

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: 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: No, ok. How many years has it been since you got your PhD?

P: A year and three months.

I: Ok. And what is your field within psychology? For instance social or cognitive psychology.

P: Cognitive psychology.

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

P: **I did do experiments, but not the Stroop task.**

I: Ok. And could you describe your experience or your knowledge about the Stroop task a bit?

P: Well, the Stroop task is well known. So, I don't know how exactly I know it. But, of course, we learned about it in the study already. But, yeah, I couldn't say where specifically I got the knowledge about the Stroop task - except that it is really well known, so yeah.

I: Yeah, ok. So, you would say you have some basic knowledge about it, but you have never used it for your research?

P: No, yes, that is true.

I: 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: Nowadays, I mostly use 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: I would say a 7.

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: A 6.

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 spread it over several days.

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

P: 3 days, I would say.

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

P: I think, 4-5 hours.

I: 4-5 hours, 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: An 8.

I: Oh, from 1 to 7.

P: Oh, sorry, I wasn't listening. Then a 6.

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

P: Yes. I prepared by looking up articles about the Stroop task, so research articles.

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

P: 1.5 hours.

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

P: No.

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

P: No.

I: Ok. Could you like explain a bit more how the preparation that you did influenced your approach to fabricating the data?

P: Well, I had some knowledge on how, for example, Diederik Stapel did it, because he was quite well-known. And I know he made some mistakes with like ordering the data, like from top to bottom and bottom to top. So, I knew that I had to take care not to do that. But I thought it would be easier if I just would go about it as I do - as I would if I was going to fabricate the data myself. And I thought, it would be the best starting point to at least know what sort of normal data would be. So, I started looking up the articles, some articles on the Stroop task, to get a general feel of the response times in both conditions. And I looked up if there were data sets available on the Stroop task to see if I can get some trial data from participants. Did that answer your question or?

I: Yes, yes. 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, for the timeline, I think that's it. I just looked at articles, got some idea of the minimum, maximum sort of response times that are usual in these both condition so I had some sort of knowledge on how it could vary and then I just started.

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

I: Ok, let me just check whether it is still running, yes, looks good. 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, the first question is: Could you name specific characteristics that would make data look fabricated or more fabricated in your opinion?

P: Yeah, so what I mentioned. Yeah, I knew that Diederik did some weird things where he would just order columns one specific way, which would look very obvious, apparently. I thought it would be important that the results weren't too nice or too neat or that every participant was too perfect, so there should be at least some weird results or outliers. Yeah, what I struggled with more was that the instructions were not that specific. So, it - a lot of questions came into my head like what kind of participants are these? Are these young people or patients or old people or a very varied group? Or was this the vocal version of the task or a computerized version of the task? I thought I had to at least try to make the participants a little bit consistent. Because I started out with just generating random values within like certain limits, but then if you do that for the two conditions and one participant has like a huge standard deviation in one condition and a very small in the other, I would think that was kind of maybe weird that some - at least within participants, people should be a little bit consistent. What else did I look at? I think those were - yeah, maybe later I will come up with others, but yeah.

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

P: More what? Genuine?

I: Yeah like more true, more real.

P: More true. Well, as a researcher I know that real data is ugly or random or makes no sense. So, I did think about that. But then I thought well if you do - I mean, the more real something is, the better the data. And the Stroop task is such a clear effect that I didn't have to worry that much about people who showed the totally opposite effect, for example. But if you would have a less strong effect that you would research, then what you always find is that people have weird results that make no sense to your hypotheses or that are weird or ... When it comes to response times I know at least that people are never quicker than 150 milliseconds. I just know that from my work, I did [REDACTED]. So, I know that, you know, what is a very long response time that you would always - that would almost always be too long, like 1.5, 2 seconds is really long, less than 150 milliseconds is too short. So, you would have to at least take care of that a little bit. But I would say that there is always an outlier or 2 both ways, maybe 3, that there has to be at least some outliers also.

