

Legend

1. [REDACTED] means that the original word/fragment was deleted to ensure the anonymity of the participants.
2. [?] is a placeholder for words/fragments that could not be transcribed.
3. (?) means that the transcriber was not completely sure what the last word/fragment was, but had a guess.
4. Sentences that begin with “I:” were said by the interviewer
5. Sentences that begin with “P:” were said by the participant

Block 1: General Information

I: Ok. So, then we will start with the first block now. The goal of this block is to get some general information about you. So, the first question is: **Are you a PhD Student?**

P: Yes.

I: Ok, and what is your field within psychology? For example social psychology, cognitive psychology...?

P: Neuropsychology or Neuroscience.

I: Ok. And did you conduct any experiments including a Stroop task in your career?

P: Eh no.

I: Do you have any -

P: But I did conduct experiments but not with the Stroop task.

I: And do you have any experience with the Stroop task?

P: Yeah, I am pretty sure I have participated in studies that had the Stroop task. I knew what it was and I know what it feels like to do it. But I can't remember the specific times that I conducted - or participated in an experiment but I am pretty sure that I have participated.

I: Ok. And which statistical analysis programs do you use at least once a week? Multiple answers are possible. For instance, SPSS, R, Stata, SAS, Matlab, Python, or any other?

P: Matlab and R.

I: Ok, thank you. And how would you rate your knowledge of statistics relative to your peers on a scale from 1, which means extremely poor, to 10, excellent?

P: My peers - what are my peers? Because we are in the [REDACTED] Department - so compared to my peers I think -

I: So, the definition is: Relative to other researchers or scientists in your field.

P: Yeah. So, I think it is pretty good. So maybe an 8.

I: Ok. And how confident are you that your fabricated data will go undetected as fabricated? Again on a scale from 1 to 10 where 1 means extremely insecure and 10 means extremely confident.

P: Ehm, I am not aware of the methods that you use to detect these false data. So that makes me a little bit less confident. But I think it is still a 6.

Block 2: Timeline of Data Fabrication Process (When?)

I: Ok, thank you. Then this is the end of the first block about general information. Now, we will start with the second block. The goal of this block is to get some information about the timeline of the data fabrication process. So, did you fabricate the data in one day or spread the data fabrication over several days?

P: Ehm, I did most of it in one day and then finished it a little bit - on a different day.

I: Ok. Ehm and so, the number of days that you worked on fabricating the data were two or?

P: Well, I spread it over two days but I didn't work two days full-time on it.

I: Ok. And how much time do you estimate that it took you to fabricate the data in their entirety?

P: I think maybe two hours.

I: Two hours?

P: Yeah, I guess.

I: And how much effort do you feel you invested in fabricating the data on a scale from 1 (no effort at all) to 7 (a lot of effort)?

P: Ehm, well not that much effort, I think. 3.

I: 3, ok. Did you prepare in any way before starting to fabricate the data?

P: Yes, I looked up in what kind of scale the reactions times are. So, I tried to find papers in which they just did a Stroop task without any additional questions or just the Stroop task with congruent and incongruent and just looked at how much sort of the reaction times and the standard deviations were. These two things together gave an idea in what range I should make my data.

I: And how much time do you estimate you spent on preparing?

P: Maybe half an hour.

I: Half an hour, ok. And did you also read literature on detecting data fabrication?

P: No. But I thought about doing that but, yeah, took a little bit too much time to do it for this purpose.

I: Ok. And did you look into previous cases of data fabrication or so?

P: No. I thought it was a little bit cheating also to do it. But I was curious what methods you use - and probably try to get around it but, yeah, it was too much effort to do it. So, I thought I just simulated the [data?] in straightforward fashion.

I: And - so, you said that you looked up the scale of like previous response - or like the response time in previous studies - and how did this influence your approach to fabricating the data?

P: So, it influenced my - just I wanted to draw a sample from a normal distribution with some mean and some standard deviation depending on if it is a congruent or an incongruent condition. And I didn't know what kind of reaction times to expect. So, I just looked at how much is the reaction time in general. If I could find an appropriate mean and standard deviation to fabricate the data.

I: Ok, then this is the end of the second block. Do you have any other comments about the timeline of the data fabrication process that you think could be interesting for us to know?

P: Ehm, I don't think so.

Block 3: Broad Framework of Data Fabrication Process (What?)

I: Ok, then, we will now start with the third block. The goal of this block is to get some information about the broad framework of the data fabrication process and the first question is: Could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: Ehm, yes probably. To make it look more fabricated... Maybe if the difference between the conditions is equal for all participants - or between conditions that would make it look fabricated, I guess. And yeah probably if it is maybe too normally distributed so maybe in fabricated data you would not expect any outliers and in real data there might be outliers. Yeah. So it might be have skewed distributions in some directions.

