

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: No.

I: And how many years has it been since you got your PhD?

P: 8.

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

P: I am employed in the section of social psychology in the department of experimental and applied psychology but I would probably best describe myself as in between social psychology, personality psychology, and health psychology.

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

P: No.

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

P: I have been a participant in a pilot study that used the Stroop task when I was a PhD student. And I have seen demonstrations of the Stroop task but that's it.

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 only **use SPSS at least once per week.**

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: With peers being general academic colleagues in my department, an 8.

I: With peers, we mean relative to other researchers or scientists in your field.

P: Ok. I will say an 8.

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

P: 2. I have high confidence in your ability to detect it.

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 actually fabricated it a few times. I did it a few different times. Each time that I did it - and in fact I actually did it again today to keep better records and notes - each time I did it it only took me about 10 minutes but there (?) were single session fabrications.

I: Ok and how many fabrication sessions did you have?

P: 3.

I: 3, ok. And all of those on different days?

P: Yes.

I: Ok, so in total you worked on 3 days and the total time would be 30 minutes-ish.

P: Yeah, I would say probably 60 in total. The first time I thought about it quite a bit more and the second two times I just replicated what I had done with better notes and records.

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 data?

P: I did a little bit of review of meta-analyses in the Stroop task.

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

P: 20-30 minutes.

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

P: No.

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

P: No.

I: Do you know some cases of data fabrication?

P: Yes.

I: And did your knowledge of those cases influence your approach to fabricating the data?

P: Yes.

I: And how?

P: I am familiar with general aspects of what went into detecting Diederik Stapel's data fabrication and alleged fabrication of Jens Foerster with unrealistic consistency in results and standard deviations. Yes, so, I am also aware that individuals are intuitively poor at generating random values. And so I took that into account in generating data.

I: Ok. And you said that you thought about like approaches how to do it. What were the approaches that you considered?

P: I thought about simulating data in R, but my R skills are pretty poor. And I realized it would have taken me several hours to do that and I didn't think that I could invest that amount of time. Maybe, that answered your question. Can you repeat the full question?

I: Like, so, the questions was more like, you said earlier that you thought about like ...

P: Yes.

I: ... how to do it and my question was then which kind of approaches you considered?

P: Ok. I thought about generating data in R from two distributions with the means that I had estimated based on the meta-analyses. I rejected that based on the time investment it would have taken me. I thought about just coming up with numbers out of the top of my head. I also rejected that because I thought that those numbers would be very artificial. So, I did a bit of a compromise in that I took real data that I had observed with a completely different variable and I transformed that in an - manipulated the values, created new values based on those to match with the means and standard deviations that I was hoping to get.

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.

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. So, the first question is: Could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: Excessive similarity in standard deviations or variances between two groups. Too little variation compared to what under (?) real data we should expect in terms of a Stroop task. Excessive similarity in values between the two conditions. So if there was simply a linear transform of group 1 relative to group 2 - I just added a value of 5 in milliseconds to everyone in group 2 and (?) everyone in the incongruent relative to the congruent task. Those are some of the top of my head.

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

P: Not other than not violating those.

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

P: Only to the extent that I decided not to fabricate it of the top of my head and instead relied upon data that I collected for another context that did not have those types of biases that I might introduce.

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

P: I don't fully understand the question.

I: Ok, sure. So, we mean for instance the distribution of the scores or other aspects that could be inspected with the data set.

P: No.

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

P: I simply relied upon having used real data as a baseline that I would then linearly transform. And since the real data that I originally had were not fabricated, I thought that by transforming those I would avoid some of the pitfalls of - that might be detected based on me coming up with values.

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

P: Yes. I based estimated means on what I had seen in meta-analyses. And so I differentially transformed the two blocks of 25 individuals so that their means were what - not exactly what meta-analyses had indicated - congruent versus incongruent means should be - but approximately.

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

P: Nothing that comes up since I thought about this this morning.

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: 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: Yes. So, I took 50 cases from an existing data set where participants had completed an individual differences scale. I took the - the scale had 21 items - I took the mean of those 50 participants for the 21 items. For the first 25 participants, I divided the grand mean that I targeted for the congruent group by the mean of those participants and then I multiplied every individual mean by that value. And I did the same process for the second group. I took the mean of all individuals, took the grand mean of that, I divided what I had wanted to be the grand mean of the incongruent group by that value and I multiplied each person's mean score for the scale by that value.

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

P: Yes. I simply multi- I took the standard deviation of the scale scores for the 21 items on the individual differences measure for every participant and I multiplied those standard deviations by the average of the two transforms that I had made for the means so that the variances would be roughly equal between groups. It wouldn't be exactly equal but I wouldn't be multiplying the variance by a larger number for the incongruent group than I had for the congruent group.

I: Ok. And how did you select this individual difference measure?

P: Not with a lot of thought. It was something that was readily available, that I have used in a lot of my research.

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

P: Not entirely. I did do a few other modifications. I switched the highest value - I initially got a p-value of about .02 and I reduced the p-value to below .01 by replacing the largest value in milliseconds in the congruent category with the

smallest value in the incongruent category. So, I tinkered with the data a little bit further.

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

P: I eyeballed to see if there were any variances that looked extremely low or extremely high that might be suspicious but I didn't systematically compare them to a [?] variances that I would expect.

