

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

Note

This interview was conducted while another person was in the room.

Block 1: General Information

I: 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: Yeah, PhD Student.

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

P: I would say Methods.

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

P: No.

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

P: Well, I only fabricated some data for the Stroop task and I heard of it before.

I: So, you heard about it in your lectures during your studies?

P: Yeah, my undergraduate, I would say, yeah.

I: And did you read any papers about it so far or so?

P: No.

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: Only 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: Peers? So like psychology peers?

I: Other researchers or scientists in your field.

P: Wow, that is very difficult. If it is psychology, I would say 10. If it is like statistics, I would say maybe 3 or so.

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: 10.

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, the first question is: Did you fabricate the data in one day or spread the data fabrication over several days?

P: Yeah, three days.

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

P: In total, maybe 2 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. And did you prepare in any way before starting to fabricate the data?

P: Well, first I thought about how I gonna do it. And then I did it.

I: Ok. And how much time do you estimate you spent on like thinking about how you would do it?

P: Yeah, maybe one third thinking and two thirds fabricating it.

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

P: No.

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

P: No.

I: Ok. And like could you describe in a bit more detail like how your thought process looked like?

P: Yeah, ok. So, I am assuming your methods ... So, the goal was basically to be undetected by your methods. Now, I am assuming that your methods are good. Which means that you can't detect like properly generated data, right? Because otherwise your methods are shit basically, right? So that is like where I am starting. And then, it is very easy because the Stroop effect is something that is actually there. It is one of the most replicated things out there. So, my strategy was to just do the Stroop task myself. And then just generate data like that. And the only requirement is that we get a significant effect. So, I just played the Stroop task for 25 times - at different times - and I just take that. And that's it. So, that was the whole reasoning. And then - I was thinking about this - how to generate this ideally. But then the best is just to actually create the data how it is supposed to be created because then if your methods are working you won't detect it as artificially because it is not. So, that's it.

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.

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

I: 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: Yeah. So, I mean you have these things with like how often integers are - different integers are appearing, I think. And if things are fitting too well some well known distribution. For example, if I would generate it myself, I would use something which looks like a response time. And then if it fits the data - if it fits some model and it fits too well, then probably something is a little bit fishy. But who knows? And, so, if I would do it in a way that I would generate it from some distributions, I would like add all sorts of random stuff to kind of avoid this way of detecting it.

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

P: I think that is the same, I think.

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

P: No, that is not necessary because I did not generate the data in an artificial way.

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

P: No.

I: Ok and -

P: Ok maybe. I wanna add one thing: So I was thinking of intra-individual differences: And I did it like at different times to make sure - when I like tired and not tired and stuff like that - to make sure that there is some variation. And there is lots of variation. So, it worked well, I would say.

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

P: Well, that was like the initial reasoning.

I: Ok. So you did not have specific criteria for the means and standard deviations?

P: No because like however - whatever the dimensions are that - how [?] determine whether it is a true dataset or fabricated. Because I generated them in like a genuine way all of these dimensions have to kind of point in the direction that [?] that it is not fabricated, right? Because it is not fabricated. So, basically, I kind of sidepassed (?) all these problems by just generating the data how it is supposed to be generated.

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

P: No.

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: Yeah. So, I went online to find a Stroop task which I could do in my browser. And I found one. And then I just played it 25 times for the 25 participants. Then, it was actually 60 trials required but then in this game it was only 40. So what I did is for each participant I calculated - or for each round of me, basically - I calculated the ratio of errors and then I scaled up from 40 to 60 trials. And then again - and the missing ones I sampled from the whole pool of all trials of all participants. And that is how I scaled it up from 40 to 60. Because that

was the only thing where I was thinking maybe you can find something because maybe of some variance or something like that that if less variance within person if you have 60 trials compared to 40 trials. So, I scaled it up. And that is the only thing what I kind of changed in the data after generating them in a genuine way. Well, and then I just calculated for each participant the mean in the two conditions and the standard deviation. And that's it.

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

P: Yeah. So, that is - I basically calculated from the data. So that is answered.

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

P: No, I just did one go because I knew that I was generating something which is not fabricated. So, there was no need for that.

I: So, did you determine in some way whether you were satisfied with the fabricated data or that they needed to be adjusted?

P: No, because I knew that the process is a genuine one. So, I didn't look at anything.

I: Ok, so you had only one mean-sd combinations before getting to the final fabricated dataset?

P: Exactly, yeah.

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

P: Yeah. So, I used a script online to get the experiment. And then I paste them into Excel, I [?] to R [?] of the pre-processing. And calculated the means. Exported it again to Excel. And that's it.

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

P: No, because I generated properly by doing the task.

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

P: Well, in the sense that I created them in the experiment.

I: Ok, and like all of the cases that you had were all sort of like real data in the sense that you created them yourself?

P: Yeah, exactly. All of them.

I: 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: Then, we will now start with the fifth block - and the final 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: Can you say that again?

I: Did you consider fabricating these data a difficult task to complete?

P: Difficult task? I don't get this. What is the question?

I: Was it like ...?

P: Oh whether I think it was difficult?

I: Yeah.

P: No, I think it was very easy because you were asking to fabricate an effect which actually exists. It would have been much more difficult - and much more interesting, I think, for you - to give me the task to create an effect which is actually not there. Because then I couldn't do it that way.

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

P: Yeah. It is going to be impossible.

I: Ok. So you can't think of any way how it could be detected?

P: No but like by logic: If you detect it, your method is flawed.

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

P: Yeah, I thought it would be fun just to think about how to do it. And then I already read everything and then [?], now, I can do it as well.

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

P: Yeah, a couple of people. Like when you sent the email around, people were like talking about it.

I: But did these people help you in fabricating the data?

P: Nonono.

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 think I told you exactly how I did it.