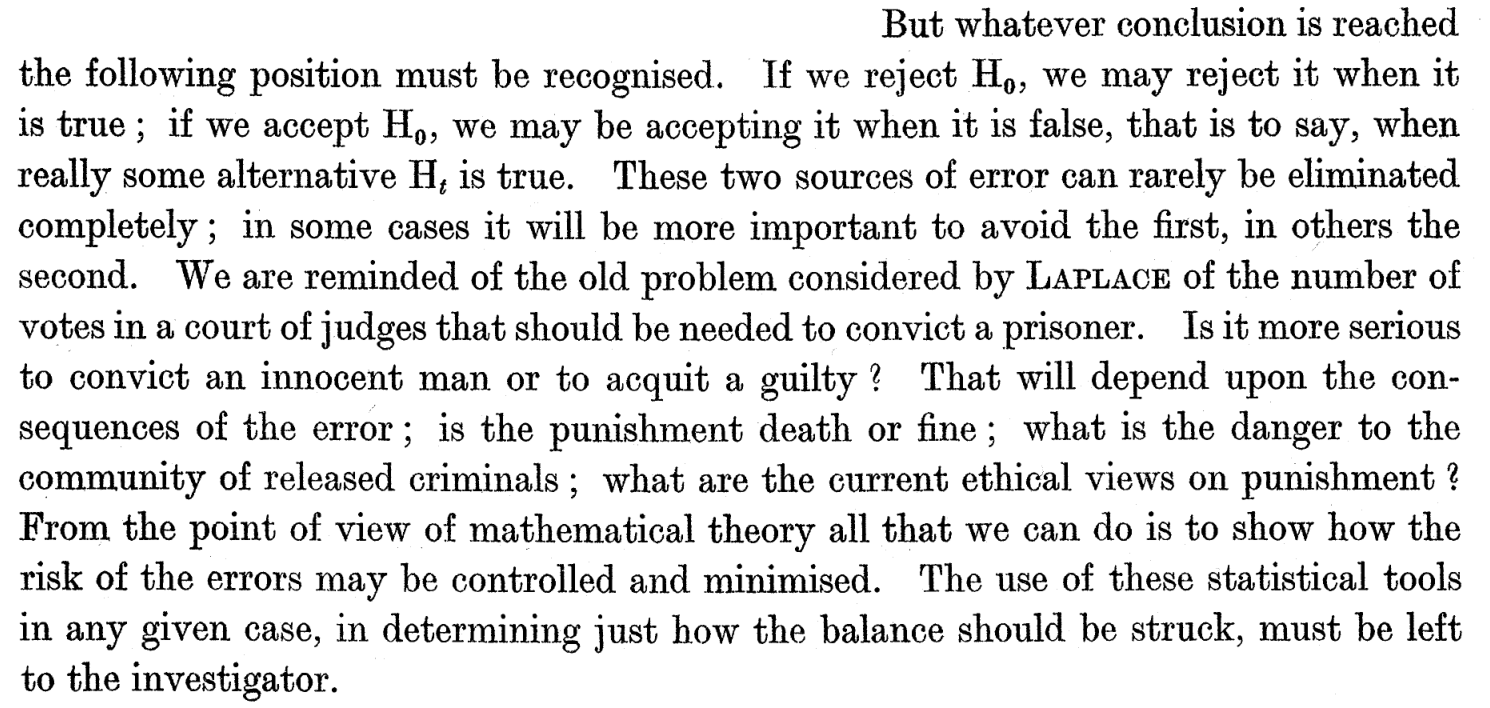
The New Heuristics

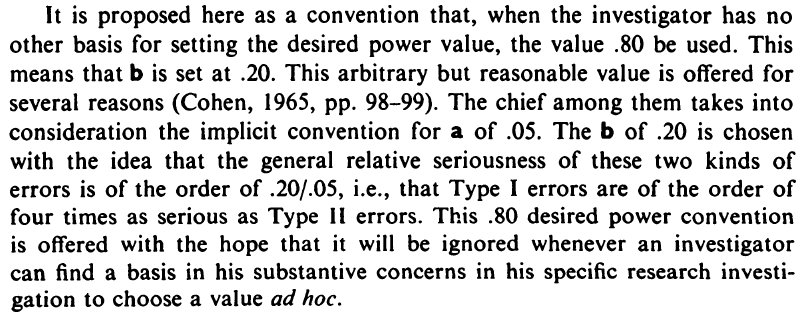
You can derive the age of a researcher based on the sample size they were told to use in a two independent group design. When I started my PhD, this number was 15, and when I ended, it was 20. This tells you I did my PhD between 2005 and 2010. If your number was 10, you have been in science much longer than I have, and if your number is 50, good luck with the final chapter of your PhD.

All these numbers are only sporadically the sample size you really need. As with a clock stuck at 9:30 in the morning, heuristics are sometimes right, but most often wrong. I think we rely way too often on heuristics for all sorts of important decisions we make when we do research. You can easily test whether you rely on a heuristic, or whether you can actually justify a decision you make. Ask yourself: Why?