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

P: No. I didn't look at a histogram or boxplot of the fabricated results. I just eyeballed them.

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

P: Same answer as for (?) the previous question.

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

P: I went through the whole process three times but I did not specifically look to see if I had unique mean-and-standard deviation combinations. I briefly eyeballed the whole set of 50 values and I don't believe that I saw any identical means or standard deviations.

I: Ok and what were the differences between the three times that you did it?

P: It was mostly just that I had not taken the kind of careful, detailed records that I would want to give you in case you were asking for the notes.

I: Ok.

P: The first time I have just been thinking about the way that I wanted to do this and working through it without taking detailed notes. The second time I did take notes of the process, but when I looked over them this morning, I realized that I had not referenced the data set that I had drawn them from and specific cases that I may have switched - information that I thought that you may want - so I briefly did it again and took better notes.

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

P: No.

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

P: No.

I: Ok. So, the next question is: Did you use real data during the fabrication process?

P: Yes.

I: And so all of the data that ended up in the spreadsheet is based on real data and is only transformed?

P: Yes, yeah. Transformed and then two values switched between groups, but yes. And differently transformed for the two groups, yeah.

I: Ok, and did you use any like additional random noise or so or ...?

P: No.

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: Only that I didn't - I didn't use any additional procedures to mimic real noise because I assumed that the real noise would be apparent in the original data file which is - yeah - I didn't have the - I didn't have the ability to do that quickly or the time to learn how to do something to generate random noise for this procedure. So, I assumed that by taking and transforming real data there would be random noise inherent to it.

I: Ok, and the data that you used is it published data?

P: Yes.

I: Is it also openly available data?

P: These are not.

I: Ok, so do you think that like all the steps that you wrote down with your notes have been covered by the questions or do you think ...?

P: Yes.

Block 5: Underlying Rationale of Data Fabrication Process (Why?)

I: Ok, great. 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: Ok. So you have indicated earlier that you thought about different approaches and that you are not very confident in that your fabricated data will go undetected. Do you think that like fabricating it in a way that would have a higher likelihood to go undetected would be a difficult task?

P: Yes. Yes, I think that - I have trust that whatever process you guys are using to detect fabricated data would require a lot of sophistication to circumvent.

I: Ok. And do you have an idea how you could improve the quality of your fabricated data?

P: Yeah. I mean I think that I probably could have done a better job of - well, first of all, I didn't take into account anything related to the number of trials that you discussed (?) - I thought briefly about what the consequence of the number of the trials would be on the variance that I put in in the two groups. And I didn't take that into account at all. I am not sure that that actually would influence the variance of the two. I also wasn't sure about different variances between congruent and incongruent for the Stroop task. I didn't look into that. And it seems plausible that there could be more variance in incongruent because there might be greater individual differences in how the incongruency is processed. I didn't model that. And yeah, I am sure that there is other things that if I thought further, I could potentially try to neutralize some of the ways that these data might look fabricated or not correspond with expectations. But ultimately, I do have quite a bit of confidence in data [?] that you are going to be pretty good in detecting these things unless so much effort is put into fabricating data that researchers would be better off just collecting the data.

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

P: Wow, that is an interesting question. I guess I have a few thoughts on that. One is that I would guess that you guys are using an algorithm that will be equally difficult for different tasks - in terms of human hours to doing (?) - I assume that most of the effort is in developing an algorithm that can be implemented across different data sets with equal effort. I would guess that - well, depending on who has answered this call - so I don't know if you got a representative sample of academics to answer this or if only got people with a higher degree of quantitative knowledge. If you got a representative sample, I think that my approach would probably be - will more closely correspond with real data than average. If you got a sample of highly quantitatively sophisticated respondents, I guess that it would be in the lower half.

I: Ok, and could you think of ways how your data could be detected as fabricated?

P: Yeah, I would refer back to previous answers. I don't know if variances are expected to be systematically different across the two conditions. I suspect that they might be but I didn't look into it. Yeah, I don't know if there is some consequence of the number of trials that I didn't take into account in transforming the standard deviations. I am actually not even entirely sure - I didn't think about whether the linear transform for the means from the original values should also correspond with the linear transform of the standard deviations. So, I might have systematically over- or underestimated the standard deviations based on the real data that I transformed.

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

P: I generally support initiatives of detecting data fabrication. I think that is important. So, I am happy to contribute data points to you. I thought it would be fun. I used to teach statistics when I was a PhD student - just basic ANOVA and regression classes - and I had to "fabricate" - I am using the air quotes

for fabricate - example data sets so that we could have easy hand calculations for computing sums of squares and F tests by hand with small data sets. And so, yeah, I had to make up those numbers to make them good for instruction purposes. And I am kind of nostalgic of times when I did that. It was a lot of fun to teach those classes. And so I thought, yeah, I could use a similar approach to fabricate a data set for this project.

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

P: Only the professor I share an office with. We got the email invitation at the same time. And he expressed being too busy to participate, I believe. And I said that this sounds like fun and a worthwhile initiative and that I was going to do it.

I: Ok, and did this person help you in fabricating the data?

P: No, no one helped me

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 process that you think is worth mentioning?

P: No. I should say that I am happy to provide the original data and the notes that I took in case you want to reproduce what I did.