

This is a section of [doi:10.7551/mitpress/12200.001.0001](https://doi.org/10.7551/mitpress/12200.001.0001)

# The Open Handbook of Linguistic Data Management

**Edited by:** Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller, Lauren B. Collister

## **Citation:**

*The Open Handbook of Linguistic Data Management*

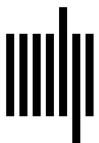
**Edited by:** Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller, Lauren B. Collister

**DOI:** 10.7551/mitpress/12200.001.0001

**ISBN (electronic):** 9780262366076

**Publisher:** The MIT Press

**Published:** 2022



The MIT Press

## 2 Situating Linguistics in the Social Science Data Movement

Lauren Gawne and Suzy Styles

### 1 Introduction

Linguists spend a lot of time working with data, but we do not always give much thought to the role that data plays in building the larger research culture in our field. We can learn a lot about good data management in our own discipline by learning from what is happening in related fields, both in terms of innovations and new benchmarks, as well as when things have not gone right. We look at how data have been conceptualized and managed in other areas of the social sciences, particularly social psychology, and how current attitudes are shaping the future of research. The fundamental theme of this discourse is the centrality of openness, both in terms of transparency of methodology and making primary data more accessible to people beyond the original researchers. This move toward open research aims to reduce biases, both for individual researchers and for the discipline, and encourages more considered data collection and presentation.

We understand that these developments can feel challenging, particularly to researchers who have already established their workflow. Researchers can feel very protective of the data they collected, and sometimes this is for very good reasons. As we will see, the discussions around open research in social psychology have, at times, exacerbated tensions and left some researchers feeling singled out. We believe that linguists can learn from social psychology about new movements toward good open data practice, but we can also learn about building an open research culture that is inclusive and encouraging.

Building an open approach to research is the most ethical way to respect the contribution of the individuals who take the time to participate in your work, by making sure their data contribute as much as possible

to global research outcomes. Making data accessible also brings its own rewards, as openness allows linguists to receive credit for their data collections, increases the likelihood of subsequent research on the original data, and provides incentives via increased citations.

In this chapter, we begin by positioning linguistics within the social sciences, to make a clearer case for what lessons other areas of the discipline can offer us (section 2). We then look at research replicability and reproducibility, particularly with regard to how they are conceptualized in social psychology (section 3), and at the data crisis that has dominated methodological discussions in social psychology, focusing on stories that highlight key issues when research does not have an open foundation (section 4). We then address how these lessons from social psychology apply to our own research through answering a series of frequently raised concerns about open data (section 5).

Many of the issues discussed here are by no means unique to social psychology. We have decided to focus on this area for three reasons. The first is that that data social psychologists work with is of similar complexity to that in linguistics, in that it is data about human behavior that is often nuanced and context-dependent. The second is that the public discussions being had by social psychologists, and the changes in research behavior being implemented, are further along than discussions and changes in other fields of social science. The third reason is that coauthor Suzy Styles works across linguistics and social psychology and is in a good position to “translate” the lessons of social psychology for linguists.

We have also seen discussions of issues regarding a lack of transparent research in other social science subfields including education (Makel & Plucker 2014), sports psychology (Schweizer & Furley 2016), and management (Goldfarb & King 2016; Bergh et al. 2017); as

well as scientific fields including medicine (Goldacre 2010), neuroscience (Grabitz et al. 2018), and behavioral ecology (Jennions & Møller 2003); as well as in the broader literature on the scientific method (Baker 2016; Ioannidis 2005, 2012; Zwaan et al. 2018). The social sciences are in a unique position to consider these issues, given that our work inherently focuses on the nature of human behavior (Bollen et al. 2015).

We are writing this chapter together as two researchers who both care about the thoughtful use of data in linguistic research, but come from different research traditions and subfields of linguistics. Lauren Gawne's research is in the language documentation and description model, mostly focusing on Tibet-Burman languages of Nepal. Suzy Styles is a psycholinguist, whose work mostly focuses on cross-sensory processing, particularly in infants and children. We have also worked collaboratively on experimental research investigating the cross-sensory processing of lexical tone for speakers of Syuba and English.

## 2 Linguistics as a social science

Linguists use data to try to understand the nature of language and how it is used. The next chapter in this Handbook will give an illustration of the breadth of data that is included in linguistics. This heterogeneity is a strength of the discipline, but it also means that there is greater imperative to think critically about the role of data within our work.

