Statistics and Psychology: what psychological scientists want to know from statistics and the hidden influences of statistics on the discipline

Sven Ulpts

S3300854

Behavioral and Social Sciences Research Master Master Thesis University of Groningen First supervisor: Rink Hoekstra

Second supervisor: Henk Kiers

Abstract

Throughout the history of psychology numerous scholars have claimed that the dominant statistical techniques, such as NHST, in this discipline do not provide researchers with the kind of answers that they are looking for and many have speculated about what psychological researchers really want to know from statistics. However, there is a general lack of actual study into what psychological scientists want to know from statistics. In the current study I interviewed one former and two current APA division presidents and one editor for a psychological science journal to inquire what they want to know from statistics. A second question that the interviews were intended to answer is whether and how the statistical methods that psychological scientists are used to influence the kind of research questions that are formulated and investigated in the discipline. The results indicate that there is not just one thing that psychological scientists want to know from statistics, but rather a variety of different things that depend on the context and the subject or problem of investigation. However, there was some support for the claim that between NHST and Bayesian statistics the Bayesian approach does provide kinds of answers and information that are more in accordance with what three of the four interviewees want to know. Furthermore, interviewees regretfully stated that the statistical methods they are used to influence what kind of research questions they consider and how they formulate them.

Keywords: Research Questions, NHST, Statistics, Bayesian statistics

Statistics and Psychology:

what psychological scientists want to know from statistics and the hidden influences of statistics on the discipline

Statistical methods are the backbone of psychological science (at least experimental psychology), and they have been for decades (Danziger, 1985; 1990). What these methods are good for is mostly a mathematical question and out of scope for this article. Furthermore, the question of what can be concluded, inferred, and what kind of knowledge can be gained, based on certain methods, such as Bayesian and frequentist statistics, might generally be of secondary importance. The prime question in this paper, is what it actually is that psychologists (active scientists in the field of psychology) want to know from statistics? Numerous scholars, especially statisticians active in psychology and related disciplines, claim to know what psychologists actually want to know. For instance, according to Colquhoun (2017), what researchers really want to know is the probability of obtaining a false positive, called the false positive rate (FPR). In other words he thinks that scientists want to know what the probability is that they actually made a discovery instead of capturing a chance occurrence (Colquhoun, 2017). Lakens (2021) lists six different positions from six statisticians about what they assume psychologists want to know from statistics. Namely, "what we want to know is the posterior probability of a hypothesis, the false-positive risk, the effect size and its confidence interval, the likelihood, the Bayes factor, or the severity with which a hypothesis has been tested" (Lakens, 2021, p. 640). The speculations about what psychological scientists want to know from statistics often happen in the context of a dissatisfaction with the widespread use and misuse of NHST in psychological science (Nickerson, 2000; Spence & Stanley, 2018; Szucs & Ioannidis, 2017). The probably most direct statement against NHST and about what psychologists want to know stems from Cohen: "What's wrong with NHST? Well, among many other things, it does not tell us what we want to know, and we so much

want to know what we want to know that, out of desperation, we nevertheless believe that it does! What we want to know is "Given these data, what is the probability that H0 is true?"" (Cohen, 1994, p. 997).

However, this claim and others of its kind are made without any empirical support. They remain conjectures and pure speculation or as Daniel Lakens calls them "wishful thinking" (Lakens, 2021, p. 640). Therefore, such statements require actual empirical investigation. Likewise, gross generalizations in the form of "what we really want to know is" (Blume 2011, p. 509) might be unwarranted, without the consideration of the research questions that motivate the use of certain methods in specific contexts of investigation and sub-disciplines that coexist within psychological science (Lakens, 2021). Daniel Lakens calls those unfounded statements the "statisticians' fallacy" (Lakens, 2021, p. 640), he describes it as follows: "... a declaration of what they believe researchers really "want to know" without limiting the usefulness of their preferred statistical question to a specific context." (Lakens, 2021, p. 640). Hence, judgments about what psychologists actually want to know from statistical techniques remain outsiders' opinions without the support of empirical evidence and the consideration of the context in which said methods are applied.

Past research has mostly focused on widespread misinterpretations of statistical methods and indicators, such as the p-value (Badenes-Ribera et al., 2016; Haller & Krauss, 2002; Lyu et al., 2020) or the confidence interval (Hoekstra et al., 2014), instead of on what psychological scientists would like to know or be able to conclude. For instance, Haucke and colleagues (2021) re-investigated with a questionnaire, which was originally used by Haller and Krauss (2002), the prevalence of misconceptions about p-values among 117 scholars who published in either of five psychology journals between 2015 and 2017. However, it is important to bear in mind that it does not follow from the commitment of certain misinterpretations that the researchers who committed them actually want to be able to make

such interpretations (conclusions). Those misinterpretations might, for instance, also be a consequence of misunderstandings and bad statistical education in the form of textbooks (Cassidy et al., 2019), teachers, or even the APA¹. More importantly, Haucke and colleagues (2021) also investigated in the survey how much those participants wanted to make the conclusions (inferences, statements) that are expressed with those misconceptions. They found that on average there is a discrepancy between what researchers (the participants) think they can conclude from p-values and what they want to conclude from p-values. What they want to conclude seems to be more congruent with Bayesian statistics (Haucke et al., 2021). According to Gigerenzer (1993) this discrepancy between what they want to know stems from an inner conflict. In his Freudian analogy the superego and the ego of the psychologist remain in the confines of frequentist statistics while the id continuously tries to break out and make Bayesian conclusions.