I: Ok. And could you name specific characteristics that would make data look genuine or more genuine in your opinion?

P: What do you mean - genuine?

I: Like more true, more real.

P: Ehm. Yeah if it is appropriate values that are to be expected of some experiment. And if .. yeah if the distribution of the values are what you would expect. So are normally – or probably mostly normally distributed – even though this was also my answer for the previous question. Yeah. So I would expect that if you do an experiment like this where you have many trials – so 30 trials – then the means would be normally distributed because you have so much trials. But if you look at a trial-to-trial basis then I think the distribution would look different, would look less normal for each individual but over all individuals the distribution of the means would still look normal. So that is why I generated my data in a normal way even though I thought that it would probably for an individual not look so normal.

I: Ok. And did you take these characteristics you just mentioned into account when fabricating the data?

P: Ehm so I did take into account that I think that the data would be normally distributed over all participants. So the means of the participants would be normally distributed. But I did not take into account that the trials per individual would be skewed probably.

I: Ok. And did you take into consideration relations in the data other than the Stroop effect itself?

P: Ehm yes. So, I would expect maybe to find gender differences but I could not find - and when I looked at literature – and I did not find it or many gender differences, I think. So then it didn't take into account. Then I thought there are probably some individual differences that some people are just better in the Stroop task than others. So, I did take into account that each individual has a like personal score how well he or she performs in the Stroop task.

I: Ok. And what criteria did you use to determine whether you thought your fabricated data would go undetected?

P: Ehm, I didn't really take any additional measures other than to look at the values and see like do they look sort of in the range that I expect and, yeah, **so I did not try to do any tests or anything to try and find my own false data.**

I: Ok. And in hindsight, are there things you think you should have paid specific attention to while fabricating the data?

P: **Yeah, so** I did take the individual ... eh I did have an individual score but I noticed that in the test, the t-test that you did was sort of a paired t-test over the Stroop effect, I think. And the individual scores don't matter for that. So, I didn't pay too much attention to the size of the individual differences because I thought, well, this doesn't really matter anyway for the – at least for the statistical test that you do with it. But it could be - you might look at some different things that you didn't put into the Excel sheet. So there is a t-test in the Excel sheet but you might look at different things and I didn't paid attention to that. So, I could have paid more attention to individual differences.

I: Ok, then this is the end of the third block. Do you have any other comments about the broad framework of the data fabrication process that you think could be interesting for us to know?

P: No.

Block 4: Specific Steps of Data Fabrication Process (How?)

I: Then, we will now start with the fourth block. The goal of this block is to get some information about the specific steps of the data fabrication process. So, could you indicate what steps you took to fabricate the means for the participants?

P: Ehm, yes. As I mentioned I looked at appropriate means for just young individuals with the Stroop task in the congruent and incongruent condition. I can look at what I came up with... I think I had a mean of 545 with a difference between the conditions of 125 milliseconds. So, that is what I used as the mean. And I did add some individual noise to the means per participant which added random noise with a mean of 0 and a standard deviation of 20 for each individual.

I: Ok. And how did you add the noise?

P: Eh, so I just drew from a random distribution with Matlab – eh from a normal distribution, I mean. So for each individual, I took a standard normal distribution and multiplied it with the standard deviation and added the means.

I: Ok. And could you indicate what steps you took to fabricate the standard deviations for the participants?

P: Eh yeah, so I kind of answered it already but well I took a standard deviation of about 70 which was a little bit lower for the congruent than for the incongruent condition. I took a difference of 6 meaning that the congruent was 6 lower (so 67) than the incongruent [which was?] 73 as a standard deviation. So I drew from a standard normal distribution and multiplied it with the standard deviation.

I: Ok. And did you also add random noise here for the standard deviations?

P: No, I don't think so. No. So only – there was no individual noise for the standard deviations.

I: Ok. And did you repeatedly fabricate data until you were satisfied with the results?

P: Eh, I did vary with the means and the standard deviations, yes. So I did repeatedly fabricate the data.

I: Ok. And how did you determine whether you were satisfied with the fabricated data or that they needed to be adjusted?

P: Well, at first, I thought the effect size was really big. So I tried to make it like more exciting. So the t-score was so big that I wanted to have a little bit less difference between the groups to make the research appear more interesting. But on the other hand I thought that it must be a pretty strong effect [in?] the Stroop task [with?] a congruent and incongruent because it is just pretty difficult to do the incongruent task while the congruent is quite easy. So I would expect quite a big difference and you had quite a big sample of 30 trials per participants and 25 participants. And also a within-subjects design so you have a paired t-test which has quite a lot of power to find this difference. So, I was not really surprised by having such a big difference but I did start out with a bigger difference at first and then made it smaller. But it is still a big effect.