Loosely defined, the social sciences are those academic disciplines interested in the way people relate to each other and that use empirical data to address their research questions. Linguistics can also be conceived of as a discipline within the humanities, in its exploration of the nature of human society and culture, or as a discipline straddling the humanities and social sciences. In this chapter we frame linguistics as a social science, because it is worth thinking critically about the nature of the materials on which we base our argumentation about language. Placing linguistics within the paradigm of social sciences centers this data and allows us to consider parallels with other disciplines within the social sciences. The social sciences, such as psychology and linguistics, differ from the "hard sciences" in the nature of the data we collect. Because the social sciences deal with people, the data are typically noisy, complex, error-prone, and inconsistent—just like the people they

represent. This makes the task of analyzing human data notoriously complex (Lykken 1991:4). One way to manage this complexity of data is to be transparent about it. This has been happening in the social sciences. Likewise, the field of linguistics is increasingly engaged in a conversation about the nature of linguistic data and how it is managed. Indeed, this Handbook is a testament to how much ground there is to cover in this discussion.

Questions about the transparency of linguistic data have been raised since at least the 1990s: Thomason voiced concerns about the frequency of erroneous data in publications in an editorial for *Language* (1994), and around the same time, Himmelmann (1998) argued that language documentation and description can only be successful if the claims made about a particular language can be verified by others—that is, if readers are allowed access to the data from which conclusions are drawn. The core argument made by both authors was that research is more reliable if the methods and the underlying data are transparent.

Despite the clarity of these arguments twenty years ago, progress in linguistics has been limited. Berez-Kroeker et al. (2017) surveyed 270 articles from nine linguistics journals published between 2003 and 2012 and found that 60% of articles published did not clearly report the source of the underlying data. Of the 40% that did clarify their data sources, 25% relied on already published data, leaving fewer than 15% of articles disclosing the source of primary data. With such a low rate of data citation, it is hard for the readers of a linguistics paper to establish for themselves (a) whether examples cited in the main text contain errors; (b) whether examples cited in the text are representative of the data set as a whole; and (c) whether the sampling conditions of the original data set qualify or constrain their interpretation. Without the ability to check original data sources, it is therefore unclear how a reader should interpret the conclusions drawn by the original author.

With a growing awareness of the importance of data to the transparency of linguistic analysis, a US National Science Foundation project was set up to bring together over forty linguists from across the world to think about the future of linguistic data. The outcome of three years of these meetings was the paper "Reproducible Research in Linguistics: A Position Statement on Data Citation and Attribution in Our Field" (Berez-Kroeker et al. 2018). This position statement argues:

Linguistic data are the very building blocks of our field. Given that linguistic theories need to be borne out through data, we believe that linguistic data are important resources in their own right and represent valuable assets for the field. Therefore, our field needs to accept responsibility for the proper documentation, preservation, attribution, and citation of these assets. (11)

While linguists are beginning to discuss how we manage and share our research and data, this conversation is further along in other fields. The open access movement, which began slowly in the late 1990s, has changed the publishing landscape with open access journals (Joseph 2013) and now has its sights on open data practice (Kitchin 2014; SPARC, n.d.). See Collister (chapter 9, this volume) for more discussion about the legal issues regarding open data. Open science now includes a broad range of practices, including open data infrastructure, accessibility of data, impact of research, and collaboration (Fecher & Friesike 2014).

Along with open data has come an important conversation about ethical use of data. In social psychology, the argument can be made that participants' efforts are not fully respected if transparency around data and analysis is lacking. While we do not talk about ethical issues very directly in this chapter, we acknowledge this conversation always needs to be kept in mind and find a good basic starting point to be the European Commission's

H2020 data management plan guidelines that data should be "as open as possible, as closed as necessary" (H2020 Programme 2016). For a more detailed discussion of ethics and data in linguistics see Holton, Leonard, and Pulsifer (chapter 4, this volume).

### 3 Replicability, reproducibility, and lucky-cowboy research

Here we revisit *reproducibility* and *replicability* (see Berez-Kroeker et al., chapter 1, this volume) and examine the differences between them: Reproducibility is the use of existing materials and methods to check whether the original conclusions are supported by the evidence, and replicability is the application of the original methodology to a new sample. This allows researchers to check whether the conclusions of the original paper hold true for a different set of data or a different group of people. In other words, is the result generalizable to people who were not included in the original study?

