The research report: why, what and how

Rob Davies

02/11/2022

Table of contents

Preface

A change in approach

We can, here, explain a development in the approach we take in teaching this course. Naturally, this development in approach will require a parallel development in your approach to learning.

We are going to focus on working in research in context (see Figure Figure 0.1).

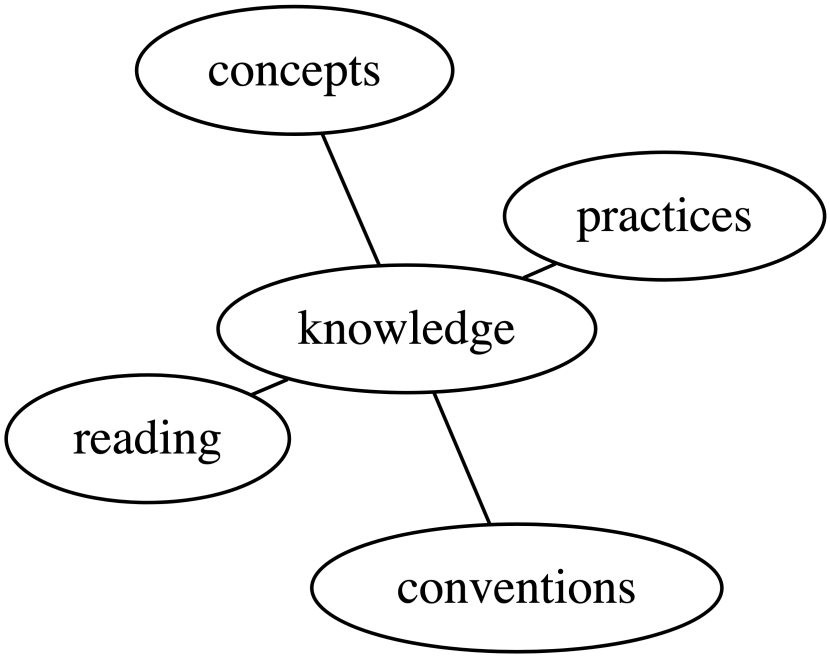

Figure 0.1: This is a simple graphviz graph.

You have been introduced to R. We know that some of you are new to R so we will practice the skills you are learning. We will consolidate, revise, and extend these skills.

We will encounter — some, for the first time – the linear model *also known as* regression analysis, multiple regression.

But the big change is this focus on the context. The reason is that *not* talking about the context has a dangerous impact on how you approach, do or think about data analysis.

In traditional methods teaching, the schedule of classes will progress through a series of tests, one test a week, from simpler to more complex tests (e.g., from t-test to multiple regression at the undergraduate level). Textbooks often mirror this structure, presenting one test per chapter. In this approach, the presentation is often brief about the context: the question the researchers are investigating; the methods they use to collect data, including the measurements; and the assumptions they make about how your reasoning can get you from the things you measure to the things you are trying to understand. In this approach, also, example data may be presented in a limited, partial, way.

The reasons for this are understandable: methods are complex, technical, subjects for learning, and teachers and students do not also have time, perhaps, to think about statistics and about theoretical or measurement assumptions. This is a mistake because it presents a misleading view of the challenge in learning methods: the challenge is *just* the (difficult enough) challenge of learning about statistical methods, or dealing with numbers. It is a mistake, also, because it implies that if you learn the method, and can match the textbook example – the variables, the state of the data – when it is your turn to do an analysis, all will be well.

Maybe. I think a more productive approach – this is the approach we will take – is to expose, and talk about some of the real challenges that anybody who handles data, or quantitative evidence, in professional life. These challenges include:

1. Thinking about the mapping from our concerns to the research questions, to the things we measure, to analysis we do, and then the conclusions we make.
2. Selecting or constructing valid measures that can be assumed to measure the things they are supposed to measure.
3. Taking samples of observations, and making conclusions about the population.
4. Making estimates and linking these estimates to an account that is explicit about causes.

1 Introduction: the why

The research report assignment requires students to locate, access, analyse and report previously collected data. This introduction is intended to answer the first question anybody might ask.

- Why: what is the motivation for the assignment?

In following materials, I will answer the questions.

- How can the assignment be done?
- What do we expect students to do?

It is going to appear, at first, that I am going a *long* way away from telling you what you need to do for the assignment. I hope you will agree that the discussion that follows is worth your time in reading it. It will help you to understand *why* we are asking you to do the assignment, and *why* we are looking for what we are looking for. It will help you to understand *how* this work will aid your development. And it will help to show *how* doing the assignment furnishes the opportunity for research experience that will help you later in your working life.

For those who are more eager to start the work, here are the links to the what information in in Chapter 2 and to the how information in Chapter 3.

1.1 The key ideas

There are two ideas motivating our approach. It will be helpful to you if I sketch them out early, here. We can demonstrate the usefulness of these ideas as we progress through our work.

The first key idea is expressed clearly in sociological discussions of science. This is that there is a difference between science "…being done, science in the making, and science already done, a finished product …" [Bourdieu (2004); p.2]. The awareness we want to develop is that there are two things: there is the story that may be presented in a textbook or in a lecture about scientific work or scientific claims; and there is the work we do in practice, as we develop graduate skills, and as we exercise those skills professionally in the workplace.

The second key idea connects to the first. This idea is that reported analyses are not *necessary* or *sufficient* to the data or the question. What does this mean? It means that the same data

can reasonably be analysed in different ways. There is no *necessary* way to analyse some data though there may be conventions or normal practices (Kuhn, 1970). It means that it is unlikely that any one analysis will do all the work that could be done (a sufficiency) to get you from your data to useful or reasonable answers to your questions.

These ideas may be unsettling but they are realistic. Stating them will better prepare you for professional work. In the workplace, the accuracy of these ideas will emerge when you see how a team in any sector (health, marketing …) gets from its data to its product. If we talk about the ideas now, we can get you ready for dealing with the practical and the ethical concerns you will confront when that happens.

We will begin by discussing psychological research, and research *about* psychological research, to answer the question: **Why: what is the motivation for the assignment?** We will then move to answering the **what** and the **how** questions.

1.2 Why: what is the motivation for the assignment?

1.2.1 The wider context: crisis and revolution

We are here because we are interested in humans and human behaviour, and because we are interested in scientific methods of making sense of these things. Some of us are aware that science (including psychological science) has undergone a rolling series of crises: the replicability or replication crisis (Pashler & Harris, 2012; Pashler & Wagenmakers, 2012); the statistical crisis (A. Gelman & Loken, 2014b); and the generalizability crisis (Yarkoni, 2022). And that science is undergoing a response to these crises, evidenced in the advocacy of pre-registration (Nosek et al., 2018, 2019), and of registered reports (Nosek & Lakens, 2014), the use of open science badges (e.g., for the journal *Psychological Science*), the completion of large-scale replication studies (Aarts et al., 2015), and the identification of open science principles (Munafò et al., 2017). We may usefully refer, collectively, to the crises and the responses, as the *credibility revolution* (Vazire, 2018)

We could teach a course on this (in Lancaster, we do) but I must be brief, here, and invite you to follow the references, if you are interested. Before going on, I want to call your attention to the fact that important elements of the hard work in trying to make science work better has been led by PhD students and by junior researchers (e.g., Herndon et al., 2014). Graduate students may, at first, assume that the fact that a research article has been published in a journal means the findings that are reported must be *true*. Most of the time, some educated skepticism is more appropriate. An important driver of the realization that there are problems evident in the literature, and that there are changes we can make to improve practice, comes from independent **post-publication review work** exposing the problems in published work (see, e.g., this account by Andrew Gelman)

> Tip
>
> - Allow yourself to feel skeptical about the reports you read *then* work with the motivation this feeling provides.

In brief, then, most practicing scientists now understand *or should* understand that many of the claims we encounter in the published scientific literature are unlikely to be supported by the evidence (Ioannidis, 2005), whether we are looking at the evidence of the results in the reports themselves, or evidence in later attempts to find the same results (e.g., Aarts et al., 2015). We suspect that this may result from a number of causes. We understand that researchers may engage in questionable research practices (John et al., 2012). We understand that researchers may exploit the potential for flexibility in doing and reporting analyses (Simmons et al., 2011a). We understand that there are problems in how psychologists use or talk about the measurement of psychological constructs (Flake & Fried, 2020). We understand that there are problems in how psychologists sample people for their studies, both in where we recruit (Bornstein et al., 2013; Henrich et al., 2010; Wild et al., 2022), and in how many we recruit (Button et al., 2013; Cohen, 1962; Sedlmeier & Gigerenzer, 1989; Vankov et al., 2014). We understand that there are problems in how psychologists specify or think about their hypotheses or predictions (Meehl, 1967; Scheel, 2022). And we understand that there are problems in how scientists do, or rather do not, comply with good practice recommendations designed to fix these problems (discussed further in the following).

This discussion could (again) be unsettling. This list of problems could make you angry or sad. I, like others, think it is exciting. It is exciting because these problems have probably existed for a long time (e.g., Cohen, 1962; Meehl, 1967) and now, having identified the problems, we can hope to do something about it. It is exciting because if you care about people, the study of people, or the applications in clinical, education and other domains of the results of the study of people, then you might hope to see better, more useful, science in the future (Vazire, 2018).

As someone who teaches graduate and undergraduate students, I want to help you to *be the change you want to see in the world* [1]. We cannot solve every problem but we can try to do better those things that are within our reach. I am going to end this introduction with a brief discussion of some ideas we can use to guide our better practices.

1.2.2 The specific context: what we need to look at, conceptually and practically

In this course, for this assignment, we are going to focus on:

1. multiverse analyses
2. kinds of reproducibility

[1] This encouragement is often attributed to Gandhi but is attributed ((here)) to a Brooklyn school teacher, Ms Arleen Lorrance, who led a transformative school project in the 1970s.

3. the current state of the match between open science ideas and practices

In the classes on the linear model, we will discuss:

4. the links between theory, prediction and analysis
5. psychological measurement
6. samples
7. variation in results

1.2.3 Multiverse analyses: multi- what?

1.2.3.1 A first useful metaphor: the pipeline

I am going to link this discussion to a metaphor (see Figure Figure 1.1) or a description you will find useful: **the data analysis pipeline** or **workflow**.

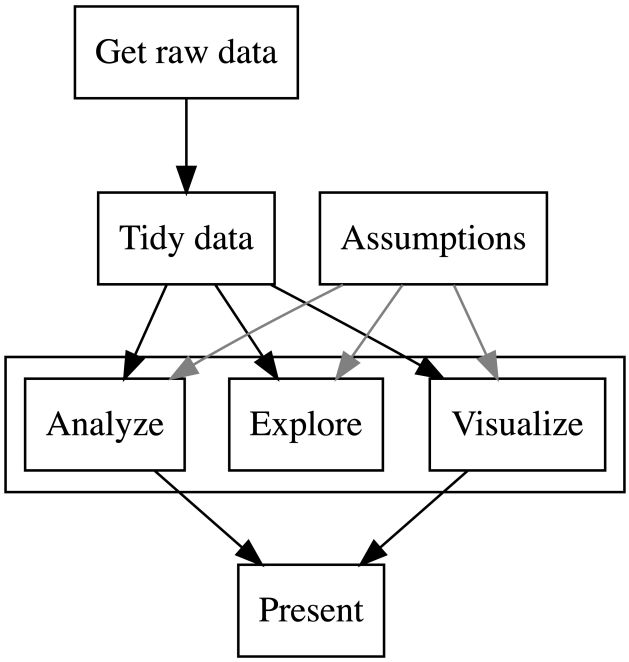

Figure 1.1: The data analysis pipeline or workflow

This metaphor or way of thinking is very common (take a look at the diagram in Wickham and Grolemund's 2017 book "R for Data Science) and you may see the words "data pipeline" used in job descriptions, or you may benefit from saying, in a job application, something like: *I am skilled in designing and implementing each stage of the quantitative data analysis pipeline,*

from data tidying to results presentation. I say this because scientists I have mentored got their jobs because they can do these things – and successfully explained that they can do these things – in sectors like educational testing, behavioural analysis, or public policy research.

