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. So, the first question is: Are you a PhD Student?

P: Eh, yeah. Well, I just finished my dissertation but yeah, officially, I am still a PhD student.

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

P: Then, I would say cognitive psychology.

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

P: I have but not for my dissertation. So, for example, I used to be a research assistant as well and I collaborated on collecting Stroop data. But I have not used the studies myself in my publications.

I: Ok. So, could you describe a bit like what your experience or knowledge about the Stroop task is? Like for instance, you said you conducted experiments but did you also analyse the data yourself?

P: Yes, yes. Of course with regular Stroop task, the analysis is quite straightforward, right. It is just comparing the incongruent to the congruent trials and calculating this congruency effect. I have never used, like for example, an emotional Stroop task or other adaptations of the Stroop task. I have only worked with the regular Stroop task. So, in experimental settings such as as a research assistant but also as a clinical interviewer that I used to - I used to work at the [REDACTED] for the [REDACTED]. I am not sure if you are familiar with it. [REDACTED]. And there we had the classical Stroop task with reading out words from a piece of paper. I also collected those data.

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: Right. So at least once a week, I would say R and Python.

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: A 5 being average?

I: Ok - yeah.

P: I don't know. Probably a 7 or something.

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 don't know. Probably a 5. I don't have the feeling that - I mean they are all knowledgeable people who were trying to fabricate these data, of course. So, I would say a 5. I don't think I would do any better or worse than my peers.

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

I: Ok. 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, did you fabricate the data in one day or spread the data fabrication over several days?

P: No, I did it in one day.

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

P: I guess about 3 hours. So, part of what I did was I had to find a large sample data set of Stroop data and that took a while to find on the internet. Because my own data is like with small groups and I wanted a large data set. So, I had to look on the internet to find one.

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: I would say 4. I mean it was more difficult than I expected. It took some thinking as to what would constitute a good fabricated data set. But, yeah, I would say a 4 out of 7.

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

P: No, not really.

I: Ok, so like you didn't read literature on detecting data fabrication?

2

P: No, no. I guess I should have. That is a good idea. But I didn't, no.

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

P: No, I just - I read a few meta-analyses on the Stroop effect. Hopefully, that gave me some insight. But no, I did not look at the literature on data fabrication itself.

I: Ok. So you said that you like started first with like thinking about how to do it, right?

P: Ja.

I: So, what were the different approaches that you considered?

P: So, first I was like, well, what I of course could do is just look at my own old data. Just look at the means and standard deviations or maybe do even the calculations per subject. And just try to - I don't know - throw some noise over it or something. But, yeah, then I thought using a large data set and treat it as some kind of population and sample from that 25 subjects would be a better way to maybe avoid detection. Yeah, that was my reasoning.

I: And did this preparation/reasoning influence your approach to fabricating the data?

P: Oh, yeah probably. Yeah, yeah definetely. I mean, the first option would be easier, I guess. But yeah, I think the approach that I took now would be less easy to detect.

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: No, the timeline was pretty straightforward. I just put a block of - I think - four or six hours in my calender and just devoted it to this thing.

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. So, the first question is: Could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: I guess, repetition of data would constitute - would be evidence for fabrication. I think the best indicator for fabrication would be means or standard deviations that would not be similar to what has been found in earlier studies. And yeah, of course, I also asked that myself: What would you guys use to try to detect data fabrication? And of course that have to be measures on the data, on the - like in the format that you actually wanted the data to be collected. So, yeah,

whatever measures you were using to detect fabrication it has to work on the means and standard deviations of 25 subjects. So, I thought means and standard deviations, maybe even cross-correlations between those measures. So, those were actually the measures that I used to fabricate my data.

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

P: I guess - I would probably say that distributions . . . . So, with data fabrication what you - what you would see, I guess, is that the distributions would be similar to a standard Gaussian or another standard distribution. Whereas, of course, real data is not perfectly distributed. So, I think you would see in genuine data deviations from perfect distributions.

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

P: The last - so, the distribution of the data I did not really take into account because in the sample data set that I found online on many subjects - all the measures looked pretty normally distributed. So, I did not go through the effort of - I don't know - adjusting my distributions. I just treated them as normally distributed. But I guess for a lot of measures you would not find data that clean.

I: Ok. And what about the other characteristics that you mentioned as indicators of fabricated data sets?

P: Yeah. I definitely took those into account. So, I looked at the - yeah, I made sure that my fabricated data set would show the same intercorrelation between all variables. The same mean and standard deviation of means and standard deviation and standard deviation of standard deviations.

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

P: No.

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

P: Yes, so like I said. I compared the characteristics of my data - so, the mean of means and the standard deviation of means, mean of standard deviations, and standard deviation of standard deviations and all the cross-correlations between those measures - and I made sure that they were quite similar to in the real data set that I found.

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

P: Well, I guess, if you wanna use bootstrapping you could probably get closer to your - to the characteristics of the genuine data that I looked at. But I did not do that. I thought that my data was good enough. But I am pretty sure that there are better ways to do it.

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, I think we get to the data fabrication later, right?