To borrow an analogy introduced by Styles (2018), imagine we asked one hundred cowboys to flip a coin one hundred times each (figure 2.1). The coin is normal, evenly weighted, and has a picture of a head on only one side. It's pretty clear that if none of the cowboys are cheating, the typical response will be fifty heads. If we



Figure 2.1

Coin-flipping cowboys.

Source: Illustration by Dannii Yarbrough.

asked all of the cowboys to line up in a barn according to how many heads they got, we'd get a big bunch of cowboys in the middle of the barn (the 50% mark) and fewer and fewer cowboys as they got near the walls. Whenever we sample from human populations, the people we sample could be anywhere along this line of cowboys. They could be typical cowboys near the middle of the barn, or they might be outliers at the fringes of the barn—lucky cowboys—whose experience of coin flipping is quite unusual. In linguistics, the same logic can be applied to the data upon which inferences are drawn—whether it is sentences selected from a corpus of closed-captioned news reports or acoustic analysis of whether different radio show hosts speak with vocal fry. The sentences selected for analysis might be more or less representative of the corpus, and the voices selected for comparison might be more or less representative of radio hosts, or of people in general.

In the social sciences, where conclusions are made on the basis of data, both replicability and reproducibility are important, because they allow us to independently verify the claims made.

### 3.1 Replicability

Replicating a study is the process of checking whether the data are representative in general (are the cowboys in this barn similar to cowboys in other barns?). Although replication may be seen as a fundamental tool of the research enterprise, it does not occur particularly frequently in the social sciences. Makel, Plucker, and Hegarty (2012) looked at articles published in one hundred journals between 1900 and 2012 and found that only around 1% were replications. Central to this dispreference for replication studies is a focus on novelty of results (Fanelli 2011; Ferguson & Heene 2012). This has driven an interest in *conceptual replications*, which test hypotheses from earlier studies with a different methodological setup (e.g., with a different cohort of participants or different stimuli) (Schmidt 2009:96). In social psychology it is not uncommon for a single paper to include several experiments that are conceptual replications of a core study, with the aim of demonstrating the robustness of the underlying principle. Although we might think that several replications with small modifications would lead to higher likelihood of direct replication, Kunert (2016) demonstrated that papers originally published with multiple conceptual replications did not replicate more frequently

or reliably than other studies. In section 4.1, we discuss the “file drawer problem” as one possible reason for this.

### 3.2 Reproducibility

Reproducing a study is the process of checking whether mistakes were made (are the cowboys in the barn standing in the right spot?) and whether the data selected for analysis are representative of the complete data set (were the cowboys representative of everybody in the barn?).

Ideally, research should be both reproducible and replicable, but documenting human behavior, particularly context-specific behavior outside of an experimental setting, can make replication unfeasible. Reproducibility is particularly useful for contexts where even having access to the original methodology makes it unlikely you will be able to replicate the original research. Reproducibility requires that the data are made available. This is not always easy. Wicherts et al. (2006) contacted authors of 141 papers published by the American Psychological Association in 2004, requesting access to their data sets for reproduction of the results, as part of a meta-analysis of outliers in research findings. Although the American Psychological Association has a policy all authors sign stating that they will share data for reproduction, Wicherts et al. only received data sets from 27% of authors. Reproducibility needs to be built on proactive sharing of data, as it is demonstrably difficult to obtain data after a study is published.

Although the terminology of replicating and reproducing studies is not commonly used in linguistics, table 2.1 contains examples of each of these types of research, applied to two famous linguistics studies.

In these examples, reproducing a study allowed errors to be detected and corrected, leading to a more detailed data set in the public domain, and replicating a study led to greater clarity in the sociodemographic characterization of New York speech, along with documenting shifts over time. It is important to note that these kinds of follow-ups can only be performed if (a) the source of the data is clearly described; (b) the methods for acquiring the data are clearly described; (c) the original data sources are accessible; (d) details about selection/inclusion/exclusion are clearly described; and (e) statistical or measurement procedures are clearly described. As such, transparent methods and data citation are central to reproducing and replicating research.