The reason this metaphor is useful is that it helps us to organize our thinking, and to manage what we do when we do data analysis, we:

- get some data;
- process or tidy the data;
- explore, visualize, and analyze the data;
- present or report our findings.

We introduce the idea that your analysis work will flow through the stages of a *pipeline* from getting the data to presenting your findings because, next, we will examine how pipelines can *multiply.*

> Tip
>
> - As you practice your data analysis work, try to identify the elements and the order of your work, as the parts of a *workflow.*

1.2.3.2 A second useful metaphor: the garden of forking paths

What researchers have come to realize: because we started looking … The open secret that has been well kept (Bourdieu, 2004): because everybody who does science knows about it, yet we may not teach it; and because we do not write textbooks revealing it … Is that at each stage in the analysis workflow, we can and do make choices where multiple alternative choices are possible. A. Gelman & Loken (2014a) capture this insight as the "garden of forking paths"[2] (see Figure 1.2).

The general idea is that it is possible to have **multiple potential different paths** from the data to the results. The results will vary, depending on the path we take. In an analysis, we could take multiple different paths simply because at point A we decide to do B1, B2 or B3, maybe we choose B1, and then at point B1, we may decide to do C1, C2 or C3. Here, maybe we have our raw data at point A. Maybe we could do one of two different things when we tidy the data: action B1 or B2. Then, when we have our tidy data, maybe we can choose to do our analysis in one of six ways. Where we are at each step *depends* on the choices we made at the previous steps.

In the end, it may appear to us that we took one path or that only one path was possible. When we report our analysis, in a dissertation or in a published journal article, we may report the analysis **as if only one analysis path had been considered**. But, critically, our

[2]The term is taken from the name of a short story by Jorge Luis Borges, "El jardin de senderos que se bifurcan".

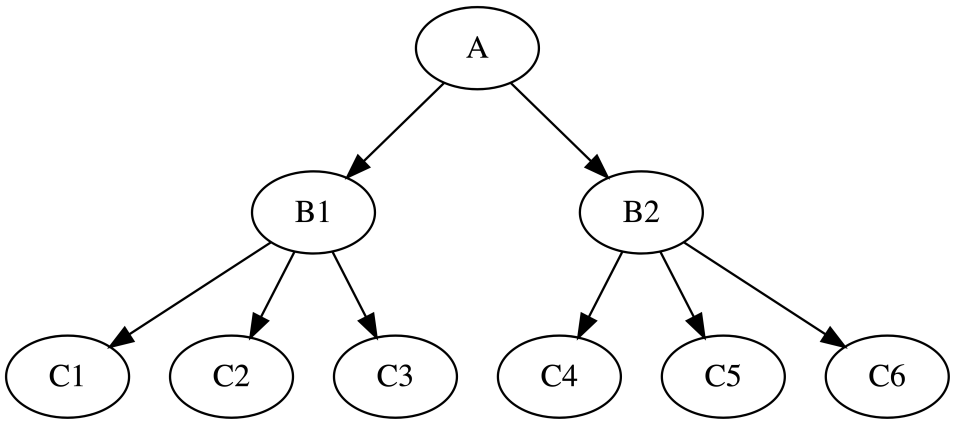

Figure 1.2: Forking paths in data analysis

findings may depend on the choices we made and this variation in results may be hidden from view.

I am talking about forking paths because the *multiplicity* of paths has consequences, and we discuss these next.

> Tip
>
> - It is about here, I hope, that you can start to see why it would makes sense to access data from a published study and to examine if you can get the same results as the study authors.

1.2.4 Multiverse analyses

I am going to discuss, now, what are commonly called *multiverse analyses*. Psychologists use this term, having been introduced to it in an influential paper by Steegen et al. (2016a), but it comes from theoretical physics (take a look at wikipedia).

I explain this because I do not want you to worry. The ideas themselves are within your grasp whatever your background in psychology or elsewhere. It is the implications for our data analysis practices that are *challenging*. They are challenging because what we discuss should increase your skepticism about the results you encounter in published papers. And they are challenging because they *reveal your freedom* to question whether published authors could have done their analysis in a different way.

We are going to look at:

1. dataset construction
2. analysis choices

1.2.4.1 The link between the credibility revolution and the multiverse

In first discussing the wider context (of crisis and revolution), then discussing the specific context (of multiverses and, in the following, of reproducibility), I should be clear about **the link between the two things**. The finding that some results may not be supported by the evidence is probably due to a mix of causes. But one of those causes will be the combination of uncertainty over data processing or the uncertainty over analysis methods revealed in multiverse analyses, as we see next, combined with the limitations of data and code sharing, and the incompleteness of results reporting (as we see later).

1.2.4.2 The data multiverse

When you collect or access data for a research study, the complete raw dataset you receive is almost never the complete dataset you analyze or whose analysis you report. This is not a story about deliberately cheating. It is a story about the normal practice of science (Kuhn, 1970).

Picture some common scenarios. You did a survey, you got responses from a 100 participants on 10 questions, and you asked people to report their education, ethnicity and gender. You did an experimental study, you tested two groups of 50 people each in 100 trials (imagine a common task like the Stroop test), and you observed the accuracy and the timing of their responses. You tested 100 children, 20 children in each of five different schools, on a range of educational ability measures.

In these scenarios, the psychologist or the analyst of behavioural data *must* process their data. In doing so, you will ask yourself a series of questions like:

- how do we code for `gender, ethnicity, education`?
- what do we about reaction times that are very short, e.g., $RT < 200ms$ or very long, e.g., $RT > 1500ms$)?
- if we present multiple questions measuring broadly the same thing (e.g. how confident are you that you understand what you have read? how easy did you find what you read?) how do we summarize the scores on those questions? do we combine scores?
- what do we do about people who may not appear to have understood the task instructions?

Typically, the answers to these questions will be given to you by your supervisor, a colleague or a textbook example. For example, we might say:

- "We excluded all reaction times greater than 1500ms before analysis."

Typically, the explanation for these answers are rarely explained. We might say:

- "*Consistent with common practice in this field*, we excluded all reaction times greater than 1500ms before analysis."

But the reader of a journal article typically **will not see** an explanation for why, as in the example, we exclude reaction times greater than 1500ms and not 2000ms or 3000ms, etc. We typically do not see an explanation for why *we* exclude all reaction times greater than 1500ms but *other* researchers exclude all reaction times greater than 2000ms. (I do not pick this example at random: there are serious concerns about the impact on analyses of exclusions like this (Ulrich & Miller, 1994).)

What Steegen et al. (2016a) showed is that a dataset can be processed for analysis in multiple different ways, with a number of reasonable alternate choices that can be applied, for each choice point: construction choices about classifying people or about excluding participants given their responses. If a different dataset is constructed for each combination of alternatives then many different datasets can be produced, all starting from the same raw data. (For their example study, Steegen et al. (2016a) found they could construct 120 or 210 different datasets, based on the choice combinations.) Critically, for us, Steegen et al. (2016a) showed that if we apply the same analysis method to the different datasets then our results will vary.

Let me spell this out, bit by bit:

- we approach our study with the same research question, and the same verbal prediction;
- we begin with the exact same data;
- we then construct different datasets depending on different *but equally reasonable* processing choices;
- we then apply the same analysis analysis, to test the same prediction, using each different dataset;
- we will see different results for the analyses of the different datasets.

Alternate constructions of the same data may cause variation in the results of statistical tests. Some kinds of data processing choices may be more influential on results than others. It seems unlikely that we can identify, in advance, which choices matter more.

Steegen et al. (2016a) suggest that we can *deflate* (shrink) the multiverse in different ways. I want to state their suggestions, here, because we will come back to these ideas in the classes on the linear model.

1. Develop better theories and improved measurement of the constructs of interest.
2. Develop more complete and precise theory for why some processing options are better than others.

But you will be asking yourself: **what do I need to think about, for the assignment?**

Tip

- When you read a psychological research report, identify where the researchers talk about how they process their data: classification, coding, exclusion, transformation, etc.
- If you can access the raw data, ask yourself: could different choices change the results of the same analysis?

1.2.4.3 Analysis multiverses

Even if we begin with the same research question and, critically, the *same dataset*, the results of a series of studies show that different researchers will often (reasonably) make *different choices about the analysis* they do to answer the research question. We often call these studies (analysis or model) **multiverse** studies. In these studies, we see variation in analysis and this variation is also associated with variation in results.

An influential example, in psychology, is reported by Silberzahn and colleagues (Silberzahn et al., 2017; Silberzahn & Uhlmann, 2015) who asked 29 teams of researchers to answer the same question ("Are (soccer) referees more likely to give red cards to players with dark skin than to players with light skin?") with the same dataset (data about referee decisions in football league games). The teams made their own decisions about how to answer the question in doing the analysis. The teams shared their plans, and commented on each others' ideas. The discussion did not lead to a consensus about what analysis approach is best. In the end, the different teams did different analyses and, critically, the **different analyses had different results**. The results varied in whether the test of the effect of players skin colour (on whether red cards were given) was significant or not, and on the strength of the estimated association between the darkness of skin colour (lighter to darker) and the chances (low to high) of getting a red card.

There have now been a series of multiverse or multi-analyst studies which demonstrate that, under certain conditions, different researchers may adopt different analysis approaches – which will have *different results* – in answering the same research question with the same data. This demonstration has been repeated in studies in health, medicine, psychology, neuoscience, and sociology, among other research fields (e.g., Parsons (n.d.); Breznau et al. (2022); Klau et al. (n.d.); Klau et al. (2021); Wessel et al. (2020); Poline et al. (2006); Maier-Hein et al. (2017); Starns et al. (2019); Fillard et al. (2011); Dutilh et al. (2019); Salganik et al. (2020); Bastiaansen et al. (2020); Botvinik-Nezer et al. (2020); Schweinsberg et al. (2021); Patel et al. (2015); see, for reviews, and some helpful guidance, Aczel et al. (2021); Del Giudice & Gangestad (2021); Hoffmann et al. (n.d.); Wagenmakers et al. (2022)).

