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: Then we will start with the first block now. The goal of this block is to get some general information about you. And the first question is: Are you a PhD Student?

P: Yes.

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

P: Experimental psychopathology, yeah. So, within clinical psychology, yeah.

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

P: No.

I: Could you describe your knowledge or experience with the Stroop task a bit?

P: Basically, just from my studies. I did cognitive psychology as well. So, there I learned about the Stroop task.

I: Ok. And so you have learned about it in your studies. Have you also like read papers about it or so?

P: Yeah, yeah, yeah. So, I have read papers about it and also read papers about the effective Stroop task, which is used in Clinical Psychology. Yeah, so, I am familiar with the task.

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: R and 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: My peers being like people within this field of psychology as well or general?

I: Yeah. Peers means other researchers or scientists in your field.

P: Yeah, in my field, ok. Then, I think an 8.

I: Ok. And how confident are you that your fabricated data will go undetected as fabricated? On a scale from 1, extremely insecure, to 10, extremely confident.

P: I think a 4. I think, yeah, it was quite difficult.

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 spread the data fabrication, but in the end I just started over yesterday. I already tried it like a few times last week, but then I didn't come with a good solution yet. So . . .

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

P: I think, 3 days. Yeah.

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

P: 3 to 4 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: A 5.

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

P: Yeah, I thought about it before and how I could best do it, yeah.

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

P: Yeah, that was just like time when you are on the bycycle and thinking about it. So, I don't know . . . 15 minutes or something.

I: Ok, and did you read literature on detecting data fabrication?

P: A bit.

I: Ok.

P: Yeah, it was also a paper from your group where - I just scanned it but people who like just came up with random numbers that that was quite easily detected.

And people who use multivariate data that was more difficult. Yeah, something like that. But I just scanned the article, but yeah ...

I: And can you ... ok, so could you name the methods that you read about?

P: No.

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

P: No.

I: Ok. So you said that you thought about or like how you would do it. Can you describe your thought process - like for instance different approaches that you considered or so?

P: Yes. So, I want to simulate data. I didn't want to like just fill in random numbers. I want to simulate it but I wasn't sure how you could ensure that the data of the congruent - to me, it made sense that the data from the congruent condition was correlated with the incongruent data. So, I wasn't really sure how you could fabricate - simulate data with a certain correlation. So, I thought about how you could do that.

I: Ok. And did you also consider other approaches to fabricate the data or?

P: Yeah. I also thought about if you would have other data from the Stroop task, I thought you might infer like the covariation matrix from and then simulate the data based on that, but I wouldn't know how to do that. And I also wouldn't know - wouldn't have a data set of the Stroop task. But I thought you might be able to infer it from an actual, existing data set, yeah.

I: Ok. So, earlier you said that you read a bit about methods to detect data fabrication. Did this preparation influence your approach to fabricating the data?

P: Does what influence my approach - the?

I: Did your knowledge or your reading about this influence your approach to fabricating the data?

P: Yeah, eventually not. Because I tried - no, no, I think not, no.

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, not really, no.

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: Yeah. So, I thought if - because my - the means from the data from the congruent and the incongruent data first had a correlation of 0. Then I tried it differently and then they had a correlation of .98, which both seem to reflect very bad data fabrication. And also at a certain point, I had a Cohen's d of like . - aeh, 1.8 or something. So, I thought these were - that is all highly unlikely. And that would give away quite easily that the data was fabricated.

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

P: Yeah, what does the data - that is a good question. Yeah, I think if the data isn't as perfect as hypothetically you would assume it would be like. If, yeah, if there would be some outliers or anything.

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

P: Yeah, I tried to.

I: And how did you do that?

P: So, I tried to come up with a way that the data was correlated by first simulating data for the congruent condition and then basing - using that data to simulate the other data. And when I found a really big effect, I tried to - I just made the differences between the conditions smaller to decrease the effect.

I: Ok, and you mentioned outliers earlier. Did you also take that into account or?

P: No, just made me realize I haven't.

I: Ok. And did you take into account - into consideration relations in the data other than the Stroop effect itself? For instance, the distribution of the scores or other aspects that could be inspected with the data set.

