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: Ok. Now we will start with the first block. The goal of this block is to get some general information about you. So, as I said some questions may seem a bit obvious but I would like to ask them anyway. So, the first question is: Are you a PhD Student?

P: No, I am not a PhD student anymore.

I: Ok.

P: I do have a PhD. I am not sure what the gist of the question is.

I: Ok. Then the next question is how many years has it been since you got your PhD?

P: [REDACTED] . . . eh [REDACTED] years - is that correct - yes.

I: Yeah.

P: [REDACTED] years.

I: Ok, thank you. What is your field within psychology? With field, we mean for instance social or cognitive or . . . ?

P: Biological psychology.

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

P: Christ. . . Eh, what eh. That is a difficult question - I mean what does one consider an experiment. Because there are also a lot student projects that one could consider do be an experiment. I don't know. I think over a hundred would be a safe estimate.

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

1

P: Nowadays, my use of statistical programs is restricted to SPSS and occasionally R.

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: 7.

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

P: 8.

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

I: Ok, thank you. 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. I will just check whether it is still running. Yeah. So, the first question is: Did you fabricate the data in one day or spread the data fabrication over several days?

P: Several days.

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

P: I think four.

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

P: Maybe three hours. Nah, that's unfair. Maybe six. Six 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: 2.

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

P: I don't undertand the question. We thought about it how we could do this and then we executed the plan. Is this what you mean?

I: Yeah. So, for instance, did you read literature on detecting data fabrication?

P: No. I read one of our own papers on the Stroop effect.

I: Ok.

P: Because I needed - we will see later - the means and standard deviations to have a starting point.

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

P: No.

I: Ok, 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: No. But I did recruit one of the postdocs.

I: Ok.

P: To discuss how we could do this.

I: Yeah, ok.

P: So, I didn't do this on my own. I recruited additional help.

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

I: Ok, 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: When they clearly come from a normal distribution.

I: Ok. Any other characteristics?

P: When they are devoid of clear outliers, that would be suspicious. When they do not meet physiological or psychological plausibility criteria. For instance, the Stroop effect is described in the literature. If the response latency would be 6 seconds, that would be very silly because the response latency is known to be - I don't know from the top of my head - but 4-500 milliseconds for a congruent trial and [?] some 2-300 milliseconds longer. So, there is kind of a range of plausibility that you can extract from the literature. If the latencies would be completely off - like by a factor(?) of 5 or 10 - then something surely must be a miss. So, distribution, absolute values, and the lack of outliers - this would be, I think, hallmarks of fabrication. That's what I look for when I get students' work. So, that is probably why I come up with these.

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

P: The reverse, essentially. So when there are outliers. When the means and the standard deviations closely resemble to what has been found previously. And when the distribution is at least semi-normal but not perfectly normal.

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

P: Of course. We started out with great reluctance not wanting to fake data but once we accepted the mission we really wanted to make sure that you couldn't

3

detect our playing this game. So, yes, we tried to think about how we could make the data look real.

I: And how did you take these characteristics into account?

P: By fabricating the data taking into account these characteristics(?). So, we deliberately introduced outliers. We deliberately deviated from a normal distribution and our starting points always were realistic value for the Stroop effect. That is how we simulated the data I can already tell you that we did that.

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

P: Yeah, we took into account the fact that the absolute latency has a relationship with the standard deviation. If you are very slow then the standard deviation will be a little bit longer. So, there was a correlation between the mean and the standard deviation introduced.

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

P: None other than the ones we - so we simply used an algorith that should produce realistic data but we didn't check ourselves whether we could detect that they(?) were somehow fabricated. No, we didn't [?]

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

P: We could have done a better job by looking into realistic standard deviations within subjects but we had no time to do that. We thought it was enough time spent already, so . . . But it could have been even better, I think. We have done a good job but it is not optimal yet. If we really . . . but then we would have to go back to our own data sets and impute within-subjects standard deviations and we were just too lazy to do that [?].

I: And could you describe in more detail how you would have changed it then? Like in what direction the standard deviations within-subjects?

P: Then, we would have probably used the - we would have used the realistic distribution of within-subjects standard deviations, which for now we didn't do. I think, we set it to 30% of the latency. Something like - a distribution around 30 % of the latency and we simply generated the standard deviations in that way. So, it is not optimal.

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 that you think could be interesting for us to know?

P: I can' think of any(?).

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

I: Ok, thank you. 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 the first question is: Could you indicate what steps you took to fabricate the means for the participants?

P: Well, we simulated normal - we used R to simulate data. And we used the means from one of our own studies but we squared them to get skew and then we introduced the random fluctuating response latency - minimum response latency. And then we did some weird things just to deliberately delete things - deliberately make it not look like a standard normal distribution. We introduced some outliers at random - ourselves by hand. And then we did the weakest part. We modeled the within-person standard deviations as a non-central chi-squared distribution with a mean of the standard deviation of around 30% of the mean. This could have been done better, I think. And then we rounded everything to milliseconds and then we wrote it into a SPSS file to hide the fact that we did it in R. That is basically the algorithm we used. I have that in writing for you with the actual R code so I think that will be very helpful - much more helpful than actually my description of the process.

I: Ok, thank you. Could you nonetheless describe the weird things that you said you did in a bit more detail?

