

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, cognitive psychology.

P: Social psychology. I don't know if judgment and decision making would be considered a separate one. But social psychology.

I: Ok, thank you. Did you conduct any experiments including a Stroop task in your career so far?

P: No.

I: Do you have any experience with the Stroop task or?

P: **I have read papers on it. I know the basic effect and the basic task from my education. But other than that, I have never used it in any way.**

I: Ok, thank you. 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: At least once a week, **R** definitely. Definitely not SPSS. I am using more and more JASP, but not really once a week yet.

I: Ok, thank you. And how would you rate your knowledge of statistics relative to your peers on a scale from 1 extremely poor, to 10, excellent?

P: So, my peers in the academic world?

I: Peers means relative to other researchers or scientists in your field.

P: Huuh. That is a tough question. I am gonna say ... **5** even though that is boring.

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

P: Let's say **4**.

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

I: Ok, thank you. 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: **One day.**

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

P: **2 hours.**

I: 2 hours, ok. 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: 2, ok. Did you prepare in any way before starting to fabricate the data?

P: **No.**

I: Ok, so you did not read like literature on detecting data fabrication or on previous cases of data fabrication or so?

P: No.

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 come to your mind and 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 process. So, could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: That is a good question. I thought about what these detection mechanisms or tools could look like and I actually have no idea. I know from the tax evasion literature Benford's law like the occurrence of different numbers that there is a certain distribution that you can look at, but I wasn't sure if that would really apply to response times data in the Stroop task. So, first, I thought about fabricating completely from scratch a dataset that kind of adheres to a certain distribution of numbers at given positions. But I was unsure if that really is necessary for Stroop task data. So, I didn't really prepare in any way. I didn't look at real data to figure out what the distribution of numbers is for example, or to figure out from a bottom up approach what makes real data look real [?] that perspective.

I: Ok aehm ...

P: Not sure if I answered the question.

I: Yeah. So like these questions are not so much about like what you used but like whether you in general could think of like characteristics that would make data look fabricated or more fabricated.

P: Yeah. I guess if we talk about data in general - and not only response time data in the Stroop task - if we look at Benford's law, the distribution of numbers - I know that in one paper, I think, the issue was - where they detected that it was fake data was that the variance in the reported standard deviations, I think, was - that it was enough variance in the reported standard deviations, whereas there was a lot of variance in the reported means. I think, that was one of Smeesters' papers. But other than that, of course, if you look at what power the design had relative to the effect size that they have, so in that sense data that looks p-hacked or too good to be true, that makes me think that the data is fake or was botched with (?) in a different way, but when I think about the actual numbers or distribution of numbers, there is nothing else where I would say this really looks fake.

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

P: I guess the noisier it is. If I can detect patterns in the data - [?] the numbers are the same in condition one and two - this might be a really obvious example of fake data - but condition one and two have exactly the same data points, only in condition two four or five are changed to way higher numbers that conform with the hypothesis, that obviously looks fake. So, the noisier the data is the more I would trust it. But that is only for this dumb-detection of 'hey, this is exactly the same data only some things are changed'.

I: Ok. And you already said that like you - but you didn't use these characteristics in your approach to fabricating the data?

P: No, because - I thought about these things, but since I didn't make data from scratch I didn't have to look at certain characteristics of what real data looks

like and then try to recreate that in data from scratch. But because I didn't fake the data from scratch, I didn't use that.

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

P: What do you mean?

I: Like for instance the distribution of the score or other aspects that could be inspected with the data set.

P: So, I looked at the mean response times for each participant and the standard deviation in response times for each participant. I looked at the standard deviation within these distributions to check if there was more variation in the mean response times compared to the standard deviation of the mean response times. I am not sure if there - a certain relationship should be expected if it is real versus if it is fake but I wanted to make sure that there is enough variation in both.

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

P: I guess common sense because - so, maybe I should mention that now because I didn't make the data from scratch, I identified real data, downloaded it, and then added noise to that real data to make it look less or not like that original data set anymore. So in - my reasoning was that there might be a lot of these characteristics that real data has and fake data doesn't have, but if I take data that is real, already, I should be fine because it is real data. If I then add random noise on top of that that is truly random, then none of these characteristics should really be detectable in the fake dataset because it originated from an original data set plus noise.

I: Ok. So after you did this, did you like go through certain checks or so to see ...ja?

P: So, I - because I - the only thing that I could really come up with, that I have read about is Benford's law, I thought that it wouldn't really apply in this case, but I did think looking at my data that I have - I don't know - 4 or 5 decimal places in there that depending of course on how these things are measured that there should - at fractions of .001 milliseconds the instruments are probably not precise anymore. So, one thing I did look at and that is also in the code is to, I think, extract the last digit that I had, which was, I think, the fourth decimal place and look at the distribution of numbers because I would expect that it is pretty random. Because just of measurement error, I guess, in precision of the methods, the measurement tools. It's - looking at the distribution, I mean of course it is only 25 subjects, I think, so it was pretty uniformly distributed. Not as much as I would have expected but I also didn't change it. So, I didn't take a new random sample. I just left it like that. Because again, I would guess that taking a random sample from a real dataset and adding noise to it that should get me - make me go undetectable - more or less.