Another reason why it is crucial to inquire what researchers in psychology want to know from statistics is the ongoing discussion, concerning to whom the conclusions and inferences made, based on certain statistical methods, (should) actually apply (Adolf & Fried, 2019; Fisher et al., 2018; Molenaar & Campbell, 2009; Voelkle et al., 2014). Talking about non-ergodicity and the "ergodic fallacy" (Speelman & McGann, 2020, p. 1) (or ergodic switch), where investigators use methods which only allow conclusions about groups (populations aggregates and samples), but frame their conclusions and inferences in a way that is tailored towards the individual, is accumulating. The claim is that the main occupation of psychology is with statistical methodology that is mostly only applicable to group data, while practitioners and clinicians are (or might be) more interested in information and

¹ The explanation of significance testing provided by the APA contains at least two misconceptions (see https://dictionary.apa.org/significance-testing) .

recommendations applicable to the individual patient, this is also sometimes seen as contributing to the scientist-practitioner gap (Falkenström, 2022). Moreover, whether the conclusions should be applicable to the individual or the group (or both) could also depend on the sub-discipline in which the investigation takes place. If psychological scientists would want to know about groups instead of individuals, then the problem of non-ergodicity would not be their problem, but if they would want to make inferences about individuals, then the problem of non-ergodicity is a problem they have to consider.

Unfortunately, the situation in psychological science seems to be that the statistical methods at our disposal determine which kind of questions we are allowed and able to ask as well as which kind of conclusions and inferences we are allowed and able to make. Lakens somewhat captured this circumstance in his bathroom tile: "Statisticians can not tell you "what you want to know " - they can only tell you how to ask a question." (Lakens, 2017, para. 12). However, should it not be the other way around? Should not the questions we want to ask determine what kind of statistical methods we need and use? If yes, then a necessary precondition is to actually know what researchers in psychology really want to know from statistical methods. With the heavy reliance of psychological science on statistical techniques the question becomes paramount not just whether but also how the statistical methods at the researchers' disposal actually influence the questions they investigate and thereby ultimately influence the nature as well as direction of a discipline.

Therefore, the aim of this exploratory qualitative study is to investigate what active researchers in psychology, who have been in the discipline for some time and know about issues in the discipline, actually would like to know from statistics. A secondary research question is: how and to what degree do researchers in psychology experience being influenced (restricted or guided) by the statistical methods at their disposal and the statistical techniques they are used to?

Methods

Interviews

Participants were informed that the online interviews would be no longer than 40 minutes. The actual interviews were 38 minutes, 39 minutes, 40 minutes and 26 minutes long. The interviews were conducted and recorded with zoom. The interviews were semi-structured and included eight predetermined questions:

- *Can you start with telling me a bit about your research?*
- What is it that you would like to know from your research?
- When you come up with a question or decide to investigate a certain subject when do statistical considerations enter the process?
- Did you ever have a question or topic that you wanted to investigate, but did not have a statistical method that would provide you with the answer you needed?
- Do you think that the statistical methods that you know and you are used to influence the kind of questions that you ask?
- When you do a statistical analysis of your data, what is it that you would like to know from statistics?
- Would you like to know about individuals or groups?
- If statisticians were to develop statistical techniques and methods based on what you would like to know from your research and the data. What would you tell them?

I used the first question as a sort of ice-breaker to start the interview and get the conversation going before the more targeted content-questions are asked. In the case that interviewees did not know how to answer a question, I rephrased the question and provided the motivation behind that question. For example, when asking them whether they would like to know about individuals or groups I would say something about the problem of non-

ergodicity and that results from aggregate or group statistics might not be applicable to individuals.

Following the interview I transcribed the recordings. I did not transcribe one of the interviews, because the interviewee was a practitioner who does not publish empirical articles that use statistical methods. Therefore, I deemed it uninformative for the research questions. Next I used thematic analysis as described by the step-by-step guide from Braun and Clarke's (2006). I coded the transcripts inductively (Mayring, 2000) based on the meanings in the transcripts and (somewhat deductively) through guidance by the two research questions. Coding was done by hand without any programs. Subsequently I constructed the themes based on groupings of codes. Thematic analysis was done in separate coding (analysis) sessions for each research question to prevent confusion. Moreover, I analyzed the interviews for each research question twice for approval and asked whether they had something to add) was done with two of the transcripts, because the other two participants did not respond (in time), to improve the credibility and validity.