In these studies, we typically see variation in how psychological constructs are operationalized (e.g., how do we measure or code for social status?), how data are processed or datasets constructed (as in Steegen et al. (2016b)), plus variation in *what* statistical techniques are

used, and in *how* those techniques are used. This variation can be understood to reflect kinds of **uncertainty** (Klau et al., n.d.; Klau et al., 2021): uncertainty about how to process data, and uncertainty about the model or methods we should use to test or estimate effects. Further research makes it clear that we should be aware, if we are not already, of the variation in results that can be expected because different researchers may choose to design studies, and construct stimulus materials, in different ways given the same research hypothesis information (Landy et al., 2020).

But you will be asking yourself: **what do I need to think about, for the assignment?**

Tip

- When you read a psychological research report, identify where the researchers talk about how they analyse their data: the hypothesis or prediction they test; the method; their assumptions; the variables they include; the checks or the alternate analyses they did or did not do.
- If you can access the data and analysis code, ask yourself: could different methods change the results of the same analysis?

1.2.4.4 What can we conclude – the story so far?

This is a good place to look at what we have discussed, and present an evaluation of the story so far.

This is not a story where everybody or nobody is right or where everything or nothing is true [3]. Instead, we can be guided by the advice (Meehl, 1967; Scheel, 2022; Steegen et al., 2016a) that we should (1.) seek better and more complete theorizing about the constructs of interest and how we measure them, and (2.) seek more complete and more precise theory so that some options are theoretically superior than others, and should be preferred, when constructing datasets or specifying analysis methods.

Not all research questions and not all hypothesis information will allow an equally wide variety of potential reasonable approaches to the analysis. As Paul Meehl argued a long time ago (Meehl, 1967, 1978), and researchers like Anne Scheel (Scheel et al., 2021; Scheel, 2022) argued more recently, the complexity of the thing we study – people, and what they do – and the still early development of our understanding of this thing, mean that what we *want* but what we do not see, in psychology, are scientifically productive tests of *falsifiable theories*. (See, consistent with this perspective, discussions by Auspurg & Brüderl (2021) and by Del Giudice & Gangestad (2021) about the range of analysis possibilities that may or may not be allowed, in multiverse analyses, by more or less clear research questions or well-developed causal theories.)

[3] There could be a story where the hero (us) ultimately learns to reject binary (present, absent; significant, non-significant) choices, and embrace variation, or embrace uncertainty (a. Gelman, 2015; Vasishth & Gelman, 2021).

Our concern should not so much be with being able to do statistical analysis, or with finding significant or not significant results. It would be more useful to do analyses to test concrete, inflexible, precise predictions that *can be wrong*.

Nor is this a story, I think, about the potential for *cheating*. While we may refer to subjective choices or to researcher flexibility, the differences that we see do not resemble the *researcher degrees of freedom* (Simmons et al., 2011b) some may exploit, consciously or unconsciously, to change results to suit their aims. Instead, the multiverse results show us the impact of the reasonable differences in approach that different researchers may sensibly choose to take when they try to answer a research question with data.

Not all alternates, at a given point of choosing, in the data analysis workflow, will have equal impact. Work by Young (Young, 2018; Young & Holsteen, 2017) indicates that if we deliberately examine the impact of method or model uncertainty, over different sets of possible choices — about what variables or what observations we include in an analysis, for example — we may find that some results are robust to an array of different options, while other results are highly susceptible to different choices. This work suggests another way in which uncertainty about methods or variation in results can be turned into progress in understanding the phenomena that interest us: through systematic, informed, interrogation of the ways that results can vary.

In general, in science, the acceptance of research findings must always be negotiated (Bourdieu, 2004). Here, we see that the grounds of negotiation should often include an analysis of the impact on the value of evidence of the different analysis approaches that researchers can or do apply to the data that underly that evidence.

But you will be asking yourself: **what do I need to think about, for the assignment?**

> Tip
>
> - The results of multiverse analyses show us that if we see one analysis reported in a paper, or one workflow, that does not mean that only one analysis can reasonably be applied.
> - If you read the methods or results section of a paper, you should reflect: what other analysis methods could be used here? How could variation in analysis method — in what or how you do the analysis — influence the results?

Making you aware of the potential for analysis choices is useful because developing researchers, including graduate students, are often not aware of the room for choice in the data analysis workflow. Developing researchers — you — may be instructed that "this is how we do things" or "you should follow what researchers did previously". Following convention is not necessarily a bad thing: it is a feature of the normal practice of science (Kuhn, 1970). However, you can now see, perhaps, that there likely will be alternative ways to process or to analyse data than the approach a supervisor, lab or field normally adopts.

This understanding or awareness has three implications for practice, it means:

1. When we talk about the analysis we do, we should explain our choices.
2. We should check, or enable others to check, what impact making different choices would have on our results.
3. Most importantly: **we can allow ourselves the freedom to critically evaluate** the choices researchers make, even the choices researchers make in published articles.

1.2.5 From the multiverse to kinds of reproducibility

Multiverse analyses and post-publication analyses, in general, show that we can and should question or critically evaluate the analyses we encounter in the literature. This work can usefully detect problems in original published analyses (e.g., A. Gelman & Weakliem, 2009; Herndon et al., 2014; Wagenmakers et al., 2011). It can demonstrate where original published claims are or are not robust to variation of analysis method or approach.

Given these lessons, and the implications we have identified, we should expect or hope to see open science practices (Munafò et al., 2017; Nosek et al., 2022):

- share data and code;
- publish research reports in ways that enable others to check or query analyses.

As we discuss, following, these practices are now common but the quality of practice can sometimes be questioned. This matters for you because it makes it more challenging – in specific identifiable locations – to locate, access, analyse and report previously collected data.

The discussion of current practices identifies where or how the assignment may be more challenging, but also identifies some of the exact places where the assignment provides **a real opportunity to do original research work**.

First, I am going to introduce some ideas that will help you to think about what you are doing when you do this work. We focus on the concept of *reproducibility*.

Gilmore et al. (2017; following Goodman et al., 2016) present three kinds of reproducibility:

- methods reproducibility
- results reproducibility
- inferential reproducibility

In looking at reproducibility, here, we are considering how much, or in what ways, the results or the claims that are made in a published study can be found or *repeated* by someone else.

1.2.5.1 Methods reproducibility

As Gilmore et al. (2017) discuss, **methods reproducibility** means that another researcher should be able to get the same results if they use the same tools and analysis methods to analyse the same dataset [some researchers also refer to *analytic reproducibility* or *computational reproducibility*; see e.g. Crüwell et al. (n.d.); Hardwicke et al. (2018); Hardwicke et al. (n.d.); Laurinavichyute et al. (2022); Minocher et al. (n.d.)].

In neuroimaging, the multiplicity of possible implementations of the data analysis pipeline (Carp, 2012a), and the fact that important elements or information about the pipeline deployed by researchers may be missing from published reports (Carp, 2012b), can make it challenging to identify how results can be reproduced.

In psychological science, in evaluating reports of results from analyses of behavioural data collected through survey or experimental work, in principle, we should expect to be able to access the data collected by the study authors, follow the description of their analysis method, and reproduce the results they report.

> Tip
>
> - For an assignment in which we ask students to locate, access, analyse and report previously collected data, we are directly concerned with *methods reproducibility*.

1.2.5.2 Results reproducibility

Results reproducibility means that if another researcher completes a new study with new data they are able to get the same results as the results reported following an original study: this often referred to as *replication*. The replication studies that have been reported (e.g., Aarts et al., 2015), and continue to be reported (see, for example, the studies discussed by Nosek et al. (2022)), in the last several years, present attempts to examine the results reproducibility of published findings.

In the classes on the linear model, we will examine if similar or different results are observed in a series of studies using the same procedure and the same materials. We shall discuss, in those classes, in more depth, what results reproducibility (or study replication) can or cannot tell us about the behaviours that interest us.

1.2.5.3 Inferential reproducibility

Inferential reproducibility means that if a researcher repeats a study (aiming for results reproducibility) or re-analyzes an original dataset (aiming for methods reproducibility) then they can come to the same or similar conclusions as the authors of the report of an original study.

How is inferential reproducibility not methods or results reproducibility? Goodman et al. (2016) explain that researchers can make the same conclusions from different sets of results and *can reach different conclusions from the same set of results.*

How is it possible to reach different conclusions from the same results? We can imagine two scenarios.

First, we have to think about the wider research field, the research context, within which we consider a set of results. It may be that two different researchers will come to look at the same results with different expectations about what the results *could* tell us (in Bayesian terms, with different prior expectations). Given different expectations, it is easy to imagine different researchers looking at the same results and, for example, one researcher being more skeptical than another about what conclusion can be taken from those results. (In the class on graduate writing skills, I discuss in some depth the importance of reviewing a research literature in order to get an understanding of the assumptions, conventions or expectations that may be shared by the researchers working in the field.)

Second, imagine two different researchers looking at the same results — picture the original authors of a published study, and someone doing a post-publication re-analysis of their data — you can expect that the re-analysis or the reproducibility analysis could identify reasons to value the evidence differently, or to reach more skeptical conclusions, through critical evaluation of:

- data processing choices;
- the choice of the method used to do analysis;
- choices in how the analysis method is used.

Where that critical evaluation involves an analysis of the choices the original researchers made, perhaps involving an analysis of other choices they could have made, perhaps reflecting on how effectively the analyses address a given research question or test a given prediction.

Tip

- We can think about the work we do, when we analyse previously reported data, in terms of the need to identify the *reproducibility* of results, methods and inferences.
- In psychological science, determining that someone can get the same results, by analyzing the same data, or will reach the same conclusions from the same results, are important – potentially, original – research contributions.

1.2.6 The current state of the match between open science ideas and practices

I have said that we should expect or hope to see open science practices (Munafò et al., 2017; Nosek et al., 2022) where researchers:

- share data and code;
- publish research reports in ways that enable others to check or query analyses.

This raises an important question: **What exactly do we see, when we look at current practices?** The question is important because answering it helps to identify where the challenges are located when you complete your work to locate, access, analyse and report previously collected data.

I break the discussion of what we see into two parts. Firstly, I look at the results of audits of data and code sharing (see Section 1.2.6.2): are data shared and can we access the data? Secondly, I discuss analyses of methods reproducibility, and shared data and code usability (see Section 1.2.6.3): can others reproduce the results reported in published articles, given shared data? can others access and run shared analysis code? can others use the shared code to reproduce the reported results? Again, I need to be brief but reference sources that you can follow-up.

1.2.6.1 The link between the credibility revolution and the reproducibility of results

I should be clear, before we go on, about **the link** between the *credibility revolution* in science, and the effort to examine reproducibility of results. Many elements of the credibility revolution emerged out of the observation that it has often been difficult to repeat the results of published studies when we conduct new studies (replication studies or results reproducibility; e.g., Aarts et al. (2015)). However, it is clearly difficult to know *what* to replicate or reproduce if we cannot reproduce the results presented in a study report (methods reproducibility), given the study data (Artner et al., 2021; Laurinavichyute et al., 2022; Minocher et al., n.d.).

1.2.6.2 Data and code sharing

Research on data and code sharing practices suggest that practices have improved, from earlier low levels.

In an important early report, Wicherts et al. (2006) observed that it was very difficult to obtain data reported in psychological research articles from the authors of the articles. They asked for data from the lead authors of 141 articles published in four leading psychology journals, for about 25% of the studies. This low response rate was found despite the fact that authors in these journals must agree to the principle that data can be shared with others wishing to verify claims.

