**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, 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: Yes.

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

P: I am kind of on the verge of cognitive and developmental psychology.

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

P: Not in my PhD, no. But . . .

I: Or beforehand or so?

P: Yes, before my PhD, yes. I have had the experience with the Stroop task in several different types of people.

I: Ok. Could you describe your experience or knowledge about the Stroop task in a bit more detail?

P: Yeah, while I was a student, I did several experiments, but those were not published or anything, so, yeah, I would - let's see, yeah, I mean, it was just as a student, so it wasn't really like a . . .

I: Ok. But so you carried out the experiment and you also analyzed the data of the experiment?

P: No, I never analyzed . . . I mean, it was a long time ago. So, it was never in like a 30 participants or something, more like 5 or 7 or something. So, no, I mean I took - I gave the experiment, like I administered it, but I never researched any data on it, no.

I: Ok. But so you would . . .

P: I am . . .

I: Yeah, sorry.

P: . . . familiar with the Stroop task enough that I know what it consists of and . . .

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: 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: 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: That - so, what is the question? How confident am I that it will be undetected?

I: Yes.

P: That it is fabricated. Well, that really depends on what your - on how the program - what the program looks at. I am - I have no idea how - what that program looks at. So on a scale of 1 to 10?

I: Yes.

P: I guess a 7. I really did my best, but I have no idea what your program looks at.


**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: Several days.

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

P: I think 3 or 4 different days.

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

P: Spread out of over these days, I think, yeah, a couple of hours. Maybe 4 or 5 hours or something.

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 a 6. I really wanted to make the data as real as possible.

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

P: Yeah. I looked up again the Stroop task and the kind of the theory behind it and the types of different Stroop tasks that are possible and yeah so I took a while to kind of get back to the theory.

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

P: Yeah, 2 hours.

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

P: No. Oh yeah, I could have done that. No, I didn't.

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

P: How they had been detected - no, I didn't - for this, I didn't look into any papers on that, no. I mean I have heard, of course, about cases about data fabrication, especially in psychology. But not on the Stroop task. And not - I didn't look for any information on that.

I: Ok. And do you think that your knowledge about previous cases of data fabrication that you already had influenced your approach to fabricating the data?

P: Yeah, a little bit, because I - or what I know about data fabrication and, you know, being able to find out that something is deliberately fabricated is that the data is too perfect. And you saw that in the case of the UvA social psychologist, I cannot remember his name, was a German guy. That his data was so statistically significant that it was just not even possible. So, I made sure that it was not too perfect. That it was significant, because you wanted the hypothesis to be able to be positive, but not so positive that it was unbelievable.

I: Ok and could you describe a bit more like what you did to make yourself familiar with the Stroop task and how this preparation influenced your approach to fabricating the data?

P: So, how did I familiarize myself with the Stroop task? I looked at a number of articles on the Stroop task and I tried to see what reaction times they had in their data - the means and also the variation in that, the standard deviations and such. And to see what - the outliers that they reported on that. The differences between the - what do you call the difference . . . ?

I: The two different conditions?

P: The different conditions, yeah. So, the . . .

I: Congruent and . . . .

P: Congruent and incongruent. So what the difference was between those. And then I went to this website that provides you with making your own tasks. So, there is this demo for the Stroop task and I took that a couple of times in a number of different ways. And then I looked at the data that came out of that. And then I used that as a kind of, yeah my mode (?) that what I would base the reaction times on.

I: Ok.

P: So, does that answer your whole question?

I: Yes, yes.

P: Ok.

I: Ok. Then, I think, 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: The timing – such as what?

I: Like what you did wrong and like how the order of the different steps was?

P: I made a couple of notes. Yeah, what did you just say?

I: If you don't have – if you can't think of further comments for the timeline that is also fine.

P: That is fine as well, yeah.

I: We will also go through the different steps.

P: Right. I will look at my notes. Just to make sure.

I: Yeah sure.

P: If there is anything else that I am forgetting. Yeah, so I did that Stroop, that demo of the Stroop task which was accidentally 30 trials as well and I looked at what happens to the reaction times if you do it really fast or like really at my very best and what happens if I just do it at my own pace and then I realized, ok, some people are still going to be much faster than me and probably some will be slower than me and so that is kind of the range that you are going to be looking at for the reaction times.

**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. Could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: Characteristics that would make it more fabricated? Just in general on data?

4

I: Yes.

P: Yeah. If - I assumed for making this data that it would look more like fabricated data if it would be just too fast, like too fast to be able to be done by humans hand. Or if the difference between congruent and incongruent condition would be too large. Of course, it does happen. There is always outliers, but if you would make 30 participants all with really large differences between those two conditions, then it would be unbelievable (?). Yeah, they have to lie beside or by each other kind of in a believable fashion. And also I realized that, of course, there are going to be some individuals that have the opposite as well. So, it always happens where some individual will have a faster incongruent time than congruent time. So, I also put those into the data.

