**Bad Psychology**

**Tutorial by Ben Ambridge** [**www.benambridge.com**](http://www.benambridge.com)

This tutorial is based on the book and newspaper column *Bad Science* by Ben Goldacre (page numbers refer to this book). Although this book is primarily about medicine, it highlights many experimental flaws that are common in psychology and other fields. It is therefore highly recommended reading (see also Goldacre’s new book: *Bad Pharma*).

Below are listed a number of experimental flaws that constitute *Bad Science* (all are common in psychology!). Your job is to read the (fictitious) reports overleaf and identify the flaw(s) in each.

**1. Retreating to untestable positions (p.5) - If a prediction of your theory is unsupported, come up with a new claim that can’t be tested.**

**2. Failure to include a control condition (p.138) - Participants’ “scores” or behaviour will often change over time even if there is no experimental manipulation, particularly if they are aware that they are being tested. With no control condition there is no way to tell if any change is caused by the experimental manipulation.**

**3. Description as an explanation (p.16) - This is very common in psychology! Researchers often “explain” phenomena by giving them fancy names. In fact, the “explanation” does no more than re-state the findings that require an explanation.**

**4. Cherry picking (p.98). When lots of studies are done on a given topic, it’s inevitable that not all will yield the same findings. “Cherry picking” is citing one study that supports a particular point of view and neglecting to mention that the majority of studies yield the opposite conclusion.**

**5. Publication bias (p.212). Studies that fail to find an effect (i.e., a difference between two conditions) are rarely published. Whilst this is pretty unforgivable in drug trials (see *Bad Pharma*), it is a little more understandable in typical psychology experiments, where the failure to find an effect could more plausibly be due to problems with the experiment (e.g., an insufficiently sensitive dependent measure). On the other hand, it could be that the experimental manipulation really makes no difference. (Obviously, it’s difficult to give an example of a study that doesn’t exist for the exercise; but one of the passages should make you at least suspicious about the possibility of publication bias).**

**6. Confusing correlation and causation (p.89). Just because X comes before Y, it doesn’t mean that X caused Y. Y could have been caused by something else entirely (possibly the same thing that caused X).**

**7. Using a surrogate outcome measure (a particular problem with drug trials – see *Bad Pharma)*. It’s difficult to get people to change their behaviour! So rather than measuring whether an experimental manipulation actually changed behaviour, researchers often – how can I put this – increase their chances of getting the result they want, by measuring a “surrogate” outcome. For example, if the study was assessing the effectiveness of anti-smoking adverts, the REAL outcome measure is how many people give up smoking (or, at least, cut down). A surrogate measure might be the number of people who said that the adverts made them think about giving up smoking (regardless of whether or not they actually did do).**

**8. A similar flaw involves extrapolating inappropriately from lab findings to real-world scenarios. An experimental manipulation may cause people to change their behaviour in some small way in the lab, but this doesn’t mean it will necessarily have a big effect on their “real-world” behaviour.**

**9, Another similar flaw is extrapolating from the particular group studied (typically students) to the population in general. Often, this isn’t a problem (e.g., when we’re measuring something on which students and non-students wouldn’t be expected to differ). But if there is reason to believe that the group studied are not typical of the general population, we can’t extrapolate findings to them.**

**10. Torture the data (p.210). Suppose that an experimental manipulation has had no effect. If you analyse the data in enough ways (e.g., splitting it into different groups on a completely arbitrary basis), you are bound to find two groups that differ significantly, just by chance. NB: Sometimes it makes sense theoretically to analyse the data separately by (for example) gender. But if there is no apparent theoretical motivation for doing so, you should be suspicious (again, this is something dealt with at length in *Bad Pharma*).**

**11. The best of five...no seven...no nine!. (p.210) “If the difference between your drug and the placebo becomes significant after 4 ½ months into a 6 month trial, stop the trial immediately and start writing up the results: things might get less impressive if you carry on. Alternatively if at six months the results are “nearly significant”, extend the trial by another three months”. This is pretty hard to spot in a paper, but sometimes there will be signs (e.g., if one group is bigger than another for no apparent reason, or the study includes lots more participants than normal for that area of research).**

