

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

P: I am a PhD student, yes.

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

P: Experimental, so cognitive psychology, yeah.

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

P: No, not a Stroop task.

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

P: I believe during my Bachelor in psychology I've learned about the Stroop task and I think I programmed one once for a course and that's it.

I: Ok. Did you read papers about it or so?

P: Yeah, I have read a couple of papers, the original by Stroop as well.

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: I mainly use R and Python, yeah.

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 about an 8.

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 am quite confident. I would say an 8 as well, yeah.

## **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. Did you fabricate the data in one day or spread the data fabrication over several days?

P: I did it in one day.

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

P: I think it took me about 10 hours, everything together.

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 about a 4, yeah, so about half way, yeah.

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

P: Yeah, I discussed it with colleagues like how they would approach the problem and I have talked to a lot of different people. So, I tried to incorporate different aspects into it, yeah.

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

P: On preparing, I think, about 5 hours, yeah.

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

P: No, I haven't.

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

P: No, I actually haven't, no.

I: Ok. So you said that you had a couple of discussions about how to approach the task. Which different approaches did you consider? Could you describe your reasoning process during the preparation time a bit?

P: Yeah, so my first instinct was to generate the data using some kind of a model like a model participant. That's what I went for in the end. And how - the other things I considered was doing it by hand - and just filling in values by hand - or using data I have lying around and change it around a bit. So, using real data, but then just, yeah, re-using it basically. Those were kind of the three approaches I have considered.

I: You mean like using data from your own studies or ....?

P: Yeah with reaction times and just changing them a little bit.

I: Ok. But not data from like Stroop tasks or so?

P: No.

I: And did this thinking influence your approach to fabricating the data?

P: Yeah, it did. But what it mainly did was it influenced my approach into generating the data, the thing I did in the end. Like you should make sure you watch out for these kind of things. Yeah, there are a couple of steps that I didn't think of myself my colleagues thought about, so I could incorporate.

I: Ok. And what was your reason to go for the approach you chose and not for the two other approaches?

P: Well, the reason is that I thought, this would be the most difficult to detect, because I basically - I generated data that should be very similar to how we make statistical models of data, so I think that is very difficult to distinguish from real data. And I also think this would be the least susceptible to, yeah, me making manual errors, repeating numbers or something like that, some kind of repeating pattern that I would do as a person, right. Yeah, so that is why I chose for this. I think this would be the hardest 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: Timeline, I - yeah, no, I think, no not timeline-wise. Sorry, I just had to think for a little bit, but no timeline wise I think we are good. I did spend quite a bit of time looking at the data, so plotting the data. Instead of 25 participants I just plotted - I generated 25 000 participants and plotted how the data looked and if it looked convincing to me. So, there was a bit of post-generation checking if everything was working ok. About 2 hours for that.

### **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: What would to me make data look more fabricated is especially if you have 25 participants if all participants show the same direction of the effect. So if the within-subject effects are very similar to each other that would to me signal that someone has been fabricating data. Or repeating numbers, like repeating patterns of numbers, because it is so unlikely that you will find 498.8 multiple times something like that, right? So, yeah, I think that would indicate to me that data is not real. Or the person (?).

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

P: Well, I think that is also a little bit with the design of this that you just give means and standard deviations per participant. I think it is quite hard to tell if the data is genuine. I would be more convinced if you had trial-by-trial - because we were supposed to simulate 30 trials - if you had the trial-by-trial data I think it would be easier to see - for me at least- if the data looks genuine. So, raw data would be the answer to that for me.

I: Ok. And like how could you - or like how do you think could you tell based on raw data whether it is real data or not?

P: Based on what - yeah, so, I think also on previous findings if it is something I know about, right. So if I - I personally look into inhibition (?) of return if I would find an inhibition of return effect that is larger than 30 milliseconds or something like that I would say this not - you did something - I would assume that someone did something wrong rather than fabricated their data, but, you know, that would indicate to me that something is up - if the magnitude of the effect isn't ok. And the raw data, yeah, if, I don't know, the amount of participants doesn't match up and that kind of stuff like, yeah, I think there is - or if the labels on the - when the data was being recorded doesn't match with the period of time that they recorded the data in, that kind of stuff would signal to me that there is something fishy going on.

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

P: Well, mainly that - well, what I took into account was that I shouldn't do stuff manually, because if I do it manually, you might be able to detect in pattern how I am entering stuff. So, I thought I should do this with a computer and not touch it myself at all. And make a very convincing model participant, but don't type in the Excel sheet myself. That was my main consideration.

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

P: Yeah, the relations that I took into effect were the relation between mean and standard deviation - that typically if you find a larger mean you also find a larger standard deviation. And the relation that - so I sampled the data from a normal distribution but that the within-subjects effects aren't sampled separately so that you don't sample from a normal distribution for the congruent condition and a normal distribution for the incongruent condition, but you sample from a normal distribution for the congruent condition and then the difference with the congruent condition that you sample from that, because otherwise there is no relation between the one condition and the other, right. So then it will - if you would draw it in a point cloud the point cloud would just be circular and it is supposed to be a little bit - there is supposed to be a little bit of correlation between the two conditions. So, those are the two things I took into

consideration.

