by other factors. If relational integration is important, then it is plausible that there should be a connection to Boolean categories. These categories differ in that information from a varying number of dimensions must be integrated to arrive at a complete description of the category. That sounds rather like relational integration.

You should explain what connections you see among these things. You then turn your attention to which question that is left unanswered by previous work you will try to answer here. This question or hypothesis should be derived from the research and theories you have discussed. You absolutely need that connection. Having stated the question you want to address then gives you a nice transition to *how* you will try to answer that question. That you explain in the Methods.

But before I comment on the methods, I again need to get specific. Here are the research question this experiment was intended to answer:

1) Is understanding of Boolean categories based on relational integration? The relevant result is the correlation between relational integration and Boolean categorisation on Sheet 2 of the Excel file. Report as: Spearman's rho = -0.199, $p < 0.01$. There is only a very weak relationship.
2) Is relational integration (as measured by the three tasks from the Oberauer lab) related to intelligence? More specifically, is it related to reasoning? Or else to memory? The relevant results are the correlations between relational integration and either the four best reasoning tasks in Hampshire et al.'s test battery (cells AK45 and AK46 of sheet 3, report as Pearson's $r = 0.307$, $p > 0.05$), or the four best memory tasks (cells AL45 and AL46 of sheet 3, report as Pearson's $r = 0.481$, $p < 0.01$). Only the correlation with memory is significant.
3) Is relational integration (as measured by ability to understand Boolean categories) related to the reasoning and/or the memory component of intelligence? We really should divide the results into the four separate groups, but we don't have enough data for that. I checked whether there was an overall effect if I ignore differences among groups. Turns out there is something: Boolean categorisation performance correlated significantly with both reasoning (Spearman's rho = -0.438, $p < 0.01$) and memory (Spearman's rho = -0.407, $p < 0.01$). Remember that because a small number of trials in categorisation means good performance, a negative correlation means those who are good at the Boolean task also tend to have good memory and reasoning. These correlations measure an overall effect, but we know (see next point) that there are differences among the Boolean conditions, so it is likely that the strengths of the relationships with reasoning and memory vary, too. The correlational analysis that doesn't check for such differences can't tell us whether that is true.
4) And finally, a question only about the Boolean categories is whether the Boolean complexity (how many elements are there in the shortest description of a category?) is really all that matters for how difficult a category is to understand. When a category that has a long Boolean description also has a recursive structure, can people spot that and use it to simplify their description of a category? Those results are in the ANOVA table on Sheet two of the Excel file. Gradual introduction makes the task easier ($F(1, 214) = 58.54$, $p < 10^{-12}$, $\eta^2 = 0.215$). The patterns do not differ overall ($F < 1$). However, the XOR is easier than the alternative pattern when introduced gradually, and more difficult when



: 12.71, $p < 0.001$, $\eta^2 = 0.056$).

You are *not* being asked to take this data set and come up with your own hypothesis. One reason is that it may mislead you into naïve data mining. Consider the cartoon on the left:

If you decide to test the relationship between acne and 20 different colours of jelly beans, you have two choices: either you divide the conventional statistical significance level of 0.05 by the number of comparisons. That would mean you only accept that there is a link between some colour of jelly beans and acne if $p < 0.0025$. Or else, you use this initial analysis with its multiple comparisons only to *generate* a hypothesis (that green jelly beans cause acne), and you test it with a *new and independent* data set. If your hypothesis is true, that should show in the new data.

Handing you an experimental design and data set and asking you to come up with a question to ask (a hypothesis) risks giving you the misleading impression that the naive data mining illustrated in this cartoon (where the data are generated first, then a single hypothesis is picked to fit the data) could be legitimate science. It is *not*. It would be junk science.

Methods

In the Methods, you should explain how you have designed your experiment in enough detail that someone else can understand your work. In the ideal case, it should be possible to reproduce your work without having to contact you for further details. If the procedure is very complicated, it may not be possible to provide all details, but at least a conceptual replication should be possible. You should include enough information that if someone correctly uses all the information you provided in the Methods section and fails to replicate your results, you cannot complain that the methods were wrong.

You should also provide some explanation why you design the experiment as you do. Which procedural detail controls for what possible alternative explanation? Which patterns of results will be consistent or inconsistent with which theories (in other words, which data are interesting, and which are boring)? If the statistical analysis is going to be complicated, you should also explain here what tests you will use on which data.

Some journals have changed their standard formats and put the Methods section at the end. In that case, you will have to provide enough information about what you have done in the Results, so that your readers can know what the data mean. So the Results section will then have the information needed to understand the data, while the Methods section will contain enough details to replicate your work.