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

P: Yeah, the outliers, I did. So, I made sure that there was a few people who showed the effect - eh, who almost did not show the effect, that there were people who were slow, and one person who was quick. So, at least that, you know, not everyone did perfect experiment, because it never happens. You can always have a participants who just goes - that just pushes the buttoms randomly without even checking the experiment or someone who thinks about it really long and is really slow. But I did - yeah, I didn't once like the opposite Stroop effect because I thought that was very unlikely. So, I just - I made sure that everyone had some sort of Stroop effect but some were smaller than others, so yeah.

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

P: I did, because - not sure what you mean with - is it like age or?

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

P: Yeah, I didn't - I thought about the distributions, trying to look up if there were some sort of knowledge on like, well, 10-15% of people score within this and this range or not. But in my mind, it really also correlates with your participant sample. So if you have a very homogenous sample, like 20 year old psychology students, they are probably more similar than if you have a group varied in age or. So, in my mind, I thought, ok, well, I would most probably be a researcher at a university with access to young people, so I will just think of them as having some sort of a similar effect in that sense. But, yeah, like I said, I did try to look at the standard deviations and such, that they are sort of a little bit similar, but they are always a little bit smaller in the congruent condition, because that

is most likely the case. But yeah - but I also didn't know like - because some researchers take out the wrong answers and some don't so that also had an effect on your standard deviation and your response latencies. So, yeah, like to answer a lot of things came to mind, but I didn't know what the task was, so I just decided to go for the [?] version.

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

P: Well, I tried to make sure that the participant data was sort of within range of what is normal in the literature so that people wouldn't be extremely fast or extremely slow. That some people would have the effect more strong than others so that not everyone showed a huge Stroop effect. That there were some outliers that were sort of almost or within 2 standard deviations of the mean, because that is the range people normally use. That normally there is someone who is at least 2 standard deviations removed. And I was hoping that that would sort of make it more real. And that the effect was consistent. Because if you would really randomly generate the data, then you would have people who have the effect, you know, the other way around and you can still have a significant result. And I thought that would be unlikely.

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

P: A little bit, yeah. Because I thought the standard deviations should be a little bit bigger in the incongruent condition. And I thought those should be a little bit similar between participants, not too much, but that at least there wouldn't be any huge differences that you would get if you really randomly generate the data. Yeah, I think, if you make a comparison between the averages and the standard deviations, that's what I looked at. And I also tried to make sense of if you had really quick - really low average like someone who was really quick, that they didn't have like a huge standard deviation or at least I thought that would make sense. But I actually thought that faking the data would be a lot harder than just programming the experiment and doing it.

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

P: Yeah, in hindsight this morning, I thought, maybe I should have put extra effort into looking up the normal distribution of scores. So, looking at how many percent of people would normally fall within a specific range. But I thought, well, let's just - because I would start all over again.

I: Yeah, 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: I don't know. I am not sure if I am the only one but I worried a lot about how many values I should have behind the comma and if they should be similar or not, if people would look at that or. So, I don't know, I became very paranoid

of oh my god, what would be - like weird or if I had one outlier, where would it - would it be participant 24 or 3, would that matter? Like normally, you don't even pay attention to these things and now I thought, oh my god, if you really were committing fraud, this would worry me like can I make participant 25, like the last one, an outlier or would that be weird? I really - I found that was quite stressful actually.

I: And how did you determine for instance how many decimal points you would have or like which participant should be the outlier?

P: Yeah, in the end - because I started out with just generating data within certain values, so I just took the one that happened to be sort of an outlier and I just made it a bit more of an outlier. I don't remember the number of the participant, but - and with the amount of values, I thought they should at least be the same. So, I wouldn't have certain values that have like 7 things behind the comma and several that have like 2. Because then it would be, I think, very obvious that I have typed in something. So, I just thought, ok, I will just make them all 2. And then I worried about it, because I thought, no, I think more would be better. But then I didn't want to do it all over again to make sure that there were too many, so yeah, weird.