Table 2.1

Examples of reproducing and replicating studies in linguistics

	Reproducing a study	Replicating a study
Original study	Peterson and Barney's American English Vowels (1952)	Labov's Fourth Floor Study ([1966] 2006)
What was done in the original study?	Peterson and Barney recorded the voices of 76 speakers of American English reading from standardized word lists. Each speaker read each word twice. Speakers included adult men, adult women, and children (both male and female). They plotted the F1 against the F2 of individual vowels to show the distribution of formants among speakers of the same language—including speakers with different sized vocal tracts.	Labov visited three large department stores in Manhattan, known to be used by people in different income brackets. He asked shop staff to direct him to a department he previously established was on the fourth floor. He pretended not to hear and asked them to repeat. He noted down the number of times each person used a rhotic pronunciation for the /r/ at the end of each word and compared the rate of rhotic use across the three stores. Labov found that staff at the three stores differed in how often they used rhotic "r". The most expensive store used "r" (the prestige variety) the most often, and the least expensive store used the "r" the least.
Follow-ups	Watrous (1991); Boersma and Weenink (2013); Barreda (2016)	Fowler (1986); Mather (2011)
What did they do? What was different?	Two published versions of the key figure contained discrepancies. Watrous (1991) requested original data, which was shared by Mattingly. A corrected figure was produced and published alongside complete data tables for F0, F1, F2, and F3. Subsequently, data tables were integrated into Praat software by Boersma and Weenink (2013) for graphing. Barreda (2016) included the data tables in a graphing tool in the R package phonTools.	Different linguists returned to the same stores in Manhattan twenty and forty years after the original study. Labov's rapid anonymous survey technique was followed precisely. Between studies, the lowest prestige store closed down, so a substitute store servicing the same socioeconomic status (SES) was chosen in each of the replications. Additional demographic characteristics were recorded in the replication samples.
Were the follow-ups successful?	A more accurate version of the figure was produced. Subsequent researchers can create their own figures from the original data sets using two open source software systems or adapt the graphing code for new purposes.	The same general pattern of rhotic use was found in each replication, with more rhotic "r" in the high-prestige store than in the others. The two replications showed higher rates of rhotic use overall, with steady increase between studies.
What was learned?	<ol style="list-style-type: none"> <li>1. Errors in original publication were minor and did not undermine the core claims made.</li> <li>2. Values in the follow-up paper are graphed correctly.</li> <li>3. Full data set was published.</li> <li>4. Inclusion of F0 and F3 along with F1 and F2 data allow future users to perform novel analyses/graphing.</li> </ol>	<ol style="list-style-type: none"> <li>1. Labov's method is reliable for eliciting SES differences</li> <li>2. Labov's sample was representative of general patterns in New York, and those patterns persist—albeit with changes.</li> <li>3. New York has undergone a gradual shift toward the prestige form over time.</li> </ol>
What was required for this follow-up?	<ol style="list-style-type: none"> <li>1. Clear descriptions of original data source</li> <li>2. Access to original data (numerical)</li> <li>3. Clear descriptions of inclusion/exclusion process</li> <li>4. Clear descriptions of the F1/F2 measurement process</li> <li>5. Clear descriptions of computations/analyses</li> </ol>	<ol style="list-style-type: none"> <li>1. Clear descriptions of the stores and their SESs</li> <li>2. Clear descriptions of the elicitation method</li> <li>3. Clear descriptions of inclusion/exclusion process</li> <li>4. Clear descriptions of counts and statistical analysis</li> <li>5. Clear descriptions of the participant demographics</li> </ol>

We do not want to imply that a lack of openness about methods means the authors are intentionally hiding anything. Brown et al. (2014) found that only around 30% of decisions about data sampling were clearly articulated in a sample of over 1,000 research publications. People are doing good research, but the whole field has to value openness to normalize best practice in data management. It is also worth stating here that openness is not a binary state. It is possible to have very transparent methods, but only share summary data, and conversely, it is possible to share all the primary data generated by research, but provide insufficient detail regarding how those data were collected or how measurements were performed.

Some work focuses on the reproducibility of the methods of a study or the reproducibility of the results. Goodman, Fanelli, and Ioannidis (2016) also discuss *inference reproducibility*, which is perhaps even more elusive. Different researchers can draw different inferences from the results of a single study, meaning it is possible for someone to replicate both the method and results of an original study, but still not agree to replicate the inference of what those results mean.