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

P: Yes, so more I guess deviations of what you actually expect. I think that if you just make fabricated data – yeah, like I said – too perfect, you would - yeah, all the data would point in the direction that you want to - that your hypothesis is in. So, sorry what was the question?

I: Could you name specific . . .

P: specific characteristics of

I: . . . that would make data look genuine or more genuine?

P: Yeah, so I think if you put some data in that deviates from the hypothesis, like for example yeah the congruent-incongruent that is different from your expectation. Or, yeah, some times that are a little bit faster or a little bit slower than you generally expect. What else? Yeah.

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

P: Ja.

I: And how did you do this?

P: Well, first of all, yeah, like I said, I kind of made this range of look (?) this is what you would expect, then I put in some cases that would deviate from that, and in the very end when I made the data and it looked like it was just really, really over the top significant, then I kind of turned (?) it down a bit so that it would not be so perfectly significant. And yeah when I had doubled data, I just kind of tweaked it a little bit so that – yeah, to make it more deviant from the actual hypothesis, so to speak, yeah.

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

P: Relations . . . what do you mean?

I: For instance, the distribution of the scores or other aspects that could be inspected with the data set.

P: Yeah I - what I wanted to do still is kind of make a distribution. I didn't do that in the end, but that might have been a good idea. No, I didn't do – I didn't make like a distribution plot (?) or anything. That is what you are asking, right?

I: Yeah or like other aspects that could be inspected with the data set. For instance like the correlation between - within person or something like that.

P: Correlations within the person. I am not sure what you mean by that.

I: Yeah no..

P: No, I don't think I did that.

I: That is fine, ok. And what criteria did you use . . . yeah, sorry?

P: What I did do is – maybe that would help for that question – is that if a person - if I made for the congruent condition, if I made the standard deviation a certain percentage so to speak of the actual reaction time, then I did approximately the same for the incongruent condition. So, I took into consideration kind of the – some people will have a large standard deviation, so I did that for the congruent and the incongruent trial - the conditions.

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

P: What criteria did I use? Yeah, I guess what I said before. So whether it would deviate from kind of the expectation. And also if the reaction times are believable. Also if the standard deviations are believable. Then I also realized that maybe it looks at whether – ok, I noticed when I was fabricating the numbers, yeah the numbers, that I had a tendency to write certain numbers and not others. For example, I often ended the number with a 9 or a 1, but less with a 0 or a 4. I thought that maybe the program looks at that. So, I kind of took that away so that it was more overall the same, so ending in the same numbers and using the same number for the whole range. And I also had a tendency to write numbers going either all up or all down. For example, 236 or 641, because somehow my fingers just went that way or that way. I have no idea if that program looks at that but it might that it can somewhat pick up like human tendencies, like deliberate tendencies. So, I tried to kind of tweak that as well so that it would look more scattered and more random.

I: Ok, and so you controlled that for all of the digits or only for the last one?

P: Yeah, the ones what I mentioned – the ending, the number ending, I controlled only for the last digit.

I: Ok.

P: But the other thing that I said about the numbers either going up or going down – that was a general thing. So, overall for the whole number. Yeah, so, I tried as much as possible to make that random – random digit.

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

P: Yeah. For the means, I just - I based it mostly around the first kind of - the test that I did on myself, kind of what, you know, what do I – what would I – if I was the participant, what would I normally come up with as data. And then I based all the means kind of around that with the thought in mind what I said before. So, kind of ranging around that. And then the standard deviation, yeah kind of the same thing so that it was believable but also with some more outlying data.

I: Ok. And did you have specific borders or did you do it more on a gut feeling?

P: I did it mostly on a gut feeling. Yeah, I didn't have a border of I shouldn't cross this or I shouldn't cross that. What I did keep in mind was that – yeah while doing the task itself – ok, it is really hard to go to this and or to that. And I mean, a reaction time of I don't know 500 milliseconds is just almost impossible and you have to be really, really, really slow to get to a reaction time of, I don't know, 1300 or something. So, whatever I came up with, that would be kind of the range. But I don't think I had a conscious stop in mind, where I shouldn't go over.

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

P: I could have done more in researching the standard deviations. That was - that is the only thing that I feel more insecure about. So, yeah, I could have paid more attention to researching the data behind that, because I found it hard to come up with that. We will see, yeah.

I: And - 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 I [?].

**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: I think, I already answered that question.

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

P: I also answered that.

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

P: Yes, yes. I fabricated multiple numbers, multiple digits and in the end, when I had 30 and it hit a certain p-value of – I think I just said it – it was lower than .001, then I thought, ok this is too far and then I kind of tweaked it back.