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

P: If - yeah, so I found some papers on Stroop task effects and I generated my data according to one paper and then I looked at other papers if it matched up. And it matched up quite well, so I thought this is a good thing. Also, I plotted the data as I said earlier to kind of get a sense of did it look too uniform or did it look like there was variation in there, so yeah.

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

P: Sorry, could you specify what you mean by that?

I: Yeah, so like the last question was about criteria that you used to like see whether your fabricated data would go undetected and then this question is more specifically did you have different criteria for means and standard deviations?

P: Well, yeah, I got all my - yeah, I got my criteria basically from papers. So, if the Stroop task paper matched my generated data, then I would find that pretty convincing. Yeah, that basically.

I: And how did you - or like what were criteria to say that it matched the other papers?

P: Ok. Yeah, criteria would be that the - so that the mean and standard deviations that I generated were pretty similar to what they found. Some papers actually include participant-data, so you can look at the distribution a little bit and compare. So, I did that. And also to kind of get an idea of the consistency of the effect. The effect is quite consistent as well, which helps in fabricating the data. So, that means you don't need to add too much noise to your sample to make it look more convincing. So, I looked at that like how consistent is this effect and, yeah, what are the means and standard deviations typically found.

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

P: Well, in hindsight, if I would have done it again, I think, I would have - what I did now is I generated means and standard deviations, but I think I would have made a model of a participant like what might be going on within the participant. But that would be a lot of more complex, that would take a lot more steps. Like including like what happens if the participant - the fake participant would guess or doesn't see it has a lapse or something like that, so model that in there as well. I think I could have made a more complex participant-model, but I think this [?], but in hindsight, I would have liked it to be a little more complex than it is.

I: Ok. And like do you have - you said you would like to make it more complex. Do you have specific things where you would like to say that like ok, with regard

to this, I would like to have a more detailed function or so or?

P: Yeah, so mainly with regard to the participant itself. Also having - sort of having guesses or well guesses, lapses in this case then. So that participants weren't paying attention and pressed to late or that participants press really quickly without really paying attention to this stimulus, where they just sometimes reflexively respond participant or what happens as well that if you present the same color multiple times, then participants will get quicker, as there is this effect of the sequence as well. So, I think for an accurate - or for a very convincing model of a participant, you need to include all those different parameters in there, yeah.

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: Yeah, one thing that I didn't mention is that - or are you gonna still ask about how specifically I generated the data.

I: Yes, yes.

P: Ok, then I am good.

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

I: Ok, because now we start with the fourth block. And the goal of this block is to get some information about the specific steps of the data fabrication process.

P: Ah, perfect.

I: So, could you indicate what steps you took to fabricate the means for the participants?

P: Yeah. I could actually look it up as well, but what I did - I think I have it open somewhere - if I can find it, sorry, let me just ...

I: No problem.

P: [REDACTED]. Yeah, so what I did is that I specified the mean for congruent condition, which I found from one paper, and then I looked at a couple of papers what typically the difference in means is for a congruent and an incongruent condition. And I specified that. And I think the mean here that I used is 672 and I found that the difference between the two conditions is about 50 - is it milliseconds - yeah, milliseconds on average. And that, yeah basically, that.

I: Ok and so this was your starting point and then what type of function did you use to ...?

P: Yeah, so I used a two-dimensional Gaussian to sample the mean and standard deviation from. So, basically, there is a certain amount of correlation between the mean and standard deviation. So if you would draw it in a plot, you would

get kind of an ellipsoid function. So, I kind of played with that until it looked about right. And then, I did the same for the difference in mean and difference in standard deviation between the congruent and incongruent conditions to make sure that there was a weak correlation between everything. And yeah, I sampled that - I sampled from that distribution for every participant basically and added some random noise to it.

I: Ok.

P: Yeah.

I: And, yeah, so the next question is: Could you indicate what steps you took to fabricate the standard deviations for the participants?

P: Yeah, yeah, ok, I kind of just - I just did that, but the standard deviations I didn't specify. I used a standard deviation of about 130, I think, and then I said in the incongruent condition, that's about 10 higher. So, that's the values that it should be centered around, yeah.

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

P: Yeah, so, I fabricated the data in a very - with high participant numbers, just so that you can plot all of the distributions and to see if there is enough variance there. And if everything is going ok. And I think it did that about 20 times, before I was satisfied with my model and reduced it down to 25 and put it in an Excel file.

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

P: Yeah, so I - well, for one, if there were no like things that I - some kind of artefacts. For example, the random noise putting that in was that was a bit tricky. What it does now is that it determines for each participant a magnitude of noise between 0 and 1, where 1 is just complete random noise and 0 is perfectly sampled from a distribution. So, that was a bit tricky to get that right because if that is correlated with each other, then you don't get these ellipsoid clouds of or ellipsoid relation between the mean and standard deviation, but you get these kind of crosses, you get kind of these axes, these weird shapes that I think would be very detectable if you looked at it. So, I made sure there was nothing weird in that. And then once I got the distribution right I looked at papers to see if it matched the distributions I found in papers. And once it did, I was satisfied.