**12. Biased interpretation of results. If a theory predicts that all participants will show some particular behaviour and, in fact, (say) 70% do, the extent to which the prediction is unsupported is unclear. The statement “the majority of the participants behaved as predicted” would technically be true, but (arguably) the prediction is not supported.**

**13. Misleading percentages (p.256). Suppose that over the course of a year the rate of cocaine use amongst primary school children increases from 1 per every 100 children to 2 per every 100 (see p.263). There are two ways of reporting this increase. In absolute terms, the proportion of children taking cocaine has increased by 1 “percentage point” (i.e., from 1% to 2%), and this may well not be a statistically significant increase. In relative terms, the proportion has increased by 100% (i.e, it has “doubled”). Guess which figure newspapers (and, to some extent, journal articles) are most likely to report.**

a) Full competence theories of language acquisition predict that children will make no errors when producing questions. As predicted, children showed impressive performance, with a correct-production rate of over 90% (M=92%, SE=2.25)

b) An investigation into “the relation of quality and quantity of illumination to efficacy in the industries” found that increasing the brightness of the factory lights increased productivity.

c) The government’s campaign to encourage people to decrease their energy usage has been a success. The proportion of people who “strongly agreed” with the statement “I try to use less energy that I did this time last year” increased from 70% to 80%.

d) Freud argued that psychological conditions are caused by traumatic events that occurred during childhood. However, these events are often repressed, meaning that the individual has no awareness of them.

e) Working mothers damage their children’s future prospects! One study found that for every year that a mother worked during a child’s primary school years, the probability of the child getting 1 or more A level decreased by 10%.

f) Children from five age groups (8-9, 9-10, 10-12 and 11-12) completed either an *exam technique* training session, or a control training session that involved general reading. Overall, children in the *exam training* group did not significantly outperform those in the control group. However, the training session did show significant benefits for boys aged 11-12 and girls aged 9-10.

g) In an experimental study, children who played a violent computer game gave higher voltage (fictious) electrical shocks to an experimenter than children who played a neutral game. This demonstrates that violent computer games cause aggressive behaviour, and that banning such games would prevent tragedies like the Columbine High School massacre.

h) A handful of studies have shown a very small effect whereby participants who play a violent computer game are less likely to help an experimenter complete a puzzle than those who played a “control” game. We are aware of no studies that have failed to find an effect of computer game violence on helping behaviour.

i) When learning a new word, children assume that it refers to a whole object (e.g., a rabbit), rather than a part or property of that object. This is due to the whole-object assumption.

j) Language acquisition studies conducted on adult speakers (English Language students at the University of Sheffield) have demonstrated that (unlike children) adults never make interpretation errors with complex sentences like “The dog that the pig kicked was bitten by the horse”.

k) Extra sensory perception is real! After hundreds of failed studies, psychologists found that participants could reliably guess which of four cards an experimenter in an adjacent room was holding.

l) If you are over 50, you have a 4% chance of having a heart attack in any given year. But if you have high cholesterol, you are *50% more likely* to have a heart attack.

m) This study investigated whether children who were assigned to an exercise programme (*N*=107) lost more weight than waiting-list controls (N=24). As predicted, children in the exercise condition lost significantly more weight (*M*=1.55 kg, *SE*=0.21) than children in the control condition (*M*=1.35, *SE*=0.22, *p*=0.049).

**References**

Goldacre, B. (2009). *Bad Science*. London: Harper Perennial.

Goldacre, B. (2012). *Bad Pharma: How drug companies mislead doctors and harm patients.* London: Fourth Estate.

NB: You will notice that these sheet does not contain the answers, in order to enable its use in

tutorials. For the answer sheet, please email [Ben.Ambridge@Liverpool.ac.uk](mailto:Ben.Ambridge@Liverpool.ac.uk)