I vividly remember talking to a researcher in 2012, a time where it started to become clear that many of the heuristics we relied on were wrong, and there was a lot of uncertainty about what good research practices looked like. She said: ‘I just want somebody to tell me what to do’. As psychologists, we work in a science where the answer to almost every research question is ‘it depends’. It should not be a surprise the same holds for how you design a study. For example, Neyman & Pearson (1933) perfectly illustrate how a statistician can explain the choices that need to be made, but in the end, only the researcher can make the final decision:



Due to a lack of training, most researchers do not have the skills to make these decisions. They need help, but do not even always have access to someone who can help them. It is therefore not surprising that articles and books that explain how to use useful tool provide some heuristics to get researchers started. An excellent example of this is Cohen’s classic work on power analysis. Although you need to think about the statistical power you want, as a heuristic, a minimum power of 80% is recommended. Let’s take a look at how Cohen (1988) introduces this benchmark.



It is rarely ignored. Not that we have a meta-heuristic here. Cohen argues a Type 1 error is 4 times as serious as a Type 2 error, and the Type 1 error is at 5%. Why? According to Fisher (1935) because it is a ‘convenient convention’. We are building a science on heuristics built on heuristics.

There has been a lot of discussion about how we need to improve psychological science in practice, and what good research practices look like. In my view, we will not have real progress when we replace old heuristics by new heuristics. People regularly complain to me about people who use what I would like to call ‘The New Heuristics’ (instead of The New Statistics), or ask me to help them convince someone they are rigidly but incorrectly applying a new heuristic. Some examples:

People who used optional stopping in the past, and have learned this is *p*-hacking, think you can not look at the data as it comes in (you can, when done correctly, using sequential analyses, see Lakens, 2014). People still use measures because previous studies used these measures, and ignore that there is no validity evidence (even though invalid measures make a study pretty much meaningless). They make directional predictions, but test them with two-sided tests (even when they could pre-register their directional prediction). They think you need 250 participants (as an editor of a flagship journal claimed), even though there is no magical number that leads to high enough accuracy. They think you always need to justify sample sizes based on a power analysis (as a reviewer of a grant proposal claimed when rejecting a proposal) even though there are many ways to justify sample sizes. They argue meta-analysis is not a ‘valid technique’ only because the meta-analytic estimate can be biased (ignoring meta-analyses have many uses, including an analysis of heterogeneity, and some bias detection techniques work under some circumstances). They think all research should be pre-registered, or published as Registered Reports, even when the main benefit (preventing inflation of error rates due to flexibility in the data analysis) is not relevant for all research psychologists do. They think *p*-values are invalid and should be removed from scientific articles, even when in well-designed controlled experiments they might be the outcome of interest, especially early on in new research lines. I could go on.

Change is like a pendulum, swinging from one side to the other of a multi-dimensional space. People might be too loose, or too strict, too risky, or too risk-averse, too sexy, or too boring. If you can’t justify your decisions, you will just be pushed from one extreme on one of these dimensions to the opposite extreme. What you need is the weight off a solid justification to be able to resist being pulled in the direction of whatever you perceive to be the current norm. Learning The New Heuristics (for example setting the alpha level to 0.005 instead of 0.05) is not an improvement – it is just a change.

If we teach people The New Heuristics, we will get lost in the Bog of Meaningless Discussions About Why These New Norms Do Not Apply To Me. This is a waste of time. From a good justification it *logically follows* whether something applies to you or not. Don’t discuss heuristics – discuss justifications.

‘Why’ questions come at different levels. Surface level ‘why’ questions are explicitly left to the researcher – no one else can answer them. Why are you collecting 50 participants in each group? Why are you aiming for 80% power? Why are you using an alpha level of 5%? Why are you using this prior when calculating a Bayes factor? Why are you assuming equal variances and using Student’s *t*-test instead of Welch’s *t*-test? Part of the problem I am addressing here is that we do not discuss which questions are up to the researcher, and which are questions on a deeper level that you can simply accept without needing to provide a justification in your paper. This makes it relatively easy for researchers to pretend some ‘why’ questions are on a deeper level, and can be assumed without having to be justified. A field needs a continuing discussion about what we expect researchers to justify in their papers (for example by developing improved and detailed reporting guidelines). This will be an interesting discussion to have. I expect we will never see researchers write in their papers that ‘we expect the central limit theorem to hold’. For now, let’s limit ourselves to surface level questions that were always left up to researchers to justify (even though some researchers might not know any better than using a heuristic). In the spirit of the name of this blog, let’s focus on 20% of the problems that will improve 80% of what we do.

My new motto is ‘Justify Everything’ (it also works as a hashtag: #JustifyEverything). Your first response will be that this is not possible. You will think this is too much to ask. This is because you think that *you* will have to be able to justify everything. But that is not my view on good science. You do not have the time to learn enough to be able to justify all the choices you need to make when doing science. Instead, you should be working in a team of as many people as you need so that within your research team, there is someone who can give an answer if I ask you ‘Why?’. As a rule of thumb, a large enough research team in psychology has between 50 and 500 researchers, because that is how many people you need for one of them to be able to justify why research teams in psychology need between 50 and 500 researchers.

Until we have transitioned into a more collaborative psychological science, we will be limited in how much and how well we can justify our decisions in our scientific articles. But we will be able to improve. Many journals now require sample size justifications. Expert peer reviewers can help by pointing out where heuristics are used, but justifications are possible (preferably in open peer review, so that the entire community can learn). The internet makes it easier than ever before to ask other people for help and advice. And as with anything in a job as difficult as science, just get started. The #Justify20% hashtag works just as well for now.