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

P: Yes, that is another thing where I - I guess I could have read up (?) on this. But I didn't partly because of the time investment, but also partly because I think many people who fake this data - they are also not 100% sure of this methods. So, just adding noise to the data - the - I drew random numbers from a normal distribution with a mean of zero and added those numbers to the real data. I - the distribution for the mean had - I specified a standard deviation of 10 whereas for the standard deviation I specified a standard deviation of 8. Just looking at the numbers so that they are not just by random chance skewed too much. Because of the small sample it might look unreal or weird in the end because there is only a sample of 25 subjects, but that was more gut than real statistical thinking.

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

P: I guess another way which would have cost more time which I didn't do in the end is - what I could have done is inspect the real data set more and there was much - there were thousands of subjects in that dataset, I could have kind of from a bottom up approach look at certain distributions or characteristics of that data set and check if in my final data set these characteristics are there as well. I could have done that but I thought, hey, it's random sampling from that and I should be fine.

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 process that come to your mind and you think could be interesting for us to know?

P: I guess - since it's Chris - thanks to Open Science I had a lot of real data that I could sample from. So, I didn't have to do it from scratch because thanks to Open Science there is a lot of real data out there that I can just manipulate. So, ironically, even though, of course, this is one disadvantage that we should happily take - one cost - thanks to being more open, I have a lot more real data to sample from and make my data, make my fake data out of.

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: So, I remembered that in the many labs project they replicated the Stroop task. I googled that, found the many labs free data set where, I think, they had - I don't know 4000, 2000 participants who accross many labs engaged in the Stroop task. I loaded that data set into R, sampled a random subsample of 25

subjects from that. That was also within-subjects, so 25 in each condition, or (?) 25 who completed both conditions. And I took those means.

I: Ok. And you said that at the end you added then some random noise ... ?

P: Yeah, so I took a random sample of 25 subjects and I created - and also saved for each of these 50 means a random number drawn from a distribution with a mean of 0 and a standard deviation of 10, added that on top of it, and that was the last step. That was the fake data, then.

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

P: Same approach. Real (?) - start with real data. I mean because it was not on the subject level. So, I took for the whole data and I made mean latencies and mean and standard deviation of the latencies for each subject. I sampled the 25 standard deviations that were connected to the means and, again, put noise over it sampled from a random distribution with a mean of 0 and a standard deviation of 8.

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

P: Nope. I basically took the first sample, the first set of random numbers I got. And - looking at the distribution of the last numbers for example - even though it was not perfectly uniformly distributed - not like any test would probably detect another distribution - but I figured it was random sampling, it should be fine. It is how it works in the real world. And if I would have sampled 25 subjects in the real world, that's the same. It has the same characteristics you should get.

I: So, you were satisfied with the first fabricated data set that you had?

P: Yes.

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

P: Yes. Like we talked about already. I looked at some distributions, but not many and even I did not have strong expectations of what it should look like so I looked at it and said, yeah, ok.

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

P: Same answer there.

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

P: There were all different since they were sampled from real data.

I: Ok but like in the end you had like one combination and you didn't like change it afterwards that you ... ?

P: Yeah, I didn't make multiple sets of 25 sets of data and then chose the best one.

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

P: R.

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

P: Only for the noise added on top.

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

P: Yes.

I: How much real data did you use?

P: All of it.

I: Ok, and the next question is: How did you use these real data?

P: Many labs project, take a random subsample, add noise on it.

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.

I: Ok. And why is that?

P: I guess the - what I wanted to do first is really look at the characteristics of real data and then create fake data from scratch that resembles these characteristics would have been more interesting from a programming/coding standpoint. But since I took real data it was easy.

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

P: I mean people are always overconfident. So, I don't wanna be overconfident because I don't know what tools are out there at all. My intuition says it is

going to be hard to detect because it originates from real data but I don't want to be too confident.

I: Ok. Like any specific reasons why you think that like it could be detected or not detected?

P: I believe in the superior competence of statisticians to know things we don't about data and, I mean, in the last years, you have seen some very smart tools that looking at detecting p-hacked or fake data so I don't have any - I am totally in the blind here what could be out there, what kind of methods.

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

P: First reason was I wanna write some cool code that simulates data because we have done some simulations in the last months where we do have to simulate data with certain characteristics. So I - that was probably the biggest reason. There was another step of simulating data with certain characteristics. I didn't end up doing in but that was probably the main motive. Of course, the money also helps.

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

P: My exact method, no. But I talked to a bunch of people about this, yeah.

I: Ok, and did these people help you in fabricating the data?

P: 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: 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 don't think so.