P: Yes, so just within-subjects, the data is correlated.

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

P: Excuse me. Can you repeat the question?

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

P: The effect size, correlation. I just looked at the data to see whether it looked alright. Yeah, but my biggest conclusion was that it is really difficult to fabricate data.

I: Ok. And so you mentioned the effect size - like did you have a feeling about based on the past literature that you read or how did you like estimate . . . ?

P: So, yeah, I thought based on - that it is a quite robust effect so I would figure a medium to large effect size would be alright.

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

P: No, for both I just wrote a for-loop to simulate data.

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

P: Outliers. Yeah, I think - I think it is better to simulate the data from existing data to let it go undetected.

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: 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. In R, I made a matrix of - with 25 rows and 30 columns to simulate 30 observations per participant. And then wrote a for-loop which would fill up the matrix with the function rnorm. So, 750 data points and then took the mean for every row.

I: Ok and how did you like how did you chose the parameters for the function?

P: I tried to look into the literature but it wasn't really clear. So eventually, it was just taking a guess, I think, yeah.

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

P: Same, the same method, yeah. But this was only for the congruent data. And then for the incongruent data, I also did - I basically did the same but then used the means from the congruent data and added a random amount to it with rnorm again - if that makes sense.

I: Ok. And for the standard deviations, did you do these for the congruent and incongruent condition in the same way - like you first did it for the congruent and then for the incongruent or?

P: Yeah, yeah, I did it the same way, yeah. But I had some - yeah, that is what I did, yeah.

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

P: Yeah, yeah, because, so, first I got like a correlation of 0 and then a correlation of 1 and so I want a correlation of somewhere in between. So, now, I first added numbers to the means of the congruent data but then subtracted again and then eventually I got a normal correlation.

I: Ok. And how did you determine what was a normal correlation? Was it a gut feeling or?

P: Yeah, yeah, yeah, I tried to look it up but I couldn't find it.

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

P: Yeah, just at a certain point I thought it is enough. Yeah, I think if you really - yeah, and I realized by doing this process that it is really difficult to make actually good fabricated data. So, I thought if I really, really, really want to make this data look good, then it would take me probably a week full time to actually do it. So, then, I realized that - that, yeah, I would have to read a lot more on it and I wouldn't be able to do that.

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

P: Yeah, I did look at the normal distribution, I think, yeah.

I: Ok and what was your reasoning then?

P: Yeah, that probably the data - I could imagine this data looking like a normal - normally distributed. So, and obviously it was as I created my data using rnorm, yeah.

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

P: In what way? Or any way?

I: Yeah like whether you tried to see whether it looked real like real data, whether you had like some criteria to check that or so?

P: No, I - no, all I have done is just looking at the actual numbers, not really, no.

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

P: I would guess 20 or something.

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

P: Yeah, so an R script.

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

P: Excuse me.

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

P: Yeah, rnorm would [?], ja.

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

P: Yeah, unfortunately not.

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, not really.

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, the first question is: Did you consider fabricating these data a difficult task to complete?

P: Yeah.

I: And what was it that you thought made it so difficult?

P: Because there are so many aspects of data that can be detected and then, yeah, I participated because I really liked the idea of the study and because I thought it would be a challenge to fabricate the data, but I didn't expect it to be this much of a challenge. Because I am - I only thought probably of the simple things that you can easily detect and probably there will be so many other things you will look at and you will easily - might easily find out that the data is fabricated.

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

P: No . . .

I: And . . . yeah, sorry?

P: Yeah, no, I don't think so. I think you guys are smart enough to do that, yeah.

I: And so like can you think of ways how you think that it could be detected?

P: No. Yeah, no.

I: Or why do you think that your fabricated data will not be super difficult to detect as fabricated?

P: Because I think I don't know enough about data in this way to, yeah, to keep the data undetected. And I think as you are all methodologists, you will [?] beat me.

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

P: Yeah, because I was really interested in the topic. And because, at first, I didn't realize how difficult it would be. I thought it would be challenging but I didn't realize it would be so difficult, yeah.

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

P: I did. Not in detail, though.

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

P: No, 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: 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, I don't think so.