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

I: Yes. 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. So, I wrote down the whole procedure and - so what I did - and I will run you through what I did and then you can decide if I answered your question. So, what I first did is I went looking for a large data set on the Stroop task with a lot of subjects, more than 25. So, that took a while and then I found a sample Stroop data set with 121 subjects and I thought ok, that's enough. And what I then did is per subject calculate the means for the two different conditions and the standard deviations for the two different conditions - basically the format that you provided in your template - per subject. So, that left me with 121 means and standard deviations for two conditions. Then, I looked at the data characteristics. So, first I looked at the distribution of the data and like I said that looked pretty normal. So, I didn't go through the effort of doing anything weird with distributions or skewing it or anything. Of course, I looked at the mean of the means, the standard deviation of the means, the mean of the standard deviations of the - standard deviations. And then, I thought, well, what else could you guys use to detect fraud and then I thought, maybe there is something known in the Stroop task about the correlation between those measures. Probably, there is - I don't know. So, I also looked at the correlation between those four measures, right - mean and standard deviation from the two condition. And then I generated a new 121 sized data set with those characteristics. So, just in R - I just generated a correlated data set with the given correlations and the mean and standard deviation of the means and standard deviations. Giving me a 121 sized data set with those characteristics. From that data set, I sampled randomly 25 subjects. Or I think it was 25 that you want, right? I randomly sampled 25 subjects and I put them in your template. Then, manually, I checked for crazy outliers - like I don't know - maybe, there were response times of below 100 milliseconds and that would of course be not possible. But there weren't any. So, then I decided that that would be a good fabricated data set.

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

P: Yeah, it is like I said I just used a random number generator that of course just generates a Gaussian with a given mean and standard deviation. So, I did

not do anything special to generate mean and standard deviation. The only constraints were of course the correlations between the variables. But there is a nice R package to build correlated data sets.

I: Ok. And could you say from where you got the data set on which you based your fabrication?

P: Yeah, I just googled 'stroop.csv'. And at some point, I came up to this website that I think for a statistics course offered a sample Stroop data set collected on 121 students. So, I thought, well, that looks good.

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

P: No. It actually only took one try. Apparently, my random number generator is good enough to generate data sets that are pretty close to the constraints that I set. But of course, if the constraints wouldn't have been met, I would just have repeated the procedure and probably would have found a good data set.

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

P: Yeah, so I just looked at the mean and standard deviation and the mean and standard deviation of the mean and standard deviation of the sample data set, of my own data set and they looked pretty similar. I did not have a fixed threshold that I used but the numbers were pretty close. So, then I was satisfied. And then, yeah, like I said for the 25 samples I took I did not look at those other measures. Because I think, yeah, this would be more authentic, I think. Of course, 25 samples - 25 subjects doing the Stroop task is also some kind of subsample from a population, right? So, as long as there are no crazy outliers, I was pretty happy with the results.

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

P: Oh yeah, of course. I just checked the histrograms. But it all looked pretty ok-ish.

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

P: Well, I - no. I mean, I don't really have the gut feeling or instinct at the moment to determine by eye if something looks genuine or not. But, I mean, the means and standard deviations looked similar to the real data set. So, I was happy with that.

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

P: So like I said, just one. I generated the data set of 121 and from that I sampled 25.

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

P: I used R.

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

P: Yes.

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

P: Yeah. Well, the data set I found online. I mean I don't know if it is real. They said it was real. I have no way of verifying that but I take their word for it, yeah.

I: Ok. And so - but you - so, you did not directly like copy-paste it but you used it like a - more or less - as the baseline for your own fabrication?

P: Right. So, I only basically took the mean and standard deviation of the mean and standard deviation and used that for my random number generator to generate a new data set.

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: Not really. I am just very curious as to the methods that other people used. I mean, I was thinking about, maybe, I should use like an LCA model or an LBA model to generate data. And I thought: No, I am way overthinking it. It doesn't have to be that hard, I guess.

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

I: Ok. 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: Not really. I mean I guess I underestimated it at first. But in the end it was not that difficult, I think, no. It is a pretty straightforward method, I guess.

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

P: I think so. I think, I mean, there are only so many things you can calculate from the template that you gave. So, the only thing I can think of is distribution of data, correlation between data and means and standard deviations. So, I am very interested in what the methods are for actually detecting fabricated data. Yeah, I couldn't think of anything else to keep in mind.

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

P: Yeah, because it is, I guess, fun to do something that is really not allowed in science. And yeah, I mean, I think [REDACTED] also did it and was like

that sounds like a good challenge. <u>Wonder if I can beat [REDACTED] with generating data.</u> So, yeah, it seemed like a fun challenge.

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

P: Yes, but not the exact method that I used. So, I knew that [REDACTED] also were interested in participating but <u>I think we are too competitive to actually discuss our methods with each other.</u>

I: Ok, so these people didn't help you in fabricating the data?

P: No, definitely not, 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: No, I just walked you through my method. There is nothing more to say. If you have any questions, feel free to come back to me. I am willing to give you the R code that I used as well if you are interested.

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: No.