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 looked up quite a lot of articles with the, yeah, Stroop data. I just noted all the means and standard deviations that were reported. So, I had a table with, I think, 10 to 12, I think with 12, 12 results and they had a certain range which looked quite normal to me. So, I had - I think between 500 something and 800 something milliseconds for the congruent and a little bit higher for the incongruent condition. I also tried the Stroop task on myself a few times just to see if my results would be comparable to that and they were. And then, so yeah, for the means I did that. And then I sort of started out with generating random values between sort of the two normal response times that you could find. And then I looked at the standard deviation and developed some outliers based on the standard deviation.

I: Ok. And how did you generate the values? Did you do it by yourself or ...?

P: In Excel.

I: In Excel, ok. And like what kind of function did you use?

P: It was the `aselect` function. I did it in Dutch, so it was - and I always forget the names in Dutch, but you have the - like the `rand` - in Dutch, it is `aselect`

and in English it is rand with open closed brackets. And if you then - because that normally generates things without any decimal points, but then if you do it - make the function like the rand() by sort of the difference between the two values plus the minimum, then you get decimals. So, I thought it would be better to have decimals.

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

P: I did the same. Because I had all the standard deviations that were reported in the literature. They were a little bit more - I took a little bit more care with them, because with the averages it is sort of ok where they are and also with the incongruent condition. I just made sure that I checked that people had really the Stroop effect, because there were a few that did not have the Stroop effect. So, I changed those, but so at first I also generated the standard deviations in the same way. And then they were already of course for the congruent condition a little bit lower than for the incongruent because that is also reported in the data. And then I just made sure that - I just changed a few to make sure that there are a little bit more similar to each other, because sometimes there was a very big difference within the participant. So, I sort of checked that and then I plotted the data just to see how it looked.

I: Ok. And did you create the values for one condition independent of the other condition or did you like first create it for one condition and then for the other or?

P: Yeah, I first started out doing them independent, but then you get sort of weird results. So then, I thought, no, within-participants, it should be at least a sort of correlation between both the averages and the standard deviations. And so - and I didn't want also - because I didn't want the data to be too neat, so I did want some big standard deviations in there as well. But yeah like I said, I made sure that there were several degrees of strength in the Stroop effect, but that everyone had the Stroop effect. Yeah, so in the end, I did - I mean I did generate it randomly, but then I did change a few to make sure that they made more sense.

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

P: Yeah, I did tinker with it, I did - I plotted it and yeah.

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

P: Well, at some point I thought I was happy with them and then I put them in the template and then I had the biggest significant effect ever. And I thought no, this is too much. So, I did try to make it - to then turn it down a little bit. So, I think it is a little bit with that as well. And I tried to make sure that there were several sorts of outliers like one person who would be very slow in both conditions and one person who would be like way slower on the incongruent

condition and the rest sort of that kind of things. To not make it look too nice. And I wondered how strong I should make the effect. Because it is quite easy to make a really significant effect with a really high t-value, so I tried to lower it a little bit. But then I didn't know actually what, because a lot of the old literature doesn't really report that many t-values anyway. So, I was like hmm, then I struggled with that. Like ok, because this is such a well-known and such a strong effect that it could very well be really strong, so yeah, at some point to be honest I thought you just now have to quit, because they said you should spend like half an hour on the data fabrication and you are already ongoing like 4. So maybe, you should just let it go and be happy with it now.

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

P: I really don't know where to get fabricated data - or do I answer your question wrong?

I: Like whether your fabricated data looked weird?

P: Oh yeah, I plotted it. And I thought, now it looked sort of ok. So, I hope that it is.

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

P: Yes, I did. I compared it, it was within my limits that I took from the literature. So, I had - I did have a strong idea of what the averages should be sort of or within what limits they should be and the same with the standard deviations. So, I thought if an article reported now my averages and my standard deviations, I don't think people would find them strange. It is more the trial data, the participant data that worried me. Because that was also hard to find, people always report their averages but never their trials.

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

P: How many, I wouldn't know.