Sample

A total of 143 past, present and elected APA division (society) presidents and journal editors were contacted by email. The invitation emails contained the motivation behind the study, what is expected from the potential participants and whether they would be interested and have the time to participate. The reminder emails reminded the potential participants of previous invitations and again asked whether they would be interested in participating. All emails had the consent form and the study information attached to them. I interviewed five participants, four of which have been transcribed. I chose this sample, because I assumed that presidents of APA divisions (societies) and editors had to be in the field for a certain (longer) time to attain such a position. Therefore, I thought that those researchers would be immersed

in their research culture and have something like a birds-eyes-view due to their position (see Wijsen & Borsboom, 2021 for a similar sampling approach). Similarly, I assumed that those researchers would have been exposed and accustomed to the statistical methods that are typical in their discipline. Furthermore, participants should still be publishing empirical research that uses statistical methods (within the past five years). One of the interviewees did not fulfill this criterion which is the reason for the exclusion. This results in a response rate of three percent. One participant was a former APA division president and two participants are current APA division presidents while one participant is an editor. Participants were a developmental psychologist [4] an educational psychologist and methodologist [1], a coaching psychologist [3] and a clinical psychologist [2]. Email addresses were retrieved from the APA divisions' (society's) homepage or from the researchers' university's staff website. Participants were contacted via email between June 1st and July 21st. In the case of no response two reminder emails were sent. All participants were from the United States.

Results

After four coding sessions numerous themes and some subthemes were identified. The themes are, for instance: *Statistics influences the types of questions that are considered and formulated* (which was also the only theme that all participant agreed on), *NHST does not tell me what I want to know, prefer Bayesian* and *there is not one thing I want to know*. The themes are divided in two categories based on the research question they are related to. I formed nine themes and four subthemes that are related to the question regarding what psychological scientists want to know from statistics, while I identified five themes about how and to what degree researchers in psychology do experience being influenced (restricted or guided) by the statistical methods at their disposal and the statistical techniques they are used to. The themes and subthemes related to the first research question are: *There is not just one thing I want to know, what does my data look like, probability of replication, help make*

decisions, prediction, NHST does not tell me what I want to know, prefer Bayesian, priors, probability of hypotheses, strength of the evidence, Individual and/or group, individual and both, it depends on the situation and the research question. The themes in relation to the second research question are: limited by demands of the statistical methods, statistics influences the types of questions that are considered and formulated, the research question has primacy, statistics changes the question and too narrow set of statistical methods. In the following quotes from the interviews are used to support the themes.

What do scientists in psychology want to know from statistics?

Theme: There is not just one thing I want to know

One theme I identified is that there is not just one thing they want to know from statistics. Three participants said that what they want to know from statistics cannot just be one thing, because it depends on the research question, the situation and the nature of the subject or problem of interest. This also relates to what one participant [1] mentioned about there not being one statistical approach that is universally better than the others. Participant 1 also suggested that there cannot be just one thing psychologists want to know from statistics, because psychology is so multifaceted (complex) that one statistical technique is usually insufficient.

"... it really depends on what you want to know and what you're doing." [2]

"Well, it depends on what I'm studying, right?" [3]

"It really depends on the question you're asking." [1]

Theme: What does my data look like?

Although descriptive statistics is not inferential statistics, two participants talked about how important it is for researchers to actually look at their data and how statistics should be used to show the shape of the data. These participants mentioned how data description and

visualization can provide unique and important kinds of information that are not provided by inferential and more specific statistical techniques, such as, hypotheses and significance testing. In other words statistics should provide tools that allows them to look at the data, describe the data and inspect its shape.

"... look at the histograms of your data. What does your data look like? Is it bifurcated? Is it a normal distribution? Is it flat? You know that can tell you a lot of information right there." [3]

"... look at your data" [3]

"Sometimes you employ statistics to describe phenomena that's the only thing you want to do. I mean describing that's where the word statistics comes from study of the state, describing what the state of something looks like." [1]

Theme: probability of replication

When presented with some speculations from statisticians regarding what they want to know from statistics two of the participants reacted by stating that they would like to know something about the probability of replication. One participant directly linked the importance of this kind of knowledge to the current state of psychological science, more specifically to the problem of replication (the replication crisis). The other participant also hinted at that problem by remarking that actual replications also have to be conducted.

"So, I think the combo of yes moving towards statistics that tell us about replication, but also doing replications ..." [2]

"I mean like the replication stuff would be really important and useful. Psychology does have a replication problem and being able to say something about that would be really helpful" [4]

Theme: help make decisions

One participant mentioned that an important aspect of statistics is that it should enable researchers to make informed decisions. In other words statistics should provide knowledge in a form that aids in a decision making processes. However, the interviewee also quickly remarked that statistics should not be misused as an instrument that takes away the responsibility of the scientists to make expert decisions (expert judgment). Statistics should be an aid for decisions it should not be the sole tool for the decisions.

"I mean statistics is an extremely useful set of techniques and parts of that techniques can be very useful in making decisions, but the statistics should not be the decisions makers themselves." [1]

Theme: prediction

Another participant also talked about wanting to be able to make predictions about the likelihood of certain incidences or behaviors. The participant mentioned this need for making predictions in combination with the wish to be able to gain knowledge about individuals.