#### 4 The data crisis in social psychology

In 2011 a paper emerged that shook the foundations of research methods in social psychology and carries a lesson for all of the social sciences. A paper by Bem (2011) appeared to show evidence of precognition—a psychic power that allowed people’s decisions to be influenced by a stimulus that was shown to them *after* they had already made their choice. The paper was peer-reviewed, published in a flagship journal, and, to all appearances, conducted using rigorous controls and standard research methods of the day. The data appeared sound, and yet the conclusions they seemed to support couldn’t be physically possible. The paper, and several papers published in its immediate aftermath (LeBel & Peters 2011; Wagenmakers et al. 2011; Galak et al. 2012), indicated that something was wrong with the standard methods that were common throughout the field of psychological research. The lack of transparency in reporting the complete research cycle meant a single spurious result (a lucky cowboy) could be presented in isolation from a mountain of contradictory evidence (all of the other

cowboys in the barn). In other words, the result that was selected for publication was not representative of all of the tests that were conducted.

In the wake of this paper, some theorists argued that biases in the existing literature mean we should expect the replication process to call the reputation of all research in the field into question (Ioannidis 2012; Johnson et al. 2017). This has come to be known as the replication crisis (Ioannidis 2012; Loken & Gelman 2017; Zwaan et al. 2018; although see Nelson, Simmons, and Simonsohn 2018 for the more optimistic framing of this as a “renaissance”). These discussions are the most recent part of a longer conversation about the shortcomings of research methods and transparency in social psychology, since Elms’s discussion about the “crisis of confidence” in social psychology research methods in 1975 (see also Sterling 1959; Cohen 1962; Walster & Cleary 1970; Lykken 1991; as well as Feynman’s famous 1974 denouncement of social psychology as “cargo cult science”). For a more detailed, but still very accessible, discussion of the recent period in social psychology, see Chambers (2017) and Nelson, Simmons, and Simonsohn (2018).

Next, we review some of the major themes in the ongoing conversation about data in social psychology. We chose these topics because we believe they are of particular value to linguists. The first is what is known as the file drawer problem (Rosenthal 1979), with unpublished data creating an untold story of the field. The second is selective interpretation of results. In both of these sections we draw parallels for linguists about how we work with data, but we also draw lessons on how to move the field forward in the most positive way. A lot of the discussion around methodology in the field of social psychology has led to cultural clashes and resentment as people feel that their research integrity is called into question (Meyer & Chabris 2014; Fetterman & Sassenberg 2015).

We focus on scenarios where researchers believe themselves to be working in the best interest of advancing research, rather than egregious misuse of data and outright fraud (but see Simonsohn 2013; Chambers 2017: chapter 5 for examples of this). Research transparency can also help reduce cases of fraud, as well as help well-intentioned researchers find errors in their own data (Nuijten et al. 2016).

#### 4.1 File drawer problem and the shoebox of tapes

In the 1970s Rosenthal coined a phrase for the plethora of psychology studies that are conducted but never published: the file drawer problem (Rosenthal 1979). At the time, he was interested in whether statistical methods can be used to estimate (and potentially correct for) gaps in the published literature. But where do these gaps come from? Let's say 20 researchers in different psychology departments conduct an identical study on whether women and men differ on some arbitrary variable—say, time to respond to e-mails. If, simply by chance, one of the researchers finds a significant difference, this constitutes a novel, positive finding, and the researcher might successfully publish that result. The other 19 researchers are simply less likely to publish. The failure to publish is related to explicit biases in high-profile publication outlets where results showing no difference are less desirable than novel findings of significant difference. What this means for the scientific record is that a single published study can represent an outlier (a lucky cowboy) from the total body of research that was conducted, and the scientific literature can become skewed by the absence of alternative accounts.

Historically, page and data storage limits in high-ranking publication outlets have meant that many details of the research cycle had to be omitted from the primary publication. Primary data is one of the parts of the research cycle that is not included in a traditional journal article. When it comes to the types of articles that are published, there is a long tradition of preference for novelty and bias against negative results in publication, from editors (Neuliep & Crandall 1990) and reviewers (Neuliep & Crandall 1993). Researchers are selective in their publication outputs: researchers are less likely to write up research with data that has a null result than research with a strong result (Franco, Malhotra, & Simonovits 2014), or underreport non-significant conditions in multipart research processes (Franco, Malhotra, & Simonovits 2016). In addition, the rise of competitive funding as a metric for success has also left little room for non-novel results and replications (Lilienfeld 2017). Selective reporting can therefore be understood as rational response to a biased incentive system, and valuing null results and replications (Koole & Lakens 2012) can ensure that large volumes of potentially valuable data may see the light of day.