Practice *has* changed: how?

One change to practice has involved the use of **open science badges**. In journals like Psychological Science authors of articles may be awarded badges — Open Data, Open Materials, Preregistration badges — by the editorial team. Authors can apply for and earn the badges

by providing information about open practices, and journal articles are published with the badges displayed near the front of the articles.

In theory, initiatives like encouraging authors to earn open science badges should mean that data sharing practices improve, enabling access to data and code for those, like you, who would like to re-analyze previously published data. In theory, all you should need to do — to locate and access data — is just search articles in the journal *Psychological Science* for studies with open data badges, and follow links from the published articles to then access study data at an open repository like the *Open Science Framework* (OSF) What do we see in practice?

Analyses reported by Kidwell et al. (2016) as well as analyses reviewed by Nosek et al. (2022) indicate that more articles have claimed to make data available in the time since badges were introduced. When they did their analysis, Kidwell et al. (2016) found that a substantial proportion, but not all, of the articles in *Psychological Science* can be found to actually provide access to shared data. However, critically, many but not all the articles with open data badges provide access to data available through an open repository, data that are correct, complete and usable (Kidwell et al., 2016). In their later report, the analyses reviewed by Nosek et al. (2022) suggest that the use of repositories like OSF for data sharing may be accelerating but that, over the last few years, the rate at which open science practices like sharing data, overall, appears to be substantial but not yet reported or observed in a majority of the work of researchers.

Many journals now require the authors of articles to include a **Data Availability Statement** to locate their data. Analyses by Federer (2022) indicate that Data Availability Statements for articles published in the open access [4] journal PLOS ONE often, helpfully, include Digital Object Identifiers (DOIs) or Universal resource locators (URLs) enabling direct access to shared data (i.e., without having to contact authors). Of those DOIs or URLs, most appeared to be associated with resources that could successfully be retrieved. In contrast, analyses reported by Gabelica et al. (2022) that where article authors state that "data sets are available on reasonable request" (the most common availability statement), most of the time, the authors did not respond or declined to share the data (see similar findings, across fields, by Tedersoo et al., 2021). Clearly, in the analyses of open science practices we have seen so far, data sharing is more effective where sharing does not have to work through authors.

Tip

- When you are looking for a study in order to get data that you can then reanalyze, it makes sense to look, first, for studies focusing on research questions that interest you.
- When you are looking for published reports where the authors share data, look for articles with open science badges or where you can see a Data Availability Statement.

[4] Open access journals publish articles that are free to read or download.

- Choose articles where the authors provide a direct link to their data, where the data are located on an open repository like the Open Science Framework (there are other repositories).

1.2.6.3 Enabling others to check or query analyses

Research on data and code sharing practices suggest that practices have improved but that there are concerns about the quality of the sharing. Here, the critical concern relates to the word *enable* in the objective: that we should publish research reports in ways that *enable* others to check or query analyses.

John Towse and colleagues (Towse et al., 2021) examined the quality of open datasets to assess their quality in terms of their completeness and reusability (see also Roche et al., 2015).

- **completeness**: are all the data and the data descriptors supporting a study's findings publicly available?
- **reusability**: how readily can the data be accessed and understood by others?

For a sample of datasets, they found that about half were incomplete, and about two-thirds were shared in a way that made them difficult to use. Practices tended to be slightly better in more recent publications. (Broadly similar results are reported by (Hardwicke et al., 2018).)

Where data were found to be incomplete, this appeared to be, in part, because participants were excluded in the processing of the data for analysis but this information was not in the report, or because data were shared without a guide or "readme" file or data dictionary (or codebook) explaining the structure, coding or composition of the shared data.

Potentially important for future open science practices, (Towse et al., 2021; also Roche et al., 2015) found that sharing data as *Supplementary materials* may appear to carry risks that, in the long term, mean that data may become inaccessible.

Tip

- When you locate open data you can access, look for a guide, "readme" file, codebook or data dictionary explaining the data: you need to be able to understand what the variables are, what the observations relate to (observations per person, per trial?) and how variables are coded.
- Locate and examine carefully the parts of the published report, or the data guide, where the authors explain how they processed their data.

A number of studies have been conducted to examine whether shared data and analysis code can be reused by others to reproduce the results reported in papers (e.g., Artner et al., 2021; Crüwell et al., n.d.; Hardwicke et al., n.d.; Hardwicke et al., 2018; Laurinavichyute et al., 2022;

Minocher et al., n.d.; Obels et al., 2020; see Artner et al., 2021 for a review of reproducibility studies). In critical respects, the researchers doing this work are doing work similar to the work we are helping students to do, locating, accessing, and analyzing previously collected data. In these studies, typically, the researchers progressed through a series of steps.

1. Searched the articles published in a journal (e.g., *Cognition*, the *Journal of Memory and Language*, *Psychological Science*), published in a topic area across multiple journals (e.g., social learning, psychological research), or associated with a specific practice (e.g., registered reports.
2. Selected a subset of articles where it was identified that data could be accessed.
3. Identify a target result or outcome to reproduce, for each article. In their analyses, Hardwicke and colleagues (Hardwicke et al., n.d.; Hardwicke et al., 2018) focused on attempting to reproduce primary or *straightforward and substantive* outcomes: substantive – if emphasized in the abstract, or presented in a table or figure; straightforward – if the outcome could be calculated using the kind of test one would learn in an introductory psychology course (e.g., t-test, correlation).
4. Attempted to reproduce the results reported in the article, using the description of the data analysis presented in the article, and the analysis code (if provided), in some cases asking for information from the original study authors, in other cases working independently of original authors.

What the reproducibility studies appear to show is that, for many published reports, *if* data are shared and *if* the shared data are accessible and reusable *then*, most of the time, the researchers **could reproduce** the results presented by the original study authors (Hardwicke et al., n.d.; Hardwicke et al., 2018; Laurinavichyute et al., 2022; Minocher et al., n.d.; Obels et al., 2020; but see Crüwell et al., n.d.). This is great. But what is interesting, for us, is where the reproducibility researchers encountered challenges. You may encounter the same or similar challenges.

I list some challenges that the researchers describe, following. Before you look at the list, I want to assure you: you will not find *all* these challenges present for any one article you look at. Most likely, you will find one or two challenges. Obviously, some challenges will be more difficult than others.

> Tip
>
> - When you find a study you are interested in, with open data and maybe open analysis code, your main challenge will often be to identify exactly what analysis the original study authors did to answer their research question.
> - Locate and examine carefully the parts of the published report where the authors explain how they did the analysis that gave them their key result. Usually that key result should be identified in the abstract or in the conclusion.

1.2.6.3.1 Data challenges

1. Data Availability Statements or open science badges indicate data are shared but data are not directly accessible through a link to an open repository.
2. The data are shared and accessible but there is missing or incorrect information *about* the data. The documentation, codebook or data dictionary is missing or incomplete. There is unclear or missing information about the variables or the observations, or about the coding of variable values, responses.
3. Original study authors may share raw and processed data or just processed or just raw data. It may not be clear how raw data were processed to construct the data analysed for the report. It may not be clear how variables were transformed or calculated or processed.
4. There may be mismatches between the variables referred to in the report and the variables named in the data file. It may be unclear how a data file corresponds to a study described in a report, where there are multiple studies and multiple data files.

1.2.6.3.2 Analysis challenges

1. The original report includes a description of the analysis but the description of the analysis procedure is incomplete or ambiguous.
2. There may be a mismatch, in the report, between a hypothesis, and the analysis specified to test the hypothesis (maybe in the Methods section), compared to a long sequence of results reported in the Results section. This makes it difficult to identify the key analysis.
3. It is easier to reproduce results if both data and code are shared because the presentation of the analysis code usually (not always) makes clear what analysis was done to get the results presented in the report.
4. Sometimes, analysis code is shared but it is difficult to use because it requires proprietary software (e.g., SPSS) or because it requires function libraries that are no longer publicly available.
5. Sometimes, there are errors in the analysis. Sometimes, there are errors in the presentation of the results, where results have been incorrectly copied into reports from analysis outputs.

1.3 This is why

The research report assignment requires students to locate, access, analyse and report previously collected data. At the start of the introduction, I said I would explain the answer to the question:

- Why: what is the motivation for the assignment?

I summarize, following, the main points of the answer I have given. When you review these points, I want you to think about two things, returning to the ideas of Bourdieu (2004) and Kuhn (1970) I sketched at the start.

Often what we do in science is guided by convention, the assumptions and habits of *normal practice* (Kuhn, 1970). These conventions can work in our minds so that if we encounter an *anomaly* or *discrepancy* between what we expect and what we find, in our work, we may usually blame ourselves: it was something wrong that we did or failed to do. It can cause us anxiety if we do not reproduce a result we think we should be able to reproduce (Lubega et al., n.d.). But I want you to understand, from the start, that sometimes, if you think you have found an error or a problem in a published analysis or a shared dataset, **you may be right**.

If there is anything we have learned, through the findings of replication studies, multiverse analyses, and reproducibility audits it is that people make mistakes, different choices are often reasonable, and we *always* need to check the evidence.

1.3.1 Summary: this is why

1. We are in the middle of a credibility revolution. The lessons we have learned so far oblige us to think about and to teach good open science practices that safeguard the value of evidence in psychology.
2. This matters, even if we do not care about scientific methods, because if we care about the translation into policy or practice – in clinical psychology, in education, health, marketing and other fields – what we do will depend on the value of the research evidence that informs policy ideas or practice guides.
3. Focusing on data analysis, it is useful to think about the whole *data pipeline* in analysis, the workflow that takes us from data collection to raw data to data processing to analysis to the presentation of results.
4. At every stage of the data pipeline, there are choices about what to do. There are not always reasons why we make one choice instead of another. Sometimes, we are guided by convention, example or instruction.
5. The existence of choices means the path we take, when we do data analysis, can be one path among multiple different *forking paths*.
6. For some parts of the pipeline – dataset construction, data analysis choices – reasonable people might make different decisions to sensibly answer the same research question, given the same data. This variation between pathways can be more or less important in influencing the results we see.
7. If results tend to stay similar across different ways of doing analysis, we might conclude that the results are reasonably robust across contexts, choices, or other variation in methods.
8. To *enable* others to see what we did (versus what we could have done), to see how we got to our results from our data, it is important to share our data and code.

9. Everyone makes mistakes and we should make it easy for others, and ourselves, to find those mistakes by sharing our data and code in accessible, clear, usable ways.
10. We need to **teach and learn** how to share effectively the data and the code that we used to answer our research questions.

In constructing the assignment – in asking *and supporting* students to locate, access, analyse and report previously collected data – we are presenting an opportunity to really investigate and evaluate existing practices.

You may find that this work is challenging, in some of the places that reproducibility research has identified there can be challenges. Where the challenges cannot be fixed – if you have found an interesting study but the study data are inaccessible or unusable – we will advise you to move on to another study. Where the challenges can be fixed – if data require processing, or if analysis information requires clarification – we will provide you with help or enabling information so that you fix the problems yourself.

Tip

- Maybe the main lesson from this exercise is a reminder of the *Golden rule*: **treat others as you would like to be treated**.
- If it is frustrating when it is difficult to understand information about an analysis or about data, or when it is difficult to access and reuse shared data and code.
- When it is your turn, do better, reflecting on what frustrated you.