"I really do think it's about the probability or odds that a person is going to in the future have some sort of challenge or outcome. Really that would be, can I predict at least the near future. That would be mostly what I'm after." [4]

Theme: NHST does not tell me what I want to know

Two participants mentioned that null hypothesis significance testing does not provide them with the kind of answer that they are looking for. The problems with NHST that were mentioned are that it only provides a dichotomous type of answer (yes or no), in the form of rejecting or not rejecting the null hypothesis and that data and hypotheses are analyzed in isolation without acknowledging what has already been done as well as known. "I can remember when I was in psychology grad school when I was reading all these papers and they were talking about null hypothesis and blah blah blah. And I'm like what is this? All that is telling you is yes or no!" [3]

"we need to get more and more away from null hypothesis significance testing ..." [2] "... what I know is null hypothesis significance testing and I know that it's not telling me, it's not really answering the question." [2]

"So that when I'm doing null hypothesis significance testing it's just as if we know nothing about that topic. And of course we know lots about a lot of topics, some topics not so much. And it seems crazy that we go into each new study saying hmmm we have no idea." [2]

Theme: Prefer Bayesian

Directly linked to the theme NHST does not tell me what I want to know is the theme that captures a general preference for Bayesian statistics. Three of the four participants indicated that they prefer Bayesian statistics over NHST, because they seem to think that it is more in line with what they want to know from statistics. I divided this overarching theme into two subthemes that outline what Bayesian statistics tells them what they want to know. Those subthemes are the fact that Bayesian statistics is able to incorporate priors into the calculation process and its ability to provide information about the likelihood (probability) of certain hypotheses. First I will provide quotes for the theme and then I will provide quotes for the subthemes next.

"Yeah I am much more in the Bayesian camp." [3]

Interestingly, although three of the four participants stated that they think Bayesian statistics seems to agree more with what they actually want to know from statistics, they self-admittedly do not know much about Bayesian statistics. Importantly, then these are somewhat uninformed opinions of the participants and not a representation of facts about the

appropriateness of Bayesian statistics or the accordance of Bayesian statistics with the interviewees' expectations for statistics.

"number one I don't know Bayesian thoroughly. I haven't dug into it yet." [3]

"... what I know is null hypothesis significance testing ..." [2]

"I've always been fascinated by Bayesian statistics and I know very little about it, but it seems to be closer to the way I think." [4]

However, one participant at a slightly more neutral position on than the other three about whether the Bayesian approach is generally more appropriate than NHST.

"I think they both can provide unique information I don't think one is better than the other they are different and so employ them knowledgeably know the information that they are providing you and interpret it within their particular context. I don't think one particular approach is universally going to work out better." [1]

Subtheme: Priors. Two of the participants suggested that the ability of the Bayesian approach to include prior knowledge into the process makes it more appropriate than NHST. This ability was valued, because according to those participants, they usually already know something about a topic that they investigate prior to the study and the ability to connect and incorporate more information seems more realistic than the isolated calculations in NHST. *"I think the fact that it takes priors into account."* [2]

"I already know a lot at least somewhat about what I expect to see and it seems that Bayesian statistics takes that into consideration." [4]

Subtheme: probability of hypotheses. The other kind of knowledge that is wanted which Bayesian statistics provides and one participant mentioned is the probability of a

certain hypothesis. The participant mentioned this expectation for statistics in combination with the wish to be able to get insights about replicability from statistical methods.

"What's the likelihood that this particular hypothesis is true? That is definitely what I am trying to do." [4]

Theme: strength of the evidence

The next theme is the strength of the evidence. Three participants talked about how they want to have some indication about the strength of the evidence. This strength indication was suggested by comments about the accuracy of tests (positive predictive value and negative predictive value), talking about wanting statistics to enable them to have some kind of confidence in conclusions instead of a yes or no decision or knowledge about differences between group means as in NHST. They seem to want to know how much the data suggests something.

"I want to know how much!" [3]

"Well I want to know about, yes, about the strength of the data, but also secondly the shape of the data" [3]

"really at the end of the day giving us some confidence in something" [2]

"Much more important are positive predictive values, negative predictive values, the area under the curve. I mean that kind of stuff is a lot more informative than whether or not two groups have a different mean value" [1]

Theme: Individual and/or group

Another theme that was prompted by the questions and identified in the responses from the participant was whether they would like to gain knowledge and make conclusions about individuals or groups. This overarching theme can be divided into two subthemes. Those subthemes are that participants want to know about individuals and that it depends on the situation as well as the research question.

Subtheme: individual. Two of the participants proclaimed their desire to be able to make conclusions about and gain insights into individuals instead of groups.

"Really I want individuals." [4]

"if you just look at the average you're losing a lot and you need to know what does that data look like. You need to get a feel for that data. So I guess to your question my answer is the individual, but the difficulty is like, what is that phrase, every drop in the ocean is the ocean or something like that." [3]

Interestingly, at the same time one interviewee also outlined some difficulties with using statistical methods that actually target the individual and how that makes that participant use group methods as a potentially improper proxy for gaining knowledge about the individual. I said: *"yeah it's interesting how you kind of describe it that as you kind of said you use group methods out of convenience and actually what you're after seems to be more like the individual things."* To which the participant replied:

"yes very much." [4]

Subtheme: both, it depends on the situation and the research question. Similar to some of the responses to the more general question what they want to know from statistics, the more specific question about knowledge regarding the individuals or groups prompted the answer that they want both it just depends on the subject and the problem that is under investigation which they want in each situation.

"I mean it really kind of depends on the situation there." [1]

"I guess that depends on the problem, right? [2]

One participant although previously claiming to be after knowledge concerning the individual, responded *"yeah and that's why I said in the beginning I can't separate them"* [3] after directly being asked whether insights into both group and individuals would be more appropriate.

How does statistics affect the research question?