I: Ok, and did you have other criteria to determine whether you were satisfied?

P: Yeah, so I did look at the standard deviations to see if they remained close to [the – like if?] the standard deviation remained close to what I thought it should be - like the 70 plus/minus 6 – or plus/minus 3 that I generated. And it appeared to look - to be in that area and I thought it was good enough, so.

I: Ok and did you try to inspect whether the fabricated data looked weird?

P: Ehm, I did not do any specific tests to look at if they were weird or not. I did make a histogram but – well, I did make a histogram. So that might be – yeah that was probably to look at [weird things?].

I: Ok and did you try to inspect whether the fabricated data looked genuine - or?

P: Ehm yeah. Well, no, no specific steps for that either. So, ehm.

I: Ehm. Yeah sorry?

P: So, that was what I had constantly on my mind. To make it look like real data. I kind of had it in the back of my mind all the time but no specific steps for it.

I: Ok, and how many different mean-sd combinations did you fabricate before getting to the final fabricated dataset?

P: Hmm. Probably quite a bit. Maybe 5, 10. But I did not after each time inspect my data thoroughly. I just looked it at and thought well, just throw in a different one and go ahead and run it again.

I: But why did you do it again then? You mentioned that it was partly because of the too large effect size?

P: Yeah, that was at the beginning - was mostly my concern that I thought it was too big. So, but I did try to run a t-test to see if it was the same value as the one coming out of the Excel sheet but it wasn't. And then I noticed that you had a different way of calculating the effect and then I did not bother to replicate it and to find out for all the means and standard deviations what it

was. So I just thought well, it makes sense that it is big – that it is that big. Well, I just thought - after all, I just accepted that it is such a big effect.

I: And besides the supplied spreadsheet, did you use any other computer programs to fabricate data?

P: Matlab.

I: And did you use a random number generator to simulate data during this study?

P: Ehm, yes. Well, yes. I drew randomly from a normal distribution so that uses a random number generator in the background. So I did not – so, I guess, yes.

I: Ok and did you use real data during the fabrication process?

P: No. Just looking at the means of real data.

I: Ok, then this is the end of the fourth block. Do you have any other comments about the specific steps of the data fabrication process that you think could be interesting for us to know?

P: No.

Block 5: Underlying Rationale of Data Fabrication Process (Why?)

I: Then, we will now start with the fifth block. The goal of this block is to get some information about the underlying rationale of the data fabrication process. So, did you consider fabricating these data a difficult task to complete?

P: Ehm, well, ehm, no not really. All I thought the way I generated it was kind of – I started out with doing it like this and then I thought of all these things like should I add random noise and vary the standard deviations within a person and add a gender effect and make it – eh, add outliers and stuff. And then I kind of thought, well, you probably won't notice that too much anyway except for maybe the standard deviations might be a little bit bigger for some people when there is outliers – but I didn't think that it would affect it enough to make it worth it. And also I thought, well if we were replicating like with young students or something they probably do the task pretty seriously. And I thought perhaps it is not that strange to not have any outliers. So then I just kind of stuck with my original simple plan of drawing from normal distributions with appropriate means and standard deviations.

I: Ok. And do you think that your approach to data fabrication will be difficult to detect as fabricated?

P: Ehm, I kind of think so. But that would mean that most data that is fabricated is pretty difficult to detect but I kind of think that that might be the case. So, I think it is quite difficult.

I: And can you say why you think that it is difficult to detect it?

P: Ehm, because I think that if you have these samples of quite a lot of trials then the means of those subjects would look quite normal. Probably normally distributed. So, I think it is hard to distinguish it from an actual normal distribution [?]. Yeah. So, I think it is quite a difficult task to do because yeah ... yeah.

I: Ok. And why did you decide to participate in this study?

P: I thought it was an interesting question of how to fabricate data and to think about it and yeah. I thought it was fun to do. And, yeah, we also get the reward, of course, which is also nice.

I: Ok. And did you discuss this study or the fabrication of the dataset for this study with other people?

P: Yes, so I asked a colleague who uses Stroop tasks also with healthy participants as a control. And I asked what her standard deviations and means were.

I: Ok, and did these people help you in fabricating the data?

P: No.

I: Ok, then this is the end of the fifth block. Do you have any other comments about the underlying rationale of the data fabrication process that you think could be interesting for us to know?

P: I don't think so.

I: Ok, then this is the end of the interview or is there anything else you can recall about the data fabrication that you think is worth mentioning?

P: I don't think so.

I: Ok. Then, I will turn off the recording.