One last question: why not just do less demanding or challenging tasks? Because this is part of what makes graduate degree valuable, what will make you more skilled in the workplace. Most of the time, we work in teams, we inherit problems or data analysis tasks, or are given results with partial information. The lessons you learn here will help you to effectively navigate those situations.

2 What

2.1 PSYC401 Project – research report – what you are expected to do

We present the following guidelines to help you to complete the coursework assessment. If you have any questions, email Padraic Monaghan at: `p.monaghan@lancaster.ac.uk`

Note the information mirrors exactly the information provided on Moodle:

https://modules.lancaster.ac.uk/mod/page/view.php?id=1921399

2.1.1 What data can I analyse?

Reports will concern, usually, findings from analyses of data-sets we have provided to you. Some students may wish to analyse data collected in previous studies or data accessed from online sources: they should correspond with Padraic Monaghan or Rob Davies if they wish to do so.

The evaluation of reports will focus on clarity, read the following for discussion of what is required.

We expect students to use one of the analysis methods taught in the module. Marks will be awarded depending:

- on how appropriate the method is to the context, to the study design, to answering the research question, and to the features of the data; the appropriateness of methods to contexts will be taught in class;
- on how effectively the analysis is explained; students must explain the motivations for their decisions, explain their methods, and explain their findings effectively to gain points.

2.1.2 What structure should reports take?

1. The reports should include abstract, introduction, methods, results, discussion and references sections, like a short research article in the journal *Psychological Science*. You can view examples of articles here

https://journals.sagepub.com/toc/PSS/current

2. Word count limit: no more than 1500 words are allowed for all materials.
3. Unlike a published research article, for PSYC401, the Results and Discussion sections must be written in full, but the Introduction and Methods sections can be written in the form of notes.

2.1.3 What content should reports present?

2.1.3.1 Introduction and Method sections

The focus of marking will be on the quality of the Results and Discussion sections. This means you can write your notes in the Introduction and Methods sections as short answers to the following questions:-

2.1.3.1.1 Introduction

- What did the researchers do and why did the researchers do it?
- What was the question addressed in the study and why is it interesting?
- What were the hypotheses?
- What results were expected and how would they relate to the hypotheses?

How can you write this as a set of notes? We require main points of information on the hypotheses concerning expected results. We will ignore the absence of citations, or of explanations of critical previous experimental work, in the Introduction.

2.1.3.1.2 Method

Note the origin of the data at the start of the method section. As for the Introduction, your method section writing needs to furnish answers to questions like the following:-

- What was done to collect the data?
- Who were tested (Participants)?
- What materials were used in testing (Materials)?
- What was the design of the study?
- What procedure was used?

How can you write this as a set of notes? We require main points of information, especially the main features of the data analyzed – what were the variables, how many observations were recorded, what exclusions or other data treatment steps were applied?

2.1.3.2 Results and Discussion sections

The focus of marking will be on the quality of the Results and Discussion sections. This means you must write in complete sentences in full paragraphs in a style appropriate for a research article appearing in a journal like *Psychological Science*. You must not use notes for these sections. You must write text that explains to the reader the analysis you did, why you did it, the results you found, and the implications of those results. You should write the text for the sections so that the questions listed following are answered fully.

If you use a data set that is already published in a journal such as *Psychological Science*, then your presentation of the results must differ from that in the article in ways that highlight new features of the data.

2.1.3.2.1 Results

Be clear on what the outcome measure or dependent variable for analysis was, and on what factors or predictor variables were brought into the analysis of that outcome. You then need to ensure the results section answers the following questions:-

- What hypotheses were tested?
- What methods were used to test the hypotheses?
- Why are they appropriate?
- What were the results? What were the direction and relative size of effects?

Do what seems reasonable using one or more of the analysis methods practiced in class, or practiced in association with the workbooks, and explain your reasoning.

2.1.3.2.2 Discussion

What the reader must be able to do, given your report, is understand the answer to the following questions:

- What are the theoretical implications of the study findings?
- What are the practical implications?

Reports should present **enough information that the reader can understand**: the background and motivation for a study; the features of the data analyzed and the methods of data collection; the approach taken in analysis, the analysis steps, and the results; the relationship between the observed results and the expected results, and the interpretation of findings in relation to previous work. To be clear about clarity: explain, spell things out (decisions, reasoning, interpretations) as if you were explaining them to a reasonably intelligent reader, a Psychologist who is not a specialist in the area of study occupied by the study reported, i.e. me. The main point is that you should keep in mind what the reader should get out of (what benefit) reading your report.

2.1.4 What format?

2.1.4.1 Statistics, tables and figures should follow APA guidelines. See here for a free guide:

For general APA formatting of reports:

https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_style_introduction.html

And for APA formatting of statistics and numbers:

https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/apa_numbers_statistics.html

Though the APA guidelines are the authoritative guide.

2.1.4.2 Add a link to the data analysed for the report

3 How

The research report assignment requires students to locate, access, analyse and report previously collected data. Here, we answer the question:

- How can the assignment be done?

We outline the workflow you can follow, proceeding through a series of steps to complete the essential tasks. Look at this outline, make a plan, and then follow the advice, taking it **one step at a time**.

3.1 The variety of things students do

Students have taken a variety of approaches to the assignment.

- Some students choose to complete an analysis of a publicly available dataset, analyzed previously, data for which the report has been published in a journal article.
- Some students choose to complete an analysis of a publicly available dataset that has been made available (for a report published as a data journal) but has not been analysed previously.
- Some students choose to complete an analysis of one of the data-sets used for practical exercises in class: the example or demonstration data we collect together as the *curated data*.

Ask in class or on the discussion forum for advice about any one of these approaches.

Here, I offer guidance on what to do if you want to locate, access, and analyse previously collected data where those data are presented in a journal article. I consider, first, working with datasets where an analysis of the data has been presented in the article (see Section 3.2). I then look at working with datasets where the data are presented without an analysis (see Section 3.3). Our advice on working with datasets presented without an analysis will overlap in key respects with our advice on working with curated data.

3.2 Working with data associated with a published analysis

In the following, I split our guidance into two parts. I look next at the task of locating, accessing and checking the data (Section 3.2.1). Then I look at the task of figuring out what analysis you can do with the data (see Section 3.2.2). Obviously, you cannot consider an analysis if you cannot be sure that you can work with the data (Minocher et al., n.d.).

3.2.1 Locate, access and check the data

At the start of your work on the assignment, you will need to (1.) locate then (2.) access data for analysis, and then you will need to (3.) check that the data are usable. I set out advice on doing each step, following. Work through the steps: **one step at a time**.

3.2.1.1 Locate

It is usually helpful to find a dataset where the data have been collected in a study within a topic area you care about, or could be interested in. It is helpful because you will need to work with the data and it will be motivating if you are interested in what the data concern. And it is helpful because, often, you will need to do a bit of reading on related research to learn about the context for the data collection, and you will usually want to read research sources that interest you.

> Tip
>
> The task here is:
>
> - Do a search: look for an article with usable data in a topic area that interests you.

There are at least two ways you can do this. Both should be reasonably quick methods to get to a usable dataset.

1. Do a search on Google scholar).
2. Do a search on the webpages of a journal.

Most psychological research is published in journals like *Psychological Science*. If you want, you can look at a list of psychology journals here.

In a journal like *Psychological Science* you can look through lists of previously published articles (in issues, volumes, by year) on the journal webpage. Here is the list of issues for *Psychological Science*..

3.2.1.1.1 Key words

In both methods, you are looking for an article associated with data (and maybe analysis code) you can access and that you are sure you can use. In both methods, you need to first think about some **key words** to use in your search. Ask yourself:

- What are you interested in? What population, intervention or effect, comparison, or outcome?

Then:

- What words do people use, in articles you have seen, when they talk about this thing?

You can use these words, and maybe consider alternate terms. For example, I am interested in `reading comprehension` or `development reading comprehension` but researchers working on reading development might also refer to `children reading comprehension`.

You want to be as efficient as possible so combine your search for articles in an interesting topic area with your search for accessible data. We can learn from the research we discussed on data sharing practices (see Section 1.2.6.2) by looking for specific markers that data associated with an article should be accessible.

If you are doing a search (1.) on Google scholar), I would use the key words related to your topic plus words like: `open data badge`; `open science badge`. So, I would do a search for the words: `reading comprehension open data badge`. I have done this: you can try it. The search results will list articles related to the topic of reading comprehension, where the authors claim to have earned the open data badge because they have made data available.

If you are doing a search (2.) in a journal list of articles, then what you are looking for are articles that interest you and which are listed with open data badges. In the listing for *Psychological Science* (here)) a quick read of the journal issue articles index shows that article titles are listed together with symbols representing the open science badges that authors have claimed.

In other journals (e.g., *PLOS ONE*, *PeerJ*, *Collabra*), you may be looking for interesting articles with the words `Data Availability Statement`, `Data Accessibility Statement`, `Supplementary data` or `Supplementary materials` in the article webpage somewhere. Journals like *PeerJ* or *Collabra*, in particular, make it easy to locate data associated with published articles on their web pages.

In *Collabra*, you can find published articles through the journal webpage (here). If you click on the title of any article, and look at the article webpage, then on the left of the article text, you can see an index of article contents and that index lists the `Data Availability Statement`. Click on that and you are often taken to a link to a data repository.

3.2.1.2 Access

If you have located an interesting article with evidence (an open data badge or a data accessibility statement) that the authors have shared their data, you need to check that you can access the data. Most of the time, now, you are looking for a link you can use to go directly to the shared data. The link is often presented as a hyperlink on a webpage, associated with Digital Object Identifiers (DOIs) or Universal resource locators (URLs). Or, increasingly, you are looking for a link to a data repository on a site like the Open Science Framework (OSF).

> Tip
>
> The task here is:
>
> - Access the data associated with the article you have found.

Here are some recent examples from my work that you can check, to give you a sense of where or how to find the accessible link to the shared data.

Ricketts, J., Dawson, N., & Davies, R. (2021). The hidden depths of new word knowledge: Using graded measures of orthographic and semantic learning to measure vocabulary acquisition. Learning and Instruction, 74, 101468. https://doi.org/10.1016/j.learninstruc.2021.101468

Rodríguez-Ferreiro, J., Aguilera, M., & Davies, R. (2020). Semantic priming and schizotypal personality: Reassessing the link between thought disorder and enhanced spreading of semantic activation. PeerJ, 8, e9511. https://doi.org/10.7717/peerj.9511

These are both open access articles.

If you look at the webpage for, Rodríguez-Ferreiro et al. (2020), (here)), you can do a search in the article text for the keyword `OSF` (on the article webpage, use keys `CMD-F` plus `OSF`). You are checking to see if you can click on the link *and* and if clicking on the link takes you to a repository listing the data for the article. The Rodríguez-Ferreiro et al. (2020) article is associated with a data plus analysis code repository (OSF))

Notice that on the repository webpage, you can see a description of the project plus .pdf files and a folder `Dataset and Code`. If you can click through to the folders, and download the datafiles, you have accessed the data successfully.