Theme: limited by demands of the statistical methods

Some participants describe how the requirements for certain statistical techniques restrict the type of questions they can investigate. For instance, a limitation that is mentioned is how the need for large sample sizes to reach a certain amount of statistical power will make certain studies unattainable. Because, the number of variables that would have to be included in combination with rare characteristics in the target population result in a required sample size that is practically impossible. Therefore, some topics and questions cannot or will not be investigated and the complexity of the actual phenomena has to be ignored or gets lost. In other words it is not just that statistics restricts what can be investigated, but it also leads to simplifications of the actual phenomena.

"... often it feels like I'm going to have to ignore a lot of that complexity, because if I really want to run some statistics with enough power to detect some differences or correlations or something then for each analysis I have to ignore a lot.... So, sometimes I'm feeling like there are a lot of questions that I'm just not going to be able to address, ..." [4]

"In that case the statistics are constraining it, because I don't want to end up with something completely messy that is not going to be that useful or for which I'm going to need way more participants than I can get in whatever amount of time."[2]

Theme: Statistics influences the types of questions that are considered and formulated

At some point or another all four participants talk about how the statistical methods they know and are used to influence the kind of research questions they consider and formulate. One participant states the wish that this would not be the case, but still admits that it is. And another asks whether this kind of influence might even limit the kind of questions asked. In other words there seems to be an awareness about the influence of the statistical techniques on their practices and research directions that go beyond mere data analysis.

"I think yes. The statistics that I know that's a really important thing right now." [2]

"I'd like to say no it doesn't that I let the questions always drive those things, but I'm sure at some level there is some implicit stuff going on there where the questions I'm interested in are informed by what I know how to do or what I'm familiar with doing." [1]

One participant when asked whether the known statistical methods influence the research questions that are being asked, just replied: *"absolutely!"*[3]

One interviewee even mentions this influence spontaneously: "It's interesting, because then I've noticed that a lot of my research questions in some ways are already being channeled towards particular statistical designs analyses and so I've often wondered whether or not that kind of limits the questions I ask given my particular understanding of statistics." [4]

The same interviewee also provided a reason why statistical considerations have such an influence on the research questions, by linking the influence of statistical considerations on the research question to the fact that any empirical research usually has to be planned and statistics has to fit the question to go through ethical review, before any study can be conducted.

"Right, yes it is and that's mostly driven by thinking about okay I need to get this project through ethics review right that's the first step. And they will want to know how am I going to analyze this data, do I have a plan for it so that I am not wasting the time and energy of participants. So, I'm very conscious about will I need to have a plan before I even start." [4]

Theme: The research question has primacy

When asked whether the consideration of statistical methods leads to changes in the research questions three participants indicated that the research question has or should have primacy over statistical considerations. In these cases it is described in a way that statistics serves or at least should serve the research questions.

"I'd say the question is of primacy and then the method should fit the question." [1]

"I mean the statistics might change what I'm doing a little, but it's not the driving force." [2]

One participant was asked whether there is an interplay between statistical considerations and the research question that leads to alterations in the question. Or if the research question comes first and then you look for the statistics. The participant replied:

"Well, definitely the latter, because I'm a very pragmatic person." $[3]^2$

Theme: Statistics changes the question

Another theme that emerged is that statistical considerations seem to change the research questions that are investigated. Statistical considerations are either already in the mind of (some of) the investigators when the question is being formulated or after the initial formulation of the question the statistical considerations lead to reformulations and alterations in the question when thinking about which approaches could answer or approximate the question. This theme is different from the *Statistics influences the types of questions that are considered and formulated* theme, because here the statistical considerations usually lead to

² That participant also spontaneously mentioned how little time there is to work on theory: "*I have a theory* paper *I worked on for three years. I only work on it when I get time maybe a week every year generally between Christmas and the New Year.*"

changes in already considered or existing questions, while the previous theme is about the circumstance that statistics seems to have such an influence that it even affects which questions are considered and formulated in the first place. In other words this theme is more about how they are formulated and the previous one is more about whether they are formulated at all.

"And so I'm already thinking about in general what kind of statistical approach am I going to need even as I'm trying to put the research question into words." [4]

One participant who also acts as a methodological advisor said: "And so it really kind of depends on who is coming up with the question, me or someone else and then can that question be rephrased or re-conceptualized in a way that it can be answerable with known methods." [1]

Theme: Too narrow set of statistical methods

Another theme that came up that is somewhat linked to the second research question and the influence of statistics on psychology is the opinion from some interviewees that the set of statistical techniques that are routinely used in psychological science are too narrow for the discipline. When directly asked whether these statements by statisticians about what psychologists want to know from statistics might be motivated by a too narrow understanding of statistics one participant replied: *"yes I would right, because it really depends on what you want to know and what you're doing."*[2]

When I asked another participant whether the problem might not be that statistics does not provide psychological scientists with what they are looking for, but that psychological scientists need to pay more attention to different kinds of statistical techniques the interviewee replied: *"yes. You've got a nail use a hammer, you've got a screw use a screwdriver you* know? ... Yes so that's where I get irritated when I see manuscripts and they just blindly apply these ... "[3]

