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: So, then 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: Ja. We call ourselves PhD candidates, but yes.

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

P: Cognitive science, I would say. [REDACTED]

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

P: I think, in my Bachelor, we did a very basic thing. But yes, I am familiar with the Stroop task.

I: Ok, so you conducted a like Stroop task experiment?

P: Yes, well, it was sort of an in-class experiment where we would record our times and so on.

I: Ok. And did you ever analyze Stroop task data or so?

P: Yes, in a limited way.

I: Ok, but so you would say that you are quite familiar with the Stroop task or?

P: Yes.

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: So once a week when I am analyzing data, but then (?) in the data analysis phase mostly R. Sometimes if I want to get a quick answer from something I use SPSS. I would say it is 70% R, 20% Matlab, and 10% 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 would say 8. I should

maybe add that I about half a year ago started applying Bayesian statistics, so I mainly (?) work with STAN, [?] R implementations, yeah.

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 would say 7. I did my best, I think, but I think that detection methods are - they take into consideration a lot more than I do.

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

I: Ok. 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: I think, it was more a spread over about 3 days.

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

P: This is hard to say, but including everything maybe 6 hours.

I: Ok.

P: But this is a rough guess. I didn't really clock the time. So including also getting a bit into theory and a bit for finding the right functions to do what I wanted to do. Maybe - yeah, I would say around 6, but it is - yeah.

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: 6.

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

P: Well, I mean fabricating the data takes about 20 seconds. So, in this - running the code and then pasting it into the Excel file. Maybe a few more times, because it is somewhat randomized, so I am - I was choosing something that was more valuable or looked better to me. So, the actual fabrication then took maybe a minute. Everything else was preparation or writing the code. [?]

I: Ok. And did you do something before you started to write the code?

P: Yes. So, I thought about it, thought about how do - well, how to make it look as natural as possible. I looked up some papers, reviews on the Stroop task to sort of see what the means and the standard deviations normally look like. And then I looked for a data set that was already available. So, I could describe the entire process if you like or?

I: For now, it is more relevant what happened like before you started to like write the code for your fabrication.

P: Right. I think, I just started by sampling from normal distributions and I think I mostly - well, ok, I think the first thing I did was just to look up a review just to sort of see in which range the data usually is. From the literature, I did not get a lot of information, but I also didn't search too much. But I started coding pretty much right away just to sort of get the numbers, do some for-loops that I would get - you know, that would get four columns with 25 rows just to get the structure and in parallel I was looking at - well, searching for data sets and then so on.

I: Ok. And how much time do you estimate you spend on preparing? And preparing here means before you started to write the code?

P: I think it was pretty much in parallel.

I: Ok. And did you read any literature on detecting data fabrication?

P: No.

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

P: No.

I: Ok. Did you consider other approaches than simulating the data?

P: I did. So, one idea was to actually take a real data set, but, yeah, I didn't do that, because I didn't think it would actually be helpful. I started by basically just drawing 25 times 4 samples from normal distributions. But then I well started using different kinds of distributions. So, other approaches - other approaches of course would also just basically get the data. Just test myself on that amount of trials, but that again would not really be fabrication. But, yeah, that's pretty much it.

I: Ok. So you didn't do like the real data approach so to speak because . . . ?

P: Because I thought it might be interesting for you to maybe have like a potential false positive in their, but then I thought, ok, you can probably find your own real data sets. So, no, I just, yeah . . .

I: Ok. So, you - did you decide to simulate data because you thought like taking a real data would not really be the task or did you think that like your chances of getting undetected were the highest with like simulating data?

P: Right. I thought my chances would be highest. But I was thinking more in terms of what would be more helpful to the project.

I: Ok. So you - just to clarify - you think your chance to go undetected would be higher with like simulating data or with taking a real data set?

3

P: I think my chances would be highest with - so, my chances to be undetected would be highest with real data.

I: Ok, alright, thank you. 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: I think if I had more time - so yeah, I basically started a few days ago, so I set some amount of time to do this. I think if I had more time, I would - so from my perspective now, I would probably look up - you know, how fake data is normally detected, what sort of the current methods are, yeah, so with more time I might be able to apply more methods. So, I have the feeling I did sort of the time that I could spend for this, but I do think that I could do more, anyway.