Recent large-scale replication projects have demonstrated that when the literature is biased in this way, less than half of published psychology papers present effects that can be replicated (Open Science Collaboration 2015; Klein et al. 2018), which weakens public confidence in the credibility of social psychology, affecting the entire field. Although the final effect is the nature of what is published, the only way to address this problem is to change research priorities and publication practices including the management and publication of data. By ensuring research has greater transparency for the entire research cycle—from hypothesis generation to methods and materials, through to publication of full and complete data sets—we can make sure that research does not languish in the file drawer.

Many linguists will be familiar with the idea that it is easier to publish work with a certain kind of theoretical “hook” or empirical novelty, and this may bias researchers to focus on linguistic features deemed more “exotic,” or languages that were previously undocumented. Publication bias aside, linguists, like other social scientists, often collect far more data than they include in their published articles, be they recorded audio, transcriptions, a collection of sentences extracted from a corpus. The average linguistics paper or grammar contains only a summary of the rich original data—this is, after all the purpose of research, to synthesize information so that insights can be drawn about the nature of language, or the way people use it. However, each data source contains vastly more information than the narrow range of features it was originally evaluated for, and these details may be of interest to generations of future linguists. As the majority of papers fail to disclose the source of their primary data, and those that do rarely make the primary data available, this represents a similar problem to the file drawer of the social sciences—although perhaps better characterized as the shoebox of tapes on the shelf.

In response to these changing priorities, novel solutions have emerged for how the outputs of research can be archived by researchers themselves. Digital archives such as the Dataverse family of institutional repositories allow researchers to upload a variety of digital file formats and create their own metadata records. This means that researchers can archive digital materials from digital documents, to video/audio files, raw text, or numeric data. Free-to-use repositories such as GitHub allow researchers



to share the code used to run or analyze their studies. Many of the new generation of repositories are digital object identifier (DOI) granting (meaning that digital documents have a unique digital identifier, and can be cited) with clear time-stamping and transparent version control. Some research archiving hubs, such as the Open Science Framework (OSF) allow in-platform browsing of content, multiuser functionality, and integrated links to other archives.

This generation of user-driven archiving solutions allows researchers to collate content along the entire research pipeline and to control access permissions at different stages of the research cycle. With contemporary tools such as these at their disposal, researchers can clear the content from the file drawer more effectively than ever before, even for data sets that are not reflected in traditional, peer-reviewed journal articles. This means that researchers can get credit for the work of creating the data, where in a traditional model of novelty, it may not have ever been made available.

To give an example from our own research, we used OSF to create an archive of project materials for a study investigating links between the senses in groups of people who use different languages (Styles & Gawne 2017). Despite being based in different countries for the majority of the project, the multiuser interface allowed both authors to upload elements to the archive, create metadata descriptions of the contents, and collate the collection in real time. We archived our test materials during the planning phase of the project. These materials included audio files prepared in the lab, as well as audio files generated in the field, and photographs of the original stimuli used in the study. After collecting data, we uploaded data tables for the individual responses to our task, anonymized to protect the identity of individuals. When we submitted our first manuscript for peer review, we were able to create a private link for anonymous peer review, ensuring that our reviewers could access all of our materials and data during the review process. When our first article was accepted for publication, we updated the status of the repository to “public,” and the links that appear in the article allow any reader to access the full set of data and materials. Our OSF repository<sup>1</sup> also contains a second data set that is currently private while we finalize the manuscript for submission, as well as the materials, analysis plan, and a preregistration of hypotheses for a third study that is currently in progress. As

this example demonstrates, self-archiving can support a transparent research process, with full data available for review. Furthermore, the OSF system allows “blind” review links where the data we present are available for checking within the existing double-blind peer-review system (reproducing results, or conducting alternative analysis to see whether the results hold under different circumstances). An added bonus of this workflow is that our research outputs are deposited in such a way that if one of the studies does not result in a traditional publication, the data become research outputs in their own right and will remain public and accessible: We are effectively clearing out the file drawer as we go along. Furthermore, as we create these archives, the DOI-generating repositories allow our documents to be uniquely identified so that future researchers can conduct novel research using our tools, materials, or data, and their engagement in our outputs will be reflected by straightforward citations that link back to the data source.

