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 participat

Block 1: General Information

I: So, now we will start with the first block. The goal of this block is to get some general information about you. Just ..., yes, still working. Ok, so, the first question is: Are you a PhD student?

P: I am a postdoc.

I: A postdoc. So, how many years has it been since you got your PhD?

P: One - not yet. I did it in May last year.

I: Ok. And what is your field within psychology? With field, we mean for instance like social or cognitive or ...?

P: Experimental clinical.

I: Ok. And have you ever conducted any experiments including a Stroop task in your career?

P: No, no Stroop task.

I: Ok. Have you heard of it or read papers with the Stroop task? Could you describe your experience with the Stroop task a little bit?

P: Yes. I studied Experimental Psychology and in this Master Program, we got to know the Stroop task quite well. So, I know it quite well. And then in - during my PhD, I was working in a research group, in which some people also made use of the Stroop task. So, there I also got some information, some background about it. So, there - that's where I know it from. So, I know the task - the general idea quite well, but I haven't used it myself.

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

P: Mostly just SPSS. And then sometimes maybe Excel for like a quick t-test or some online calculators or something, but mostly just SPSS.

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

P: I think 5, quite intermediate.

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: I think about 4.

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

I: 4, ok. Thank you. Then this is 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, the first question is: Did you fabricate the data in one day or spread the data fabrication over several days?

P: I spread it over several days.

I: Ok. And on how many days did you work on fabricating the data?

P: I think, about 3 or 4.

I: 3 or 4 days, ok. And how much time do you estimate that it took you to fabricate the data in their entirety?

P: I think like maybe 2 to 3 hours.

I: Ok. 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: Like 4, I think.

I: 4, ok. And did you prepare in any way before starting to fabricate the data?

P: That's a difficult question. I started doing it right away but I also looked up for instance like what could be like average reactions times in a Stroop task. So, I suppose that's count as preparing.

I: Yeah, so for instance, a follow-up question would be: Did you read literature on detecting data fabrication?

P: No, no, I didn't.

I: Ok. Or did you look into any previous cases of data fabrication and how they had been detected or?

P: I read a general article in The Guardian, I think, about data fabrication and I read the book of Diederik Stapel. So, I have a very broad idea of how other cases were detected but no specific, scientific articles, just popular - like general audience articles.

I: But so you said you looked up previous articles on the Stroop task to get an idea about what would be an average response time?

P: Yeah, exactly.

I: And then you used that for your approach to fabricate the data?

P: I already had an idea about the average reaction time, so, but I just checked it to validate it and it was about the same. So, and I think it varies quite over experiments so I didn't adjust it.

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

P: No, but - just that I did it sort of iteratively. So, I started filling in all the things and then after a while I came back to it to adjust some things so I did it like over a course of several days to start it and then think about things and then come back to it to adjust a little bit further, yeah. Specifically - or maybe that comes back later in the experiment, but I checked like, for instance, the correlation between the two conditions - congruent and incongruent - and I saw that there - it was very large so I adjusted the data a bit to decrease the correlation, for instance.

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

I: Ok, thank you. 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. Could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: Yes. I think large effect sizes. Also probably inflated correlations or (?) correlations, spurious correlations, for instance correlations between standard deviations, I think, that you wouldn't expect but are there. And that might indicate fraud as well. Yeah, what else? I think if you generate it with a statistical program like a too nice distribution that is a very neat, nice distribution whereas maybe real data would have a bit of a skewed distribution, maybe. I think, effect size, weird correlations, and weird distributions might be the three important ones.

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

P: Yeah, I think the Stroop task was a bit - yeah, I think acceptable reactions times. Another - too big effect size, for instance. No inflated correlations, I would say. That is what I focused on anyway.

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

P: Yes, I tried to not inflate the effect size too much. And as I said, I tried to reduce the correlation between the conditions and then made sure that there wasn't like a large correlations for instance between the standard deviations. Because - I am not sure if that is correct actually, so I did this mostly on hunches on my idea about data but I wouldn't expect that there would be correlations between standard deviations but there would between reaction times.

I: Ok. And - as I said the questions may sometimes be a bit repetitive, so yeah - did you take into consideration relations in the data other than the Stroop effect itself?

P: Yes. So, I considered the standard deviation which I - yeah, I checked for the correlation betwen the standard deviation.

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

P: Yeah, again, focusing on not too large correlations, not too big effect size, and making sure that the standard deviations and the reaction times are acceptable, that they are not too big or too small to be genuine or to be real.

I: Ok. And had you specific, different criteria for the means and the standard deviations?

P: Yes. So, mean reaction times should correlate, standard deviations shouldn't correlate or shouldn't correlate highly - shouldn't correlate at all [?] maybe.

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

P: No, I am not sure. I am trying to figure out how you would detect it but I don't know. Something that I would like to add is the fact that I also made sure that people with higher reaction times have somewhat larger effects - so bigger difference between congruent and incongruent conditions. Because that is also something I remembered that it seems - that the effect seems to be a bit more pronounced when latencies are longer.

I: Ok, and you - and did you take that into account or you think that you maybe should have taken it into account?

P: I did, I did take that into account. But I forgot to mention it before.