Responses such as the previous ones and one from another participant also somewhat support Gigerenzer's (2004, 2018) claims that statistics and especially a version of NHST is automatically and ritualistically applied in psychological science³: "So hundred percent psychologists, albeit they need to have this big extensive tool chest, but they need to know that the tool chest exists and be able to know that there is more than just one particular way to go about studying things. And so if their question of interest looks at this then let's find some method that actually gets at that as opposed to just using what you're comfortable using. What I think has been the point and center, this is what psychology has been doing. This is what I'm trained to do so that is what I do." [1]

Limitations of the study

The study has some important limitations. The sample is by no means representative of psychological scientists in general, not just because of the uniqueness of the sample concerning their positions in the discipline and the small sample size, but also because of the extremely low response rate (3%). This could suggest that the researchers who agreed to participate are more interested in statistical issues in psychological science than the general population of psychological scientists or they might even see problems in the current statistical methods in psychological science, because they for themselves identified a mismatch between what statistical methods are able to provide them and what they actually want to know. However, the sample was never intended to be representative and for a broader as well as deeper exploration into the topic a continuation of this study with further participants is in planning. Furthermore, this was only my second interview study.

³ Also see for example Porter (1995) and Davidson (2018) for their conception of mechanical objectivity.

Consequently, the quality of the study could have been lowered by my limited experience with qualitative research.⁴ Additionally, my own opinions regarding the appropriateness of certain statistical methods, such as NHST and Bayesian statistics could have influenced the interview process and the responses of the participants. Generally I tend to think that statistics are like a toolbox. Hence, depending on the topic, the research question and the context, different methods are appropriate and usually more than one are available and justifiably usable. However, for hypotheses testing I tend to think that the Bayesian approach is more appropriate than NHST if your goal is to know something about hypotheses (hence, hypotheses testing). Moreover, the power or authority (academic position) difference between me, as a master student, and the interviewees who are APA division (society) presidents and journal editors, might have influenced the proceeding of the interviews. For example, it could be that interviewees did take questions that seem abstract or foreign to them not as serious as they might have when the question would have come from a more experienced and tenured interviewer. However, at the same time, it might also be that interviewees were more patient and explained in more detail what might have stayed implicit in a conversation with someone of comparable background and academic position. Additionally, in the context of the replication crisis or crisis of confidence (Earp & Trafimow, 2015), which has also been labeled a statistical crisis (Gelman & Loken 2014) which indicates the important role statistics plays in the current crisis discussions, participants could also have answered overly careful. Lastly, the abstractness of the topic and the questions could have affected the quality of the interview. This was the reason why I chose to conduct semi-structured interviews, I thought it might enable me to explain questions and react to potential misunderstandings. However, misunderstandings and the abstractness of the topic and the questions could still have led to confusion about how to answer the questions and to a limited richness in the answers.

⁴ I also had COVID during three of the interviews which could have affected the quality of the interviews.

Discussion

The answers in the interviews indicate that the statistical methods at the researchers' disposal and the methods that they are used to, influence the type of questions they formulate and investigate. Hence, the statistical techniques that psychological scientists know and are used to are apparently among the factors, such as political orientation (see e.g. Honeycutt & Jussim, 2020), that potentially influence the kind of research questions that are considered and investigated in psychological science. It seems as if Lakens' (2017) bathroom tile does not even go far enough, because it seems that it is not just the case that statisticians and statistics guide how to ask questions in psychological science, but statistics even influence which kind of questions are considered and formulated in the first place. Importantly, it is not just the statistical methods that are available in psychological science, but maybe even more importantly the methods that psychological scientists are used to that influence the questions that are being asked. Therefore, a too narrow focus in the discipline on one or two methods (e.g. NHST and the focus on the p-value)⁵ could restrict the scope and possibilities of the discipline. Such a discipline with strong influence of statistics on the research practices and the direction of the discipline is reminiscent of Danziger's (1985) conception of a methodological imperative, according to which inferential statistics binds and guides the discipline. Put differently, according to the answers of the interviewees, it could be that the statistical methods psychological scientists are used to, influence where the discipline is going, instead of the actual interests of the scientists or the characteristics of the phenomena that they are investigating (or at least it does so to non-negligible degree). And when the discipline is only used to a narrow set of statistical techniques, then the investigation and maybe even theoretical orientation of the discipline will be restricted to what the statistical methods allow. In other words, a narrow focus on a small set of statistical techniques or even

⁵ The problem of a too narrow focus would also exist if the discipline just exchanges frequentist with Bayesian statistics.

only one method (NHST) alone could also contribute to the apparent problem of a certain kind of conformism or even stagnation in psychological research. Fiedler and colleagues (2012), for instance, argue that a neglect of false negatives and the focus on strict statistical criteria (e.g., p-value < 0.05) might lead to the outright rejection of unconventional thinking and theories in the discipline that could explore potentially fruitful directions for the field. Likewise, when the statistical methods that the research community is used to have such an immense influence on the kind of questions asked and only a narrow set of statistical techniques is widely used and known, then the discipline will also only navigate through narrow and conventional pathways. Put differently, Fiedler et al's (2012) critique about a general lack of theoretical risk taking in the discipline could then be extended to a critique of a general lack in risk taking regarding the kind of questions that are investigated in psychological science, because a narrow set of statistical methods restricts the field. It might be that due to the narrow set of statistical techniques that dominate the field it could seem or even is easier for researchers to change their research question than to find an appropriate statistical method for the original or intended research question. Tellingly, interviewees who talked about that the statistical methods they know and are used to influence the kind of questions they formulate and investigate, also voiced regrets about that. They seem to wish that it would not be the case and that the research questions should dictate the statistical methods that are used to answer the questions, instead of the other way around.