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, could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: Right. I don't know too much about this. I can only speculate on how you would detect fake data. So, one of course would be I am guessing to compare with real data what the range is, what normal standard deviations are, what real data normally looks like just in terms of absolute values, but that is a very easy step. I mean, if you would get a huge effect size that is nowhere found in the literature that would be weird. The distribution that you could fit to the data might not correspond to what you normally expect for this kind of data. So, for instance, reaction time data is not normally distributed usually. So, I would have expected more something like a inverse Gaussian. So, basically if I would [?] kind of reaction time data and it would be normally distributed that would be weird. Then, if I would be drawing from, well, sensible (?) distributions, standard deviations and means might be very uncorrelated. I would expect, for instance, that with higher reaction times you would get higher standard deviations. Yeah, that is pretty much it, I think.

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

P: I am not sure. I am not sure if I can point to specific things other than what I already described.

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

P: To some degree, yes. So, as far as I could at this point.

I: Ok. And how did you do that?

P: So should I just - I mean I could describe the whole process if that would be a good point in time to do, because that is what the process is taking those things into account.

I: Yeah, ok. So, we also have a block where we talk about the specific steps, but I would be . . .

P: Right.

I: happy to hear like if you - yeah, maybe, I mean you don't have to describe it in full detail now . . .

P: Right.

I: But maybe more . . .

P: Well, it is actually fairly simple, I would say, what I did. But I mean I looked at parameters of distributions in existing data sets. Then, I recreated these distributions and sampled from them. So, I had distributions for the means and for common variance. And then I was making two joint distributions. And essentially sampling from those.

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

P: No.

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

P: Sorry, could you repeat that?

I: Yes, sure. What criteria did you use to determine whether you thought your fabricated data would go undetected?

P. I just looked at basic histograms, but I mainly trusted that the distributions that I was sampling from made sense.

I: Ok. And did you have any specific and different criteria for the means and standard deviations?

P: It was based on what I found in the literature.

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

P: Yes. Maybe I could have looked at more different data sets. So, I am now thinking that I could have specifically produced outliers, which - but, so basically I assumed that if I would sample from distributions that are based on data sets that that would be realistic enough.

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 that you think could be interesting for us to know?

P: No.

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

I: Ok. 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: So, I looked at a data set that I found at the Open Science Framework where they were looking at various universities in the US and basically did this - the basic Stroop task and we are looking at fluctuations over the semester within students. I basically made a - so, I looked at all the means and the standard deviations they had, so they had about 3000 people, so, well, to go a step back, so first, I just used normal distributions and sampled from those for means and standard deviations, but then I saw that, well, this is not really how they would be distributed. So, I took this data set and I basically fitted distributions. So, there is some tools in Matlab to basically look at the data and find which distributions fit best. Then I took the distribution that fitted best, took those parameters, created random data from - with these parameters - from the distribution with these parameters, and then took 25 times 4 samples from that.

I: Ok.

P: Wait, so before I did that - so I had these two distributions for the means, so for congruent and incongruent, and one distribution for the variance, because I expected that to be similar and in this dataset I also found this. One description for the standard deviations. And then I created two joint distributions with, yeah for the mean and standard deviation, yeah and then took 25 samples from that joint distribution.

I: Ok. Yeah, so the next question would be: Could you indicate what steps you took to fabricate the standard deviations for the participants?

P: Yeah, so, that should explain that as well.

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

P: Yes. So, I run it a couple of times until the standard deviation of the variance within - no, between participants was sort of close to what I found in the literature. But because there is also fluctuations between studies, I didn't - I wasn't too worried about the numbers actually being too close, because I also thought that, well, in this experiment, there could be all kinds other factors that would influence the actual numbers. So, I was more concerned with how these numbers would make a distribution, how I could fit a distribution to those numbers. Does that answer your question?

I: Yeah . . .

P: Maybe you can repeat it.

I: So, the question was just whether you repeatedly fabricate data until you were satisfied with the results.

P: Right. sorry, yeah. So, I just basically run the script a few times and I made sure that they (?) chose one outcome that was - where there was a significant different and where the numbers were somewhat close - so that was, well, mainly just intuition-based. But I feel that I could have used any of those outcomes.