Results

Here you explain what you have found. Once more, honesty and clarity are your most important aims, and everything else is secondary. It is good practice to say straight away what the data mean, and then offer the statistics. The reader's job is rather harder if the author starts off with the statistics, and only then explains what they mean. Here is an example from a published paper:

> A (3) (memory response) x (3) (emotional valence) repeated measures ANOVA of mean generation latency for these memories yielded only a significant main effect of emotional valence, $F(2, 78) = 8.05$, $p = 0.001$, $\acute{\eta}_p^2 = .171$. No other main or interaction effects were significant. Participants took slightly longer to generate memories to neutral ($M = 16.54$, $SD = 8.07$) than to negative ($M = 13.45$, $SD = 7.15$) or positive ($M = 11.99$, $SD = 5.87$) category or cue words, $t(39) = 3.14$, $p = 0.003$ and $t(39) = 4.08$, $p > 0.001$ respectively; this finding may reflect the extremely mild (and perhaps personally unimportant) nature of the neutral memories.

The authors then presented the most important piece of statistics halfway down the next paragraph, without flagging it up as being important. There is good science in that paper, but the authors made the reader's job harder than it has to be. What I find easier to understand is to present the descriptive statistics (average ± standard deviation or standard error) in a graph, and in the text say first what the result is, and put the statistics second. Like this:

> Only emotional valence influenced how long it took participants to generate memories (main effect of emotional valence: $F(2, 78) = 8.05$, $p = 0.001$, $\acute{\eta}_p^2 = .171$). No other main effects or interaction were significant. Participants took slightly longer to generate memories to neutral than to negative or positive category or cue words ($t(39) = 3.14$, $p = 0.003$ and $t(39) = 4.08$, $p > 0.001$ respectively). This finding may reflect the extremely mild (and perhaps personally unimportant) nature of the neutral memories.

Discussion

The primary purpose of the discussion is to put your results in context with previous work. What does it mean? Have you found something new and surprising? Do your results change the way people should think about your area of research? Or have you confirmed and strengthened previous findings. What are the implications for the why should you care argument? Do your results make a difference to the real world?

Keeping in mind this primary purpose, one generally builds up to it. First discuss your specific findings, then put them in context with other work, then write about the general implications, if any.

What can we say about the relationship between relational integration, as measured by Oberauer and von Bastian's three tasks, and Boolean categorisation is so weak? It seemed plausible that they are both relational integration tasks, but then the correlation should be stronger. What does it mean that relational integration is only significantly correlated with the memory component of Hampshire et al.'s test battery? Does this fit with Halford's and Oberauer's ideas? What does it mean that Boolean categorisation shows modest correlations with both memory and reasoning (provided you choose to trust an analysis that ignores all the differences among the four conditions of Boolean categorisation)? What does it mean that Boolean categorisation depends not only on Boolean complexity, but also on whether people can spot a recursive pattern?

Normally, by the time someone writes a first scientific paper, at least a year or two of work has been invested in that research, and the author will have read and discussed at least 50 to 100 papers relevant to the research. A typical paper has 20 to 50 references, and you can expect that for every reference cited the author has read several papers which are related to this research, but are not quite relevant enough to be worth citing in this paper. I do appreciate that you do not have time for that much reading in a course lasting just one semester, especially when you have other things to do besides report writing. Therefore I do not expect a lot in terms of relating your findings to a lot of relevant previous work, simply because you will not

have had time to read it. I do expect to see some indication that you understand the purpose of a discussion.

I see a lot of discussions that consist of lists of issues that may affect the validity of the conclusions. When you look at scientific papers, you will find that the authors normally assume that the readers are competent to make these judgements, and they will raise validity issues only if there is something that is not obvious or easy to work out from the Methods and Results. So if a psychologist has tested 23 white, right handed, middle class Anglo-Saxon protestants of about age 20, the assumption is that the readers know enough to judge to what extent the results may be specific to that group. If you study depth perception by binocular disparity, the assumption will be that this kind of basic sensory processing is pretty much the same for most people. If you study attitudes to a patriarchal society structure you probably know that this is likely to depend on culture and possibly personality, and you assume this is also clear to the average reader of a specialist scientific journal, so this would merit at most a passing remark (with perhaps a relevant reference), not an extended discussion. If you write for Scientific American, or for Aftenposten, you assume less background knowledge, and you say more about the limits of your knowledge. You need to tailor the discussion of validity issues to the intended audience.

If you do discuss validity, the same criteria apply as when you present your results. I remember a report in which a student listed about half a dozen validity issues, but the one that was really important to this experiment was buried somewhere in the middle of the list. You should put the important stuff up front, and make clear that it is important.

Robert Biegler