Hence, psychology has to do a better job at teaching, training and demanding a wider variety of statistical methods which are used more in accordance with the research questions, than what an individual scientist or the discipline is used to at the moment. Most participants in this study also seem to agree with such a sentiment.

The answers from three of the four interviewees also provide some support for Haucke and colleagues' (2021) findings regarding psychological scientists wanting to know what Bayesian statistics provides. The things they seem to want are, for instance, the incorporation of priors and the probability of hypotheses. However, such (Bayesian) responses from the interviewees can also be the consequence of a certain lack of a broader knowledge concerning different statistical methods that can provide different kinds of answers. Furthermore, it is important to mention that in the literature Bayesian statistics is often presented (advertised) as an alternative to NHST. In other words it might be the case that when researchers have the realization that NHST does not provide what they want to know they see Bayesian statistics as the alternative, because they do not know other alternatives. Therefore, a more humble interpretation of the responses might be required, namely that between the two they seem to prefer Bayesian statistics. Moreover, Haucke and colleagues (2021) reported a discrepancy between what psychological scientists think the p-value tells them and what they want to know from the p-value. Interestingly, some interviewees in this study demonstrated another discrepancy. There seems to be a divide between which statistical methods the interviewees think would best fit their wishes regarding what they want to know from statistics (namely, Bayesian methods) and the methods they actually use in their day to day practice as researchers (namely, frequentist methods and group statistics). There are numerous potential reasons for this divide, among them are that there might be a lack of widely available statistical training and a general lack in initiative among psychological scientists to go on their own and learn a new method. It might also be the case that because of publication pressures and funding applications psychological scientists just do not have the time that would be needed to learn new or more methods. All of those potentially prevent researchers in psychology from learning more and different statistical techniques that could be more appropriate for the actual research questions they are interested in investigating and answering. One interviewee even hints at this lack of time when talking about the intent to learn more about Bayesian statistics: "I'm on sabbatical this year. I hope to learn more." [2]. So, apparently researchers not just have the feeling or impression that there is little to no time

for proper theory development available in the discipline (see footnote 2), but also that there is a lack of time that prevents scholars from learning new statistical methods that one is interested in or that might be needed to answer a particular research question.

The responses from the interviewees also give some support to Lakens' (2021) claims regarding the *statisticians' fallacy*, because the participants said that usually what they want to know from statistics depends on the topic, the research question, the problem and the context of the investigation. Such answers also suggest that statements from statisticians about what psychological scientists want to know from statistics have an under-appreciation for the complexity and diversity that is inherent to the discipline. This complexity and diversity is not just prevalent in the phenomenon of investigation in psychology, but also due to the varied aims and interests of the researchers in the discipline. The results of this study indicate that further research into what psychological scientists want to know from statistical techniques is urgently needed. Because, what kind of strange world would we live in where we let methods that don't even tell us what we want to know influence the kind of questions that are ultimately investigated and even which research questions are considered in the first place?

Conclusion

Statistics has an undeniable influence on psychology. This study hinted towards one potential source of that influence that might usually be neglected. Namely, how statistical considerations affect the form of research questions and even what kind of research questions are actually investigated in psychological science. Importantly, generally participants who indicated that statistics influences the kind of questions they ask, also showed regret about this and wished it would be the other way around that the research questions dictate the statistical methods. Moreover, the current study at least to a certain degree reinforces the notion that psychological scientists think NHST does not provide them with the kind of answers they are looking for and that Bayesian statistics might be more appropriate (at least when it comes to

hypotheses testing). Furthermore, the interviewees also indicated that generally what psychological scientists want to know from statistics is dependent on the research question, the subject or problem of investigation and the context of the research. In other words, there usually is not just one thing they want statistics to tell them. For instance, the things the participants in the current study wanted from statistics were among others: aid in decision making, information about the shape of the data, indications concerning the strength of the evidence, predictions about individuals and information about the replicability of a finding. In closing, for the sake of the field, what psychological scientists want to know from statistics and the influence of statistics on the discipline are in dire need for further investigation and reflection. Ultimately, a wider awareness about those issues is needed and statistical training as well as education have a key role to play in spreading such an awareness. However, in the end it is up to each individual researcher to take the initiative and become aware of such issues and to broaden their horizon beyond what is usual in their discipline or seems straightforward to them.