P: Honestly? No. Just weird things, just, you know. Taking a value and just choosing another value that somehow was in the general ballpark of the values and sometimes was two times or three times outside of the ballpark. That's a couple of times. I forgot how many times we did this. We simulated hundred subjects because we forgot that it had to be thirty. I think we did it maybe like 10 times or so. Randomly picking either congruent or incongruent. But, in all honesty, it was really done by hand and also not written down. So, this was haphazard. I couldn't reproduce it the next time. If you would ask me to do the same, that's not. The R code is easily run a second time. Although it would come up with obviously slightly different solutions because it uses a random generator based process. But what we did with the outliers I don't think we can reproduce that. So this was an one-time event, essentially. And we made no record of it. So, we couldn't - I never thought to do that.

I: So, you already mentioned some steps how you created the standard deviations. Are there any other steps that you took to fabricate the standard deviations for the participants?

P: No, this was it. This is the complete algorithm. And the only weird thing is that we did some stuff by hand that could not be reproduced. The other things would probably be easily reproduced using the same script again.

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

P: No, we didn't check the final results against the actual means and standard deviations that we sort of gave the simulator. Not. We could have done that. Oh, we did plot, I think. Yeah, but we never actually deliberately checked whether the standard deviations and the means were off from the values that we put in.

[Short break as someone came in]

I: So, you said that you plotted the data. How did you use the plot to inspect whether the fabricated data looked weird or genuine?

P: No, I don't think we - we just - I think, we just looked at whether we didn't completely miss - just the very very eye balling - rough eye balling whether everything looked sort of ok-ish. But no, no formal checks if the value is like 2% outside of the original means we do it again until . . . That, we didn't do. We didn't iterate the process. Just did it one time. We wanted to beat the system but not spend too much time. Basically, that is what happened.

I: Ok, so as I said sometimes it may be a bit repetitive but I will just to continue to ask you the question: How many different mean-sd combinations did you fabricate before getting to the final fabricated dataset?

P: Only one.

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

P: R. And then we transported it to SPSS and then from SPSS I had to save it as an Excel file and then copy - actually, I had to copy it to your Excel file because it had some sort of specific format.

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

P: We did. This is just a multivariate normal function from R which - as you know - uses a random generator [?].

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

P: No. That was actually my first idea to simply permutate a real existing dataset. But we only used the means and the latency coming from a realistic study. But not the individual data from subjects.

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: 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: No. It was more thinking about which approach would be most hard to detect. No, it is not difficult to fabricate data, unfortunately. But it is difficult probably to fabricate data that look like real data. That is not an easy task.

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

P: We hope so. That was the aim.

I: And why do you think that it will be difficult to detect?

P: Because we deliberately did the things that we think would give away a dataset as being fabricated. We introduced outliers, we deliberately did not go for an all too normal distribution, and we stayed close to realistic values, namely those that are based on a real paper of our group.

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

P: Well, I said, no, I don't want to participate. But then Chris told me, no, but it is really - of course, we will not disclose who is participating and it is really meant for improving science etc. So, my first gut instinct was: I am not going to fabricate data. That is really bad. But I understand why it is necessary that we actually increase our ability to detect fabricated datasets that will be in the benefit - in the long-term benefit of science. That is basically it. And furthermore, since - although you never know for sure - neither I nor anybody of my team has ever fabricated data I am much in favor of catching those who did. We have nothing to fear from a good data fabrication detection system. So, that is why I ultimately thought, of course, I should participate.

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

P: One postdoc.

I: Ok, and this postdoc also helped in fabricating the data?

P: Absolutely. He was the one who actually crafted the original R script to do the simulation. Although we did discuss the approach. He certainly was very helpful. He is also a very much better mathematician than I am. So, that helped, too.

I: So, could you describe your collaboration a bit more then? So, he was the one who set up the original script and then you . . . ?

P: Then I - yeah, ok - so, I asked him, if this is the mission, this is the Tilburg mission, and then his first response was, oh, we gonna beat them. So, that was fun. And then I said, I want to use this - this is the means and the standard deviations that I know are realistic, can we somehow simulate a dataset that end

up around these means and standard deviations and yet do not look like - so, we discussed what should we do and we came up with these things: shouldn't be a normal distribution, how can we deviate from a normal distribution. And then the suggestion of having kind of a mixture of a normal and a skewed distribution and just put them together. So, we exchanged ideas and he coded basically the final R script. And then he gave me the SPSS and then I did the weird things and transcribed it to the Excel [?] within your spreadsheet. But we didn't spend - I don't think we discussed it more than twice in total.

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: 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: Puh. Well, it would have been better to have individual-level data of one of my own Stroop experiments and I noticed that to actually get those data I would really need to spend a lot of time browsing through my old directories and this was a very painful discovery because it means that I don't have my own experiments from very long time ago one push from a button away but that's more of a detail. So, in terms of archiving your data so that they are completely immediately retrievable I think I have a couple of days work still ahead of me. We do a lot better these days but this is an old experiment. You can see how it changed over time. Archiving is a lot better now than it used to be like 10 years ago or so or something like that. I mean the data are there, fortunately, but it would take me a lot of hours to actually really get the final dataset that was used for the publication. And that would have been better because then the intra-individual standard deviation would have been modeled a little bit more reliable than we did now. Or at least we would have better information to do that. Now, we simply took a fixed value.