I have guided you, here, through to the Rodríguez-Ferreiro et al. (2020) data repository, can you find the data for the Ricketts et al. (2021) repository?

3.2.1.3 Check

If you have located an interesting article with data that you can access, and if you have read the introductory notes (see Section 1.2.6.3), then you will know that you need to make sure that you can use the data.

> **Tip**
>
> The task here is:
>
> - Check the data and the data documentation to make sure you can understand what you have got *and* whether you can use it.

What make data usable are:

1. Information in the article, or in the data repository documentation, on the study design and data collection methods: you need to be able to understand where the data came from, how they were collected, and why.
2. Clear data documentation: you need to find information on the variables, the observations, the scoring, the coding, and whether and how the data were processed to get them from raw data state to the data ready for analysis.

Data documentation is often presented as a note or a wiki page or a miniature paper and may be called a *codebook, data dictionary, guide to materials* or something similar. You will need to check that you can find information on (examples shown are from the Rodríguez-Ferreiro et al. (2020) OSF *guide to materials*):

- what the data files are called e.g. `PrimDir-111019.csv`;
- how the named data files correspond to the studies presented in the report;
- what the data file columns are called and what variables the column data represent e.g. `relation, coding for prime-target relatedness condition ...`;
- how scores or responses in columns were collected or calculated e.g. `age, giving the age in years ...`;
- how coding was done, if coding was used e.g. `biling, giving the bilingualism status`;
- whether data were processed, how missing values were coded, whether participants or observations were excluded before analysis e.g. `Missing values in the rt column ... coded as NA`

If these information are not presented, or are not clear: **walk away**.

3.2.2 Plan the analysis you want to do

After you have found an interesting article, and have confirmed that you can use the associated data, you will need to plan what analysis you want to do.

> **Tip**
>
> The task here is:
>
> - Identify and understand the analysis in the article.
> - Work out what analysis *you* want to do.

Students have taken a variety of approaches to the assignment.

- Some students choose to complete a reanalysis of the data, in an attempt to reproduce the results presented in the article (see Section 3.2).
- Some students choose to complete an alternate analysis of the data, varying elements of the analysis (see Section 1.2.4).

Either way, you will want to first make sure you can identify exactly *what* the authors of the original study did, *how* they did it, and *why* they did it.

You can process the key article information efficiently using the *QALMRI* method we discussed in the class on graduate writing skills (Brosowsky et al., n.d.; Kosslyn & Rosenberg, 2005). You are first aiming to **locate** information on the broad and the specific question the study addresses, the methods the study authors used to collect data, the results they report, and the conclusions they present given the results. Can you find these bits of information?

3.2.2.1 Are you interested in attempting a methods reproducibility test?

Following Hardwicke and colleagues (Hardwicke et al., n.d.; Hardwicke et al., 2018) it would be sensible to focus on identifying the primary or *substantive* result for a study in an article.

- **Substantive** if emphasized in the abstract, or presented in a table or figure.

As we discussed in the class on graduate writing skills, the article authors *should* signal what they consider to be the primary result for a study by telling you that a result is critical or key or that a result is the or an answer to their research question.

> **Tip**
>
> - An article may present multiple studies: focus on one.
> - The results section of an article, for a study, may list multiple results: identify the primary or substantive result.

If you are, then you will want to identify a result that is both substantive and *straightforward* (Hardwicke et al., n.d.; Hardwicke et al., 2018).

- **straightforward** if the outcome could be calculated using the kind of test you have been learning about or will learn about (e.g., t-test, correlation, the linear model)

Psychological science researchers use a variety of data analysis methods and not all the analyses that you read about will be analyses done using methods that you know about. The use of the methods we teach — t-test, correlation, and the linear model — are very *very* common; that is why we teach them. But you may also see reports of analyses done using methods like ANOVA, and multilevel or (increasingly) linear mixed-effects models (Meteyard & Davies, 2020).

In the research on the reproducibility of results in the literature (see Section 1.2.6.3), the researchers attempting to reproduce results often focused on answering the research question the original authors stated using the data the original authors shared. This does not mean that they always tried to *exactly* reproduce an analysis or an analysis result. Sometimes, that was not possible.

Sometimes, you will encounter an article and a dataset you are interested in but the analysis presented in the article looks a bit complicated, or more complex than the methods you have learned would allow you to do. In this situation, don't give up. What you can do – maybe with our advice – is identify *a part* of the primary result that you *can* try to reproduce. For example, what if the original study authors report a linear mixed-effects analysis of the effects of both prime relatedness and schizotypy score on response reaction time (Rodríguez-Ferreiro et al., 2020)? Maybe you have not learned about mixed-effects models, or you have not learned about analysing the effects of two variables but you *have* (you will) learn about analysing the effect of one variable using the linear model method: OK then, do an analysis of the shared data using the method you know.

You may be helped, here, by knowing about two good-enough (mostly true) insights from statistical analysis:

1. Many of the common analysis methods you see used in psychological science can be coded as a linear model.
2. More advanced common analysis methods — (Generalized) Linear Mixed-effects Models (GLMMs) — can be understood as more sophisticated versions of the linear model. (Conversely, the linear model can be understood as an approximation of a GLMM.)

There is a nice discussion of the idea that common statistical tests are linear models here.

Tip

- Identify the analysis method used to get the result you are interested in.
- If it is complex or unfamiliar, discuss whether a simpler method can be used.

- If the result is complex, discuss whether you can attempt to reproduce a part or a simpler result.

3.2.2.2 Are you interested in attempting a different analysis?

It can be interesting and important work to complete a simpler analysis of shared data. Sometimes, we learn that a simpler analysis is as good account of the behaviour we observe as other more complex analyses. This can happen if, for example, our theory predicts that two effects should work together but an analysis shows that we can explain behaviour in an account in which the two effects are independent. For example, Ricketts et al. (2021) predicted that children should learn words more effectively if they were shown the spellings of the words *and* they were told they would be helped by seeing the spelling but, in our data, we found that just seeing the spellings was enough to explain the learning we observed.

In completing analyses that *vary* from original analyses, we are engaging in the kind of work people do when they do *multiverse analyses* or *robustness checks* (see Section 1.2.4).

Tip

In planning an alternate or multiverse analysis, do not suppose that you need to do multiple analyses: you do not.

In planning an alternate or multiverse analysis, you will want to begin by critically evaluating the analysis you see described in the published article. I talk about how to do this, next.

Before we go on, note that I previously discussed an example of how to critically evaluate the results of published research in the context of Rodríguez-Ferreiro et al. (2020). Take a look at the Introduction of that article. There, we summarised the analyses researchers did previously and used the information about the analyses to explain inconsistencies in the research literature. We found limitations in the analyses that people did that had (negative) consequences for the strength of the conclusions we can take from the data.

3.2.2.2.1 Critically evaluate the analysis description

If you revisit our discussion of multiverse analyses, you will see that we discussed two things: (1.) analyses of the impact on results of varying how you construct datasets for analysis (Section 1.2.4.2) and (2.) analyses of the impact on results of varying what analysis method you use, or how you use the method (see Section 1.2.4.3). These are both good ways to approach thinking about the description of the analysis you see in a published article.

As we noted in Section 1.2.4.2, you almost always have to process the data you collect (in an experiment or a survey) before you can analyze the data. Often, this means you need to code for responses to survey questions e.g. asking people to self-report their gender, or you need to

identify and code for people making errors when they try to do the experimental task you set them, or you need to process the data to exclude participants who took too long to do the task (if taking too long is a problem). Not all of these processing steps will have an impact on the results but some might. This is why you can sometimes do **useful** and sometimes **original** research work in reanalyzing previously published data.

You can begin your analysis planning work by first identifying exactly what data processing the original study authors did then identifying what different data processing they could have done. Remember the research we discussed in relation to reproducibility studies, you need to be prepared for the possibility that it is challenging to identify what researchers did to process their data for analysis Section 1.2.6.3.1. To identify the information you need, look for keywords like `code, exclude, process, tidy, transform` in the text of the article, or look for words like this in the documentation you find in the data repository.

When you have identified this information, you can then consider three questions:

1. What data processing steps were completed before analysis?
2. What were the reasons given explaining why these processing steps were completed?
3. What could happen to the results if different choices were made?

Working through these questions can then get you to a good plan for an analysis of the data. For example, a simple but useful analysis you can do is to check what happens to the results if you do an analysis with data from all the participants tested, if participants are excluded (for some reason) in the data processing step. Obviously, if the original study authors *only* share processed data, you cannot do this kind of work. Another simple but useful analysis you can do is to check what happens to the results if you change the coding of variables. Sometimes different coding of categorical variables (e.g., ethnicity) are reasonable. For example, you can ask: what happens if you analyze the impact of the variable given a different coding? (In case you are reading these notes and thinking about recoding a factor, there are some useful functions you can use; read about them here.)

> **Tip**
>
> - Do you want to check the impact of varying data processing choices: check, do you need and have access to the raw data? can you see how to recode variables?

As we noted in Section 1.2.4.3, when we consider how to answer a research question with a dataset, it is often possible to imagine multiple different analysis methods: reasonable alternatives. Most often, this is most clearly apparent when we are looking at an *observational* dataset or data collected given a *cross-sectional* study design.

In *cross-sectional* or *observational* studies, we typically are not manipulating experimental conditions, and we are often analyzed data using some kind of linear model. We often collect data or have access to data on a number of different variables relevant to our interests. For example, in studies I have done on how people read (R. Davies et al., 2013; R. A. I. Davies

et al., 2017), we wanted to know what factors would predict or influence how people do basic reading tasks like reading aloud. We collected information on many different kinds of word properties and on the attributes of the participants we tested. (Note: the papers are associated with data repositories in Supplementary Materials.) It is an **open question** *which* variables should be included in a prediction model of the observed outcome (reading response reaction times). Therefore, if you are interested in a study like this, and can access usable data from the study, it will often be true that you are able to sensibly motivate a different analysis of the study data using a different choice of variables.

As discussed in a number of interesting analyses, over the years (e.g., Patel et al., 2015), researchers may be interested in the specific impact of one particular predictor variable (e.g., we may be interested in whether it is easier to read words we learned early in life), but will need to include in their analysis that variable plus other variables known to affect the outcome. In that situation, the effect of the variable of interest may appear to be different depending on what other variables are also analyzed. This makes it interesting and useful to check the impact of different analysis choices.

We will look at data like these, for analyses involving the linear model, in our classes on this method.

Tip

- Do you want to check the impact of different analysis choices: check, do you need and have access to a choice of variables?
- Can you think of some reasons to justify using a different choice of variables in your analysis.

3.2.3 Summary: working with data associated with a published analysis

Here's a quick summary of the advice we have discussed so far.

- At the start of your work, you will need to (1.) locate then (2.) access data for analysis, and then you will need to (3.) check that the data are usable.
- Once you have confirmed you have found interesting data you can use, you should plan your analysis.
- Students do a variety of kinds of analysis. Whatever your interest, you first will want to first make sure you can identify exactly what the authors of the original study did, how they did it, and why they did it.
- If you are interested in attempting a methods reproducibility test (can you repeat a result, given shared data?) you will perhaps benefit from focusing a result that is both substantive and straightforward.
- If you are interested in doing an alternate analysis, you can critically evaluate the data processing and the data analysis choices that the original study authors made. You

can consider whether other choices would be appropriate, and might sensibly motivate a (limited) investigation of the impact of a different analysis pipeline choice on the results.