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

P: Yes, so by plotting mainly, yeah.

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

P: Yeah as well - what I mainly looked at if - because typically if you test about 25 participants, there should be 2 participants or so that are a bit weird or not quite showing the effect or not even not showing the effect but are just doing

something that doesn't quite make sense. At least that is from my experience. So I made sure that there is enough noise in there, that there is at least 2 participants that are kind of like [?], so yeah.

I: And – ok so when you like looked at the plots and so on, like how did you determine whether this would be good or bad? Was it that you had like really specific objective criteria or did you follow more like a gut feeling or so?

P: Actually, in this case, I just - I followed kind of a gut feeling from like spending quite a bit of time of looking at data. I didn't quite think about – I am sure I could have objective it somehow, maybe tests of normality or something like that, but it is a little bit not completely normal – normally distributed. But I didn't do that. It looked good to me, so then I figured, ok, I think this is going to give you tough enough a job as it is.

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

P: So, well, yeah, so doing it with 25 participants, I think I did it about 4 times before I was satisfied.

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

P: Yeah, I used Python to generate the data.

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

P: Yeah, so this would be sampled from a random number generator, so yeah.

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

P: No, there is no real data.

I: So you were like kind of inspired by real data, but there was no case where you just like copy-pasted ...?

P: No, no, that is correct. I was, yeah, inspired by the real data, tried to do what the statistics are doing on the real data and see if it looked convincing.

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: Let me look if I mentioned everything. Yeah, I mentioned about everything, so.

### **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: No, I didn't find it very difficult. I did find it – what I found most difficult was just thinking about how are you going to find out that this is fake. I wasn't really thinking about – yeah, in a sense I wasn't trying to make convincing data, I was trying to make data that you can't distinguish from real data. That was kind of my goal. So, I was really thinking like what – how are you going to look at this? How are you going to try to detect that this isn't real data?

I: And do you have like ideas that came to your ideas how [?] ...?

P: Yeah, so mostly statistically what we were thinking about. So that you would look at if there was a relation between mean and standard deviations, for example, because if that is not there, that is weird, because there should be some kind of correlation between that. And also at the within-subjects effects as I mentioned previously that they should make sense so if someone scored quite - if someone had a quite slow reaction time, that would have an effect on their - quite slow reaction time on the congruent condition, that would have an effect on their incongruent condition as well if the size of the within-subject effect - so, I was thinking in those terms like what can you do, because means and standard deviations, it is not that much information to go on for you guys, I think. So, I was just trying to think like ok what – statistically, what can you try to get out of this? Yeah, that was my reasoning.

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

P: I think so, personally, yeah.

I: And why do you think so like what ...?

P: Well, my reasoning was that if this is - basically, the statistics we use do certain assumptions about the data and if I use those assumptions to generate data and add noise to it, I think it would be statistically very difficult to distinguish it from real data, because this is - these statistics make the model of the data, [?] it is normally distributed and compare to those two normal distributions to each other. If I basically reverse-engineer that and add some noise to it, that should be very difficult to detect in my opinion, at least with my line of reasoning.

I: And can you think of ways how it might be detected as fabricated?

P: Well, I was thinking - well, the fact that you mentioned a random number generator worried me slightly, that you might find some kind of pattern in that, that you could find the seed of it somehow. I have no idea but, yeah, so that would be – that's worrying me now a little bit. I didn't consider that. And what else could you do? I am thinking that maybe there is some kind of information in the Excel sheet if that makes any sense - if it has been copy-pasted into it or something like that all in one go, that would be a bit suspicious. I don't know if you could retrieve that somehow. But other than that, no.

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

P: I thought it was quite fun, actually. I just - [REDACTED], so I decided to sign up, because it sounded like a good exercise for myself as well to see like whether I understand statistics enough to reverse-engineer this, so it was an exercise for myself. Yeah, and it just seemed like fun.

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

P: Yeah, so with some colleagues [REDACTED].

I: Ok, and so would you say that these people helped you in fabricating the data?

P: Well, conceptually yes, but I have programmed it all myself.

I: Ok. And so you - so like did the other people with whom you discussed also participate in this study? Do you know that?

P: I think, another colleague of mine [REDACTED] is also participating in this study. So, I think another colleague of mine, but I didn't really discuss it with him. So, that's the one colleague I didn't discuss it with, coincidentally.

I: Ok. And so you discussed like different approaches how one could do it or?

P: Yeah, exactly. Different approaches. And then I would be like this would be my approach, would that be convincing to you? And then it kind of went back and forth until we were like ok. But yeah, in one sense, we all kind of approached the problem from the same angle which is statistical. We would be thinking how is this statistically indistinguishable, so yeah.

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, not really, no.

I: Ok, then this is the end of the interview or is there anything else you can recall about the data fabrication process that you think is worth mentioning?

P: Anything I can recall that is worth mentioning? No, I don't think so, no.