I: I mean like now you have a template and like did you often like change the values for the 25 participants that you had or did you just use the first final version so to speak that you created?

P: No, I think, once I had data that I was happy with and I put it in the template, I changed it maybe 4 times more to sort of try make the effect a little bit less significant.

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

P: No, only Excel.

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

P: Yeah, the one within Excel.

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

P: Well, I used the data that was reported in the literature.

I: Ok and how did you use it?

P: At least that I knew the boundaries of the averages and the standard deviations. So like the normal strength of the effect.

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: I don't think 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, the first question is: Did you consider fabricating these data a difficult task to complete?

P: Yes.

I: What did you find difficult about it?

P: There is just so much you have to think about and I thought, yeah like I said before, I - it would have been so much easier for me to program the experiment and do it with at least - or do it 25 times myself or with a few friends. I would have been done within half an hour, I think. To get real data. Because this is such an easy task to program and to do. It would have been done so quickly that, yeah - I mean, we had a laugh about this actually, because I have a lot of friends [REDACTED] and I said, oh my god, it would be just so much easier to do this for real. And then they said, oh, you would be fabricating your data for your fabrication experiment, because you have to fake it, so, yeah, that would be funny. But, yeah, I thought it was hard and stressful in a way, because you wanna do it well. And you are like, there is so much, I don't know. Like who are these people, are they quick or nor, or are they students, are they motivated to do it well, or are these people who just wanna get it over with or - because I know these are all things that really influence your data quality. So, I thought it was, yeah, it was difficult. And then I noticed, I never really thought that much about my own data in a way. That if you have your own experiment, you have such a real sense of what it - what makes sense, and you have really strict things that you think upfront, like what do I think will be an outlier, you know, who will I take out, but I never thought about my standard deviations in very abstract way like they should be this high, but so I don't know, I thought it was interesting but difficult. And I don't know why people would do this.

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

P: I don't know. I hope that it will, but I am sure that there are - that you guys will look at very different things than I would come up with. So, I am very curious to see how I did. I don't think - I think - I don't know - I don't know. To be honest, I hope you won't (?) see it, but I am not completely certain.

I: Ok. Do you have like a guess at like what could tell us that the data is fabricated?

P: Is fake? Well, I still think that maybe the effect is too strong, or maybe I should have made more outliers, or maybe I just have no real concept of how average data across participants should really look like. Maybe (?) I made really stupid mistakes with my pattern in the standard deviations compared to the averages. I don't know. I thought I was quite good in statistics, but maybe not good enough to really understand how to fabricate them. And I still worry about the decimal points like how would people normally save this data or have this data? With how many - or would it be totally normal within such a template to give it with 2 decimal points or not, I don't know.

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

P: Yeah. I thought it was really funny, like a cool challenge. Yeah because I have many friends [REDACTED], I really try to follow the research that they do and I also think Chris is doing really cool studies. I think within me, it is sort of a race like oh this would be so cool, this would be your chance to see if you could do this or not. And I don't know, I noticed with my friends that they had the same response like we thought this would be a really cool game like to see if you could beat the system to see if you could be really good at faking it. And to be honest, when I started, I thought I could do this a lot better. Then I started it and I thought, oh, this is more complicated or more stressful than I thought it would be. But I don't know, I think it is really cool. I really hope that Chris can make an algorithm that can detect it. That would be really cool. I don't know I think it is a really nice study.

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

P: No. Within - because a lot of my friends are doing it as well, but we immediately made a pact that we would not share anything and that we would really do this individually.

I: Ok, so like no one helped you in fabricating the data?

P: No. It was - I mean sometimes I would want to ask like how did you do this, but I thought no you have to do this individually, otherwise we would ruin Chris' study and we didn't want that, so yeah.

I: Ok. Then this is ...

P: But this is the first thing I am going to do when everyone is done. Compare tactics.

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 don't think so.

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: I don't know except that I thought that it was quite stressful and difficult and that it would have been a lot easier to just do the experiment. I thought, it was difficult.