What if you access interesting data that were shared without a previous analysis? We talk about that situation, next.

3.3 Working with data that are not associated with a published analysis

A number of datasets have been published online with information about the data but with no analysis. You can look for data that may be interest you in a number of different places, now, but I would focus on one. I talk about that next. Then I offer some guidance on how you might approach analyzing such data Section 3.3.2.

3.3.1 Looking for open data

Wicherts and colleagues set up the Journal of Open Psychology Data (JOPD) to make it easier for Psychologists to share experimental data. A link to the journal webpage is here) Usually, a data paper reports a study and provides a link to a downloadable dataset.

Some datasets that I have looked at in JOPD and other places include the following.

3.3.1.1 Wicherts intelligence and personality data

Wicherts did what he recommended and put a large dataset online here

You can analyse these data in a number of different interesting ways. You can explore relationships between gender, intelligence and personality differences.

The data file and an explanatory document are located at the end of the article. Read the article, it's worth your time. Wicherts reports:

The file includes data from our freshman-testing program called "Testweek" (Busato et al., 2000, Smits et al., 2011 and Wicherts and Vorst, 2010) in which 537 students (age: M = 21.0, SD = 4.3) took the Advanced Progressive Matrices (Raven, Court, & Raven, 1996), a test of Arithmetic, a Number Series test, a Hidden Figures Test, a test of Vocabulary, a test of Verbal Analogies, and a Logical Reasoning test (Elshout, 1976).

Also included are data from a Dutch big five personality inventory (Elshout & Akkerman, 1975), the NEO-PI-R (Hoekstra, Ormel, & Fruyt, 1996), scales of social desirability and impression management (based on work by Paulhus, 1984 and Wicherts, 2002), sex of the participants, and grade point averages of the freshmen's first trimester that may act as outcome variable.

3.3.1.2 Smits personality data

Smits and colleagues (including Wicherts) put an even larger dataset online at the Journal of Open Psychology Data here)

You will need to register to be able to download the data but the process is simple.

The Smits dataset includes **Big-5** personality scores for several thousand individuals recorded over a series of years. You can analyse these data in interesting ways including examining changes in personality scores among students over different years.

3.3.1.3 Embodied terror management

Tjew A Sin and colleagues shared a dataset at the Journal of Open Psychology Data on an interesting study they did to test the idea that interpersonal touch or simulated interpersonal touch can relieve existential concerns (fear of death) among individuals with low self-esteem. The data can be found here)

The Tjew A Sin can be downloaded from a link to a repository location, given at the end of the article. You will likely need to register to download the data. Note that the spreadsheets holding the study data include `999` values to code for missing data. Note also that the data spreadsheets include (in different columns) scores per participant for various measures e.g. mortality anxiety or self-esteem. The measures are explained in the paper. To use the data, you will need to work out the simple process of how to sum the scores across items to get e.g. a measure of self-esteem for each person.

3.3.1.4 Demographic influences on disgust

Berger and Anaki shared data on the disgust sensitivity of a large sample of individuals. The data are from the administration of the Disgust Scale to a set of Hebrew speakers. They can be found here)

The experimenters collected data on participants' characteristics so that analyses of the way in which sensitivity varies in relation to demographic attributes is possible. You will see that the disgust scale is explained in the paper. The different disgust scores, for each item in the disgust scale, can be found in different columns. The disgust scores, for person, are calculated overall as values: `Mean_general_ds, Mean_core, Mean_Animal_reminder, Mean_Contamination`

When you download the dataset, you may need to change the file name, adding a suffix: `.txt` (for the tab delimited file), to be opened in Excel, or `.sav` (for the SPSS data file), to be opened in SPSS – to the file name to allow you to open it in the appropriate application.

3.3.2 Thinking about analyses of open data

The availability of rich, curated, clearly usable datasets with many variables can make it challenging to decide what to do.

I would advise beginning with an exploratory analysis of the data you have accessed. You will want to begin by using the data visualization skills we have taught you to examine:

1. The distributions of the variables that interest you using histograms, density plots or bar charts.
2. The potential relationship between variables using scatterplots.

In such *Exploratory Data Analyses*, you are interested in what the data visualization tells you about the nature of the dataset you have accessed. The papers associated with the datasets can sometimes offer only outline information: how the data were collected, coded, and processed. You may need to satisfy yourself that there is nothing odd or surprising about the distributions of scores. This stage can help you to identify problems like survey responses with implausible scores.

The work you do in exploring, and summarizing, the data variables that interest you will often constitute a substantial element of the work you can do and present for your report. You may discuss, for advice, what parts of this work will be interesting or useful to present.

Then, our advice is simple.

> Tip
>
> - When working with open datasets, consider keeping the analysis *simple*.

Note that *simple* is relative. Do what interests you. Work with the methods you have learned or will learn (the linear model).

In practice, you will find that part of the challenge is located not in using the data or in running an analysis like a linear model, it is in (1.) justifying or motivating the analysis and (2.) explaining the implications of your findings.

Working on the thinking you must develop to motivate an analysis or to explain implications requires you to do some (limited) reading of relevant research. (Relevant sources will be cited in data papers, as part of their outline of the background for their data collection.) If you consider the advice we discussed in the graduate class on developing writing skills, you will see that there I talked about how you might extract data from a set of relevant sources (papers) to get an understanding of the questions people ask, the assumptions they make. That is the kind of process you can follow to develop your thinking around the analysis you will do. What you are looking for is information you can use so that you can say something brief about, for example, why it might be interesting to analyze, say, whether personality (measured using the Big-5) varies given differences in gender or differences between population cohorts. The

reading and the conceptual development should be fairly limited, not extensive, but should be sufficient that you can write something sensible when you introduce and then when you discuss your analysis results.

3.4 Summary: how

In this chapter, I have outlined some advice on how you might approach the task of locating, accessing, and analyzing previously collected data. The main advice is to think about your workflow in stages, then progress through the work one step at a time.

You will need to begin by assuring yourself that you can find a dataset that interests you, and that you can access and use the data. The usability of data will require clear, understandable, descriptions in the published article (if any) about the research question and hypothesis, the study design, the data collection methods, the data processing steps, and the data analysis (if any). Sometimes, useful information about data processing and data analysis can be found in detail in repository documentation (e.g., in guides to materials) but only referenced in the text of the article.

If you know you can locate, access and have checked data as usable, you will want to think about what analysis you want to do the data. The approach you take depending on what aims you would like to pursue.

If you are interested in attempting a methods reproducibility test (i.e. checking if you can repeat presented results, given shared data), then you will first need to identify a substantive and straightforward result to try to reproduce. If you identify a primary result to examine, you will want to check that you can work with the data that have been shared, and then that you can use the analysis methods you have learned to reproduce some or all of the result that interests you.

If you are interested in doing an alternate or a different analysis (from what may be presented), you may need to consider the information you can locate on data processing and on data analysis choices. Did the original study authors process the data before sharing it, how? are the raw data available? What analyses did the authors do and why? When you consider this information, you may critically evaluate the choices made. In the context of this critical evaluation, you may find good reasons to justify doing a different analysis, whether to examine the impact of making different data processing choices, or to examine the impact of using a different analysis method, or of applying the same method differently (e.g., by including different variables).

In considering an analysis of data shared without a published set of results, you may want to keep your approach simple. Focus on what analysis you can do using the methods you have learned. And think about the understanding you will need to develop, to justify the analysis you do, and to make sense, in the discussion of your report of the analysis results you will present.

It is always a good idea to explore your data using visualization techniques throughout your workflow.

> **Tip**
>
> - You can always get advice, do not hesitate to ask.
> - We are happy to discuss your thinking, especially in class.

4 Summary

To complete when book is completed.

References

Aarts, E., Dolan, C. V., Verhage, M., & Van der Sluis, S. (2015). Multilevel analysis quantifies variation in the experimental effect while optimizing power and preventing false positives. *BMC Neuroscience*, *16*(1), 1–15. https://doi.org/10.1186/s12868-015-0228-5

Aczel, B., Szaszi, B., Nilsonne, G., Akker, O. R. van den, Albers, C. J., Assen, M. A. van, Bastiaansen, J. A., Benjamin, D., Boehm, U., Botvinik-Nezer, R., Bringmann, L. F., Busch, N. A., Caruyer, E., Cataldo, A. M., Cowan, N., Delios, A., Dongen, N. N. van, Donkin, C., Doorn, J. B. van, … Wagenmakers, E.-J. (2021). Consensus-based guidance for conducting and reporting multi-analyst studies. *eLife*, *10*, e72185. https://doi.org/10.7554/eLife.72185

Artner, R., Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F., & Vanpaemel, W. (2021). The reproducibility of statistical results in psychological research: An investigation using unpublished raw data. *Psychological Methods*, *26*(5), 527–546. https://doi.org/10.1037/met0000365

Auspurg, K., & Brüderl, J. (2021). Has the Credibility of the Social Sciences Been Credibly Destroyed? Reanalyzing the "Many Analysts, One Data Set" Project. *Socius*, *7*, 23780231211024421. https://doi.org/10.1177/23780231211024421

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., Chen, M., Chow, S.-M., Jonge, P. de, Emerencia, A. C., Epskamp, S., Fisher, A. J., Hamaker, E. L., Kuppens, P., Lutz, W., Meyer, M. J., Moulder, R., Oravecz, Z., Riese, H., Rubel, J., … Bringmann, L. F. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of Psychosomatic Research*, *137*, 110211. https://doi.org/10.1016/j.jpsychores.2020.110211

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review*, *33*(4), 357–370. https://doi.org/10.1016/j.dr.2013.08.003

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Bourdieu, P. (2004). *Science of Science and Reflexivity*. Polity.

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H. V., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H. K., Auer, D., Azevedo, F., Bahnsen, O., Balzer, D., Bauer, G., Bauer, P. C., Baumann, M., Baute, S., Benoit, V., Bernauer, J., Berning, C., … Żółtak, T. (2022). Observing many researchers using the same data and hypothesis reveals a

hidden universe of uncertainty. *Proceedings of the National Academy of Sciences*, *119*(44), e2203150119. https://doi.org/10.1073/pnas.2203150119

Brosowsky, N., Parshina, O., Locicero, A., & Crump, M. (n.d.). *Teaching undergraduate students to read empirical articles: An evaluation and revision of the QALMRI method*. https://doi.org/10.31234/osf.io/p39sc

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Carp, J. (2012a). On the plurality of (methodological) worlds: Estimating the analytic flexibility of FMRI experiments. *Frontiers in Neuroscience*, *6*, 149.