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

P: I mainly trusted the process.

I: Ok, so you had - so and then like when you did it again and again, you had sort of a gut feeling of what felt like could be correct or not or?

P: Yeah, it was a gut feeling. So, I was looking at sort of the means of the means and whether that made sense. Usually, so what I found was for responses, a response time of 507 (?) milliseconds approximately and it was sort of in this range. So, I mean by chance of course I could have gotten a result that was very extreme. So, I just wanted to make sure that that did not happen, but yeah it was a gut feeling mainly, yeah. So, I looked at the numbers in the literature that were reported, but that was about it.

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

P: Well, I plotted histograms, but with 25 data points, well, it looked ok. Yeah, so in retrospect I could have produced 1 or 2 extreme outliers in there, but I felt it was not really necessary.

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

P: Genuine?

I: Yeah.

P: Well, again, I just trusted the process.

I: Ok. Like did you take a look at the relationships between the means and standard deviations or the correlation within participants or so?

P: So, I looked at some scatterplots. So the relationships did not seem very strong, but I did not look too closely into that.

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

P: Once I was satisfied with code, I think it took me about 4 runs.

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

7

P: Matlab.

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

P: Yes. In the end, mostly the random function. I do now realize that I forgot one thing which I actually wanted to do, but I run out of time and was not really thinking about that anymore. Which is I wanted to get a more random seed, but, yeah, I did not. So, it was basically the default Matlab process.

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

P: Yes. So, it was based on this large data set.

I: Ok. But the cases that you have in your final spreadsheet, did they correspond to like original numbers or were you more inspired by real data?

P: More inspired. So they are all randomized.

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: I don't think so, no. [?] Let me just have a look at my notes in case (?) I forgot something. I can give you the URL of the data set that I used if that helps.

I: Ok.

P: So that is pretty short. So, I can just ... osf.io/ct89g. Somewhere there is a zip file in there that contains - their (?) thing is called final Stroop data. So, I can also send this to you if you want to.

I: I think like if we have it on the dictaphone, it should be fine.

P: Ok. I could describe the distributions a bit more. So, what I used was - what I found the best fit was something that was also new to me: generalized extreme values - this type of distribution. Yeah, and inversed-Gaussian also fitted fairly well for both the means and the standard deviation data, but the GEV seems to have - fitted even better, so I [?] used that.

I: Ok.

P: Yeah, that's pretty much it.

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

I: Ok, great, thank you. 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?

8

P: I was somewhat surprised of how long it took me to figure these - all the details out. So, I kind of knew that I would have to use a joint distribution, but in this context I have not made one before. So, that took some time.

I: Ok.

P: Ja, so I mean, of course, I could have just gone with the most simple way to do it - just draw, you know, from some random distributions, normal or otherwise. But I thought it would be best if I take the parameters for the distribution from the data set.

I: Ok. And yeah so like otherwise, did you think that it was difficult or?

P: I was surprised that I was unable to find any data sets by googling, but the Open Science Framework was very useful.

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

P: Well, if I had to do a binary guess, I would assume that you can detect it. I would assume that I am slightly less detectable than average, but I don't know what other people are doing, of course. But considering that I didn't extensively look at, yeah, how fake data is detected usually, I am sure that I overlooked a lot of things that would be obvious.

I: Ok. And like could you think of some ways how your fabricated data could be detected as fabricated? Do you have like some guesses about that?

P: No, no.

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

P: I think this is an extremely good initiative. Data fraud in science is really bad considering the importance of the whole [?]. The monetary aspect was very appealing as well. I did end up spending more time than I thought I would, but I think it is quite worth it. So, yeah, it is mostly wanting to support the study and also earning some [?] easy money. And also the exercise itself is a good practice, I think, to think about distributions and how data works, roughly speaking, yeah.

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

P: With my colleagues just briefly. I mean, we mentioned it, but there was not a lot of discussion, no. [REDACTED].

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

P: I think with the joint distribution - I am not even sure - it was sort of in my mind, but when I was briefly mentioning it, I think one of my colleagues was also mentioning that, and that made it more clear for me that I definitely have to use that. I am not sure if this is very clear, but I would say that mostly I did this without help.

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 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: If I think about it, I think - no, I don't think so, no.