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

P: Well, partly by the things that I said before like tweaking the digits because I thought it was too – it was not random enough and partly by checking the p-value and the t-value just so that it wouldn't be too high - or low.

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

P: Yeah, yeah, I mean, weird . . . I just especially paid attention to whether it was, yeah, believable, I guess. That it looked like real data. So weird, yeah, in a way, yeah, weird would be a good description of that, because you don't want it to look like it looks like you know a human typed that in. So, in that case, yeah, weird would be something that I stayed away from.

I: And you had a gut feeling sort of what would be like believable and what wouldn't?

P: Yeah, I guess that was just my gut feeling.

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

P: Yes. Yes, I did.

I: Was this different from the other things that you have mentioned so far?

P: No this is – I have mentioned this before. I don't think I can elaborate more than I did already on that.

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

P: Yeah, I think – say this is the 30 combinations that I ended up with, then I made maybe 60 so that I would have discarded half of what I made.

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

P: No.

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

P: No, I thought about that. I did think about that, but I didn't do that, no.

I: Ok, let me just shortly come back to the earlier question about how many different mean-sd combinations you fabricated before getting to the final fabricated dataset: You said that you changed it around 60 times, but like did you mean like that you changed individual numbers or did you change the entire combinations of means and standard deviations?

P: No, what I meant was that: If I make an approximation of the amount of combinations that I made that I made approximately 60. And I would have discarded half of those. But it is not like I generated a list of 60 and then I went, ok I am taking these away and these I am keeping. It it more that I made a couple and then I thought those are good, then I made a couple more, then I discarded some and then – yeah, so it was more kind of random, not really stepwise.

I: Ok. And you adapted them within the process and not . . . ?

P: Yes, exactly. Yes, that is what I did.

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

P: I used - first I used real data, yeah, of the data that I did on myself, but the data that are in here they are all fabricated. So, there is no real data, because you asked for fabricated data. But it was based on – yeah, the first step was based on real data, yeah.

I: Ok and so did you base it on both your own trials and the numbers that you found in the paper or primarily on your own or?

P: Oh yeah. Yeah, that is true. I also based it on the numbers that I found in the papers, yeah. Don't remember what papers they are, but I could look them up if that is necessary.

I: No, it is fine.

P: That is fine, ok.

I: And then like you put those numbers together with your own and . . . .

P: Yes.

I: Ok.

P: Yeah exactly and then kind of compared them to each other, but of course I also realized that some Stroop tasks are different than others, right? Because on one you might need to – for example the one that I used, you had to press a key for the different colors, but there are also Stroop tasks where to press a key for . . . yeah, I don't know - there is different methods of asking someone to respond. So, of course, the reaction time is highly dependent on that and also the language in which it is done (?). I mean the one that I did was in English, while, of course, if you do it in your own language, then it is going to be easier as well. So yeah - so I took that into consideration because if you create reaction

times based on different tasks then, of course, it is going to be different. So, I kept that in mind.

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, I think I mentioned everything . . . 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. Did you consider fabricating these data a difficult task to complete?

P: Not difficult, no, I wouldn't say it was difficult. Maybe if I – no, I wouldn't say it was difficult.

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

P: Sorry that my approach was?

I: Will be difficult to detect as fabricated?

P: I am sure you have a program that knows exactly how to see that, but I did all I could to make it look like it was non-fabricated. So, I hope it looks genuine enough.

I: Ok. And do you have a guess about how it might be detected as fabricated? Could you think of criteria?

P: Yes, as I said before, it probably looks at the – if the data is too much in the direction of positivity – so yeah, if it is too much like the hypothesis. And, like I said, also, it might look at whether the data is typed by human hands rather than random – well random – rather than what you would expect if someone actually got this data from actual reaction times. So, those were kind of the assumptions that I had of the program that you use.

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

P: Because I think it is really important that there is research in this. I think it is a really interesting study, I have never heard of any study like this, so I was intrigued. I think, in the last few years, there have been some cases of data fabrication [REDACTED] and I think it is important that these tendencies are stopped for different reasons. So, yeah, if I can help in that respect, then yeah,

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

P: No.

I: Ok, so you had no help in fabricating the data or so?

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: Let me see if I wrote anything down in my notes. I put here that I reasoned that there are a couple of possibilities where (?) reaction times can be close together or more far apart. For, example some people that will be really fast, they might also not have less - the bias of the incongruent might be less for them as well. So apparently – I don't remember this exactly, but for certain data points I took that reasoning in consideration. So, I put the reaction times of congruent and incongruent more close together if the reaction times were fast if you understand what I mean. Yes? Ok.

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

P: No I think everything that I have in my notes and that I have in my head has been mentioned.