Carp, J. (2012b). The secret lives of experiments: Methods reporting in the fMRI literature. *Neuroimage*, *63*(1), 289–300.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, *65*(3), 145–153. https://doi.org/10.1037/h0045186

Crüwell, S., Apthorp, D., Baker, B. J., Colling, L., Elson, M., Geiger, S. J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D. S., Zaneva, M., & Brown, N. J. L. (n.d.). *What's in a badge? A computational reproducibility investigation of the open data badge policy in one issue of psychological science*. https://doi.org/10.31234/osf.io/729qt

Davies, R. A. I., Birchenough, J. M. H., Arnell, R., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(8). https://doi.org/10.1037/xlm0000366

Davies, R., Barbón, A., & Cuetos, F. (2013). Lexical and semantic age-of-acquisition effects on word naming in spanish. *Memory and Cognition*, *41*(2), 297–311. https://doi.org/10.3758/s13421-012-0263-8

Del Giudice, M., & Gangestad, S. W. (2021). A Traveler's Guide to the Multiverse: Promises, Pitfalls, and a Framework for the Evaluation of Analytic Decisions. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920954925. https://doi.org/10.1177/2515245920954925

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., Hawkins, G. E., Heathcote, A., Holmes, W. R., Krypotos, A.-M., Kupitz, C. N., Leite, F. P., Lerche, V., Lin, Y.-S., Logan, G. D., Palmeri, T. J., Starns, J. J., Trueblood, J. S., Maanen, L. van, … Donkin, C. (2019). The Quality of Response Time Data Inference: A Blinded, Collaborative Assessment of the Validity of Cognitive Models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069. https://doi.org/10.3758/s13423-017-1417-2

Federer, L. M. (2022). Long-term availability of data associated with articles in PLOS ONE. *PLOS ONE*, *17*(8), e0272845. https://doi.org/10.1371/journal.pone.0272845

Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., Yo, T., Mangin, J.-F., & Poupon, C. (2011). Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. *NeuroImage*, *56*(1), 220–234. https://doi.org/10.1016/j.neuroimage.2011.01.032

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: a mixed-methods study. *Journal of Clinical Epidemiology*, *150*, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Gelman, a. (2015). The connection between varying treatment effects and the crisis of unreplicable research: A bayesian perspective. *Journal of Management*, *41*(2), 632–643. https://doi.org/10.1177/0149206314525208

Gelman, A., & Loken, E. (2014a). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Psychological Bulletin*, *140*(5), 1272–1280.

Gelman, A., & Loken, E. (2014b). The statistical crisis in science. *American Scientist*, *102*(6), 460–465. https://doi.org/10.1511/2014.111.460

Gelman, A., & Weakliem, D. (2009). Of beauty, sex and power. *American Scientist*, *97*(4), 310–316. https://doi.org/10.1511/2009.79.310

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1396*, 5–18. https://doi.org/10.1111/nyas.13325

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341).

Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B. N., deMayo, B. E., Long, B., Yoon, E. J., & Frank, M. C. (n.d.). Analytic reproducibility in articles receiving open data badges at the journal psychological science: An observational study. *Royal Society Open Science*, *8*(1), 201494. https://doi.org/10.1098/rsos.201494

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448. https://doi.org/10.1098/rsos.180448

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3). https://doi.org/10.1017/S0140525X0999152X

Herndon, T., Ash, M., & Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, *38*(2), 257–279. https://doi.org/10.1093/cje/bet075

Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U., & Boulesteix, A.-L. (n.d.). The multiplicity of analysis strategies jeopardizes replicability: Lessons learned across disciplines. *Royal Society Open Science*, *8*(4), 201925. https://doi.org/10.1098/rsos.201925

Ioannidis, J. P. a. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), 0696–0701. https://doi.org/10.1371/journal.pmed.0020124

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*(5), 1–15. https://doi.org/10.1371/journal.pbio.1002456

Klau, S., Hoffmann, S., Patel, C. J., Ioannidis, J. P., & Boulesteix, A.-L. (2021). Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework. *International Journal of Epidemiology*, *50*(1), 266–278. https://doi.org/10.1093/ije/dyaa164

Klau, S., Schönbrodt, F., Patel, C. J., Ioannidis, J., Boulesteix, A.-L., & Hoffmann, S. (n.d.). *Comparing the vibration of effects due to model, data pre-processing and sampling uncertainty on a large data set in personality psychology*. https://doi.org/10.31234/osf.io/c7v8b

Kosslyn, S. M., & Rosenberg, R. S. (2005). *Fundamentals of psychology: The brain, the person, the world, 2nd ed.* Pearson Education New Zealand.

Kuhn, T. S. (1970). *The structure of scientific revolutions* ([2d ed., enl). University of Chicago Press.

Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., Gronau, Q. F., Ly, A., Bergh, D. van den, Marsman, M., Derks, K., Wagenmakers, E.-J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., … Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451–479. https://doi.org/10.1037/bul0000220

Laurinavichyute, A., Yadav, H., & Vasishth, S. (2022). Share the code, not just the data: A case study of the reproducibility of articles published in the Journal of Memory and Language under the open data policy. *Journal of Memory and Language*, *125*, 104332. https://doi.org/10.1016/j.jml.2022.104332

Lubega, N., Anderson, A., & Nelson, N. (n.d.). *Experience of irreproducibility as a risk factor for poor mental health in biomedical science doctoral students: A survey and interview-based study*. https://doi.org/10.31222/osf.io/h37kw

Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., Reddick, W. E., Glass, J. O., Chen, D. Q., Feng, Y., Gao, C., Wu, Y., Ma, J., He, R., Li, Q., … Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, *8*(1), 1349. https://doi.org/10.1038/s41467-017-01285-x

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115.

Meehl, P. E. (1978). *Theoretical risks and tabular asterisks: Sir karl, sir ronald, and the slow progress of soft psychology*. *46*(September 1976), 806–834.

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*. https://doi.org/10.1016/j.jml.2020.104092

Minocher, R., Atmaca, S., Bavero, C., McElreath, R., & Beheim, B. (n.d.). Estimating the

reproducibility of social learning research published between 1955 and 2018. *Royal Society Open Science*, *8*(9), 210450. https://doi.org/10.1098/rsos.210450

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van?t Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2020). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 229–237. https://doi.org/10.1177/2515245920918872

Parsons, S. (n.d.). *Exploring reliability heterogeneity with multiverse analyses: Data processing decisions unpredictably influence measurement reliability*. https://doi.org/10.31234/osf.io/y6tcz

Pashler, H., & Harris, C. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531–536. https://doi.org/10.1177/1745691612463401

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530. https://doi.org/10.1177/1745691612465253

Patel, C. J., Burford, B., & Ioannidis, J. P. A. (2015). Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *Journal of Clinical Epidemiology*, *68*(9), 1046–1058. https://doi.org/10.1016/j.jclinepi.2015.05.029

Poline, J.-B., Strother, S. C., Dehaene-Lambertz, G., Egan, G. F., & Lancaster, J. L. (2006). Motivation and synthesis of the FIAC experiment: Reproducibility of fMRI results across expert analyses. *Human Brain Mapping*, *27*(5), 351–359. https://doi.org/10.1002/hbm.20268

Ricketts, J., Dawson, N., & Davies, R. (2021). The hidden depths of new word knowledge: Using graded measures of orthographic and semantic learning to measure vocabulary acquisition. *Learning and Instruction*, *74*, 101468. https://doi.org/10.1016/j.learninstruc.2

021.101468

Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLoS Biology*, *13*(11), 1–12. https://doi.org/10.1371/journal.pbio.1002295

Rodríguez-Ferreiro, J., Aguilera, M., & Davies, R. (2020). Semantic priming and schizotypal personality: reassessing the link between thought disorder and enhanced spreading of semantic activation. *PeerJ*, *8*, e9511. https://doi.org/10.7717/peerj.9511

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., … McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, *117*(15), 8398–8403. https://doi.org/10.1073/pnas.1915006117

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development*, *31*(1), e2295. https://doi.org/10.1002/icd.2295

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, *16*(4), 744–755. https://doi.org/10.1177/1745691620966795

Schweinsberg, M., Feldman, M., Staub, N., Akker, O. R. van den, Aert, R. C. M. van, Assen, M. A. L. M. van, Liu, Y., Althoff, T., Heer, J., Kale, A., Mohamed, Z., Amireh, H., Venkatesh Prasad, V., Bernstein, A., Robinson, E., Snellman, K., Amy Sommer, S., Otner, S. M. G., Robinson, D., … Luis Uhlmann, E. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes*, *165*, 228–249. https://doi.org/10.1016/j.obhdp.2021.02.003

Sedlmeier, P., & Gigerenzer, G. (1989). Statistical power studies. *Psychological Bulletin*, *105*(2), 309–316.

Silberzahn, R., & Uhlmann, E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature*, *526*(7572), 189–191. https://doi.org/10.1038/526189a

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M., Dalla Rosa, A., Dam, L., Evans, M., Flores Cervantes, I., … Nosek, B. (2017). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. *Advances in Methods and Practices in Psychological Science*. https://doi.org/10.31234/osf.io/qkwst

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011b). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011a). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Starns, J. J., Cataldo, A. M., Rotello, C. M., Annis, J., Aschenbrenner, A., Bröder, A., Cox, G., Criss, A., Curl, R. A., Dobbins, I. G., Dunn, J., Enam, T., Evans, N. J., Farrell, S.,

Fraundorf, S. H., Gronlund, S. D., Heathcote, A., Heck, D. W., Hicks, J. L., … Wilson, J. (2019). Assessing Theoretical Conclusions With Blinded Inference to Investigate a Potential Inference Crisis. *Advances in Methods and Practices in Psychological Science*, *2*(4), 335–349. https://doi.org/10.1177/2515245919869583

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016a). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016b). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712.

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, *8*(1), 192. https://doi.org/10.1038/s41597-021-00981-0

Towse, J. N., Ellis, D. A., & Towse, A. S. (2021). Opening Pandora's Box: Peeking inside Psychology's data sharing practices, and seven recommendations for change. *Behavior Research Methods*, *53*(4), 1455–1468. https://doi.org/10.3758/s13428-020-01486-1

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*, 34–80.

Vankov, I., Bowers, J., & Munafò, M. R. (2014). On the persistence of low power in psychological science. *Quarterly Journal of Experimental Psychology*, *67*(5), 1037–1040. https://doi.org/10.1080/17470218.2014.885986

Vasishth, S., & Gelman, A. (2021). How to embrace variation and accept uncertainty in linguistic and psycholinguistic data analysis. *Linguistics*, *59*(5), 1311–1342. https://doi.org/10.1515/ling-2019-0051

Vazire, S. (2018). Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Wagenmakers, E.-J., Sarafoglou, A., & Aczel, B. (2022). One statistical analysis must not rule them all. *Nature*, *605*(7910), 423–425. https://doi.org/10.1038/d41586-022-01332-8

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Maas, H. L. J. van der. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on bem (2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Wessel, I., Albers, C., Zandstra, A. R. E., & Heininga, V. E. (2020). *A multiverse analysis of early attempts to replicate memory suppression with the think/no-think task.*

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Wild, H., Kyröläinen, A.-J., & Kuperman, V. (2022). How representative are student convenience samples? A study of literacy and numeracy skills in 32 countries. *PLOS ONE*, *17*(7), e0271191. https://doi.org/10.1371/journal.pone.0271191

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. https://doi.org/10.1017/S0140525X20001685

Young, C. (2018). Model uncertainty and the crisis in science. *Socius*, *4*, 2378023117737206.

Young, C., & Holsteen, K. (2017). Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis. *Sociological Methods & Research*, *46*(1), 3–40. https://doi.org/10.1177/0049124115610347