I: Ok, thank you. Then this is the end of the third block. Do you have any other comments about the broad framework of the data fabrication proces that now you come to your mind or you think could be interesting for us to know?

P: No - yeah, just that I - I just created the reaction times myself, I didn't generate them. And just to summarize, I checked for the correlations between the conditions, or (?) the correlations between the standard deviations, and for the correlations between effects and latencies from a subject.

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

I: Ok. Let me check whether it is still running - it is, ok. And, 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: Yeah, I have an idea of what an acceptable reaction time in a psychological experiment are and, for me, that is around 5-600 milliseconds and I used this to generate the reaction times myself. So, I didn't use a program or something to generate data. And then I thought that the standard deviations would [?] be around 100-120 milliseconds so I also took those and generated them for one of the conditions and then I copied the reaction times to the incongruent condition and then I added like a few tens of milliseconds or yeah - so, sometimes I also made sure that people didn't show the effect. So, not everybody showed the effect. And then, for the standard deviations, to make sure that there is no big correlation, I copied the standard deviations of the last 12 people that I made to the first 12 people in the incongruent condition and vice versa so that there wouldn't be like a correlation because I might be biased if I had to generate it myself to still, yeah, have a correlation there.

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

P: Yeah, I adjusted the data I fabricated first several times until I was satisfied with the results.

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

P: First of, I checked the effect size whether it was acceptable. So, I decided that like a t-value around 2 and a half to 3 or something would be acceptable for this sample size. And then once I had that, I checked the data in a new Excel file and I correlated - aeh, I calculated the correlations using Excel and then just adjusted the reaction times a little bit to adjust for the correlations.

I: Ok and the range of the desired p-value - aeh t-value that you had - what was this based on - on a general feeling or on your knowledge of the Stroop task or . . . ?

P: Yeah, just on my impression what would be acceptable. Yeah, I think it can vary a bit but, yeah, I didn't really have a concrete idea how big it would be.

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

P: Yeah. So - but just using correlations and checking the reaction time difference between the congruent and incongruent condition, but I didn't use a statistical program to check for the distribution or something.

I: Ok, so you did it by eyeballing or?

P: Yeah.

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

P: Yeah, so basically on my feeling and on my, yeah, my knowledge about what reaction times look like generally but not using statistical programs.

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

P: I just generated it once, but then I adjusted it like 5 times or something until I was reasonably satisfied or (?) until I didn't really have an idea anymore how to improve it.

I: Ok, and you did this over several days, right?

P: Yeah. So, in about - not consecutively but spread over the two weeks now since I was admitted to the study, I took like four days to adjust the data or something - three, yeah, three or four days.

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

P: No, just Excel to additionally calculate correlations.

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

P: No, I didn't.

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

P: No, I considered it but I didn't, no.

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, I don't think so.

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

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

P: Yeah, it was surprisingly more difficult than I anticipated so if you know your - it is going to be checked, then you try to do it really thoroughly but, yeah, in hindsight I should have probably read some of the papers on how to detect it

and maybe improve my - yeah, just my data accordingly. But I didn't really have an idea how you would try to check it, so.

I: Ok. And what then make you feel that it is hard task - could you describe that a bit more?

P: Yeah. Because I started first with the generating the reaction times just myself. But then I noticed that I have some biases myself or that I don't know exactly how real data is supposed to look like. And then, yeah, I tried to adjust it but I was just - I noticed that I just didn't have enough background or information or knowledge about how real data structure looks like to be sure that it looks like a real one. So then, yeah, once I generated it, I always started to worry about how it should exactly look like.

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

P: Yeah - I am not sure, I can't really say. It depends on which like - how you will approach it and I think it will depend on that. Yeah, I am sort of in between whether it will be easy to detect my data or not. I feel like they are quite genuine but I also have the impression that there might be some easy things in there to catch that it is actually not real data.

I: Do you like - can you describe more specifically what you think could be detected or?

P: What could be detected in my data?

I: Yeah.

P: There is still a very high correlation between the reaction times - between the congruent and incongruent condition which is .95 which is, I think, too much. So, I think that is a bit of a catch. The data in the two conditions is too similar to be genuine.

I: Thank you. Then the next question is: Why did you decide to participate in this study?

P: I thought it was interesting but also the participation fee was really high so I thought it was nice. And I took it also as a bit of a challenge to try and do it. So, I thought it was just an interesting study. And I just started working here so I had a little bit of time to actually do it. So, that's why I decided to . . .

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

P: Yes, with my colleague. And also the study was briefly mentioned in our lab meeting.

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

P: No. A colleague of mine is participating as well but we decided to try and do it as independently as possible. Of course, we discussed it a little bit but still we

tried to do it independently. So, nobody helped.

I: Ok. So what did you discuss?

P: So, I talked about the issue that the correlation is quite high and we discussed like different ways of approaching it. I think she would try to use a random number generator, something in R to use to simulate data.

I: Ok, thank you. 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: No, 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: <u>I think it is just worth mentioning that it is quite hard and it also feels that with some motivation you can do it better, but then it also feels like it is like it is quite laborious to generate it. So I think that if it becomes so laborious to just try and generate youre data, it is probably even just easier to collect real data instead of generating false data for 35 people.</u>