References

Adolf, J. K., & Fried, E. I. (2019). Ergodicity is sufficient but not necessary for group-to-individual generalizability. *Proceedings of the National Academy of Sciences of the United States of America*, 116(14), 6540–6541.

https://doi.org/10.1073/pnas.1818675116

- Badenes-Ribera, L., Frias-Navarro, D., Iotti, B., Bonilla-Campos, A., & Longobardi, C.
 (2016). Misconceptions of the p-value among chilean and italian academic psychologists. *Frontiers in Psychology*, 7, 1247–1247.
 https://doi.org/10.3389/fpsyg.2016.01247
- Blume, J. D. (2011). Likelihood and its evidential framework. In Bandyopadhyay, P. S.,
 Forster, M. R. (Eds.), Philosophy of statistics (pp. 493–511). Amsterdam, The
 Netherlands: North Holland. doi:<u>10.1016/B978-0-444-51862-0.50014-9</u>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <u>https://doi.org/10.1191/1478088706qp0630a</u>
- Cassidy, S. A., Dimova, R., Giguère, B., Spence, J. R., & Stanley, D. J. (2019). Failing Grade:
 89% of Introduction-to-Psychology Textbooks That Define or Explain Statistical
 Significance Do So Incorrectly. *Advances in Methods and Practices in Psychological Science*, 233–239. <u>https://doi.org/10.1177/2515245919858072</u>
- Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, 49(12), 997–1003. https://doi.org/10.1037//0003-066X.49.12.997
- Colquhoun, D. (2017). The reproducibility of research and the misinterpretation of pvalues. *Royal Society Open Science*, 4(12), 171085–171085.
 <u>https://doi.org/10.1098/rsos.171085</u>

- Danziger, K. (1985). The methodological imperative in psychology. *Philosophy of the Social Sciences*, *15*(1), 1–13. <u>https://doi.org/10.1177/004839318501500101</u>
- Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge University Press. <u>https://doi.org/10.1017/CBO9780511524059</u>
- Davidson, I. J. (2018). The Ouroboros of Psychological Methodology: The Case of Effect
 Sizes (Mechanical Objectivity vs. Expertise). *Review of General Psychology*, 22(4),
 469–476. https://doi.org/10.1037/gpr0000154
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*, 621–621.
 https://doi.org/10.3389/fpsyg.2015.00621
- Falkenström, F. (2022). Data analytic advances that bridge the scientist-practitioner gap. *Cognitive and Behavioral Practice*, 29(3), 585–590. <u>https://doi.org/10.1016/j.cbpra.2022.02.007</u>
- Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The Long Way From α-Error Control to Validity Proper: Problems With a Short-Sighted False-Positive Debate. *Perspectives* on psychological science : a journal of the Association for Psychological Science, 7(6), 661–669. <u>https://doi.org/10.1177/1745691612462587</u>
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, 115(27), 6106.
- Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, 102(6), 460–465.

- Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences:
 Methodological issues (p. 311–339). Lawrence Erlbaum Associates, Inc.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033
- Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. Advances in Methods and Practices in Psychological Science, 198– 218. <u>https://doi.org/10.1177/2515245918771329</u>
- Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers. *Methods of Psychological Research*, 7(1), 1–20.
- Haucke, M., Miosga, J., Hoekstra, R., & van Ravenzwaaij, D. (2021). Bayesian frequentists:
 examining the paradox between what researchers can conclude versus what they want to conclude from statistical results. *Collabra: Psychology*, 7(1).
 https://doi.org/10.1525/collabra.19026
- Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic bulletin & review*, 21(5), 1157–1164. https://doi.org/10.3758/s13423-013-0572-3
- Honeycutt, N., & Jussim, L. (2020). A model of political bias in social science research. *Psychological Inquiry*, *31*(1), 73–

85. https://doi.org/10.1080/1047840X.2020.1722600

Lakens, D. (2017, November 11). *The 20% Statistician: The Statisticians' Fallacy*. http://daniellakens.blogspot.com/2017/11/the-statisticians-fallacy.html

- Lakens, D. (2021). The practical alternative to the p value is the correctly used p
 value. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, 16(3), 639–648. https://doi.org/10.1177/1745691620958012
- Lyu, X.-K., Xu, Y., Zhao, X.-F., Zuo, X.-N., & Hu, C.-P. (2020). Beyond psychology:
 Prevalence of p value and confidence interval misinterpretation across different fields.
 Journal of Pacific Rim Psychology, 14. <u>https://doi.org/10.1017/prp.2019.28</u>
- Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research*, 2, 1-18.
- Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2), 112–117.
- Nickerson R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*, 5(2), 241–301. <u>https://doi.org/10.1037/1082-989x.5.2.241</u>
- Porter, T. M. (1995). *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life.* Princeton, New Jersey: Princeton University Press.
- Speelman, C. P., & McGann, M. (2020). Statements About the Pervasiveness of Behavior Require Data About the Pervasiveness of Behavior. *Frontiers in Psychology*, 11, 3117. <u>https://doi.org/10.3389/fpsyg.2020.594675</u>
- Spence, J. R., & Stanley, D. J. (2018). Concise, Simple, and Not Wrong: In Search of a Short-Hand Interpretation of Statistical Significance. *Frontiers in psychology*, 9, 2185. https://doi.org/10.3389/fpsyg.2018.02185

- Szucs, D., & Ioannidis, J. (2017). When Null Hypothesis Significance Testing Is Unsuitable for Research: A Reassessment. *Frontiers in human neuroscience*, 11, 390. <u>https://doi.org/10.3389/fnhum.2017.00390</u>
- Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: building a bridge between two research paradigms. *Multivariate Behavioral Research*, 49(3), 193–213. <u>https://doi.org/10.1080/00273171.2014.889593</u>
- Wijsen, L. D., & Borsboom, D. (2021). Perspectives on Psychometrics Interviews with 20
 Past Psychometric Society Presidents. *Psychometrika*, 86(1), 327–343.
 https://doi.org/10.1007/s11336-021-09752-7