While not all projects in linguistics will benefit from drawing on all of these features, the development of these tools for social sciences research has broad implications for the field. Some fields of linguistics have embraced broad and accessible data sharing and archiving platforms, including child language acquisition via CHILDES, the Child Language Data Exchange System (MacWhinney 2000); corpus linguistics, including tools such as the International Corpus of English,<sup>2</sup> and researchers in the field of language documentation and description (see Salfner 2015 and Caballero 2017 as examples, although Gawne et al. 2017 and Thieberger 2017 note inconsistency in this field).

Because archiving your data is a kind of publication, this kind of digital filing allows you to acknowledge all of the work that you do. Many researchers in the social sciences have multiple sections in their curriculum vitae to address all of their different outputs, including open data sets, and preprints.

Alongside these developments we also need to advocate for structural solutions to the historic biases that have arisen alongside traditional publication formats. As a community of practice, we can also reward people who value a more transparent and open research environment by citing their outputs and acknowledging the important work they do. In 2018, the Linguistic Society of America adopted its “Statement on Evaluation of Language Documentation for Hiring, Tenure, and Promotion,”<sup>3</sup> which

provides a good example of valuing data, although we would argue that it is relevant to all subfields, not just language documentation. When hiring, we can create job descriptions that explicitly value open data principles; we can encourage junior researchers to include more publication types in their curriculum vitae; and if we truly come to value these practices, we can reward those with the best practice in open data with the jobs and promotions that their commitment to evidence deserves.

Get your data out of the file drawer and publish it—your data deserve it, and so do you. Where possible, make these collections open, or at least discoverable to other researchers. Using one of the new digital archiving platforms with DOIs or other stable citation practices means that you can be credited for the work of building the original data. The easier part of the journey is to build archiving into your current and future research.

#### 4.2 Selective interpretation and how we choose data for analysis

In 2016, Brian Wansink, then the head of Cornell's Food and Brand Lab, blogged about "The Grad Student Who Never Said No" in which he commented on the difference between two junior researchers in his team. He had offered them both the chance to work on "a data set from a self-funded, failed study which had null results," claiming, "there's got to be something we can salvage because it's a cool (rich & unique) data set." He praised the doctoral student who agreed to work with the data for her ability to "make hay while the sun shines" (producing five papers in the course of a six-month stint in the lab), while dismissing the attitude of the postdoc who refused, produced a smaller number of papers, and left academia. Early commenters on the blog generously assumed the story was a work of academic satire, a cruel parable designed to highlight a broken reward system in academia. Wansink clarified his position as a serious commentary, leading other researchers to query four "buffet" study papers from the data set (Just, Siğirci, & Wansink 2014, 2015; Siğirci & Wansink 2015; Kniffin, Siğirci, & Wansink 2016), as well as other work from Wansink (see van der Zee, Anaya, & Brown 2017). Because the data looked at consumption of pizza, this scandal is sometimes affectionately known as "Pizza-gate." Cornell's Food and Brand Lab specialized in consumer behavior as a subdiscipline of social psychology. The discussion that has come out of this controversy has touched on practices common across social psychology.

For those familiar with how statistical analyses can be manipulated to provide more desirable outcomes, Wansink's blog post was not about a diligent student making the best of a situation, but a demonstration of data manipulation. Wansink encouraged cherry-picking, or using only selected subsets of data that skew toward the preferred outcome, as well as *p*-hacking and HARKing (or hypothesizing after results are known). The term *p-hacking* refers to running statistics over a variety of subsets of the data until the desired statistical significance is achieved. Statistical significance is a desired aim in the quest for novelty in quantitative research. It is unclear how common this practice is, but meta-analysis suggests it is likely to be common (Head et al. 2015). HARKing is the practice of retroactively coming up with a hypothesis once a test with statistical significance has been identified (Murphy & Aguinis 2017), for example, stating that the hypothesis was about how much pizza women ate in a restaurant in particular conditions, when the original data included both men and women. These practices are known, but their prevalence is hard to quantify without transparent and open research practices. The difference is that Wansink talked about this publicly.

This has led to further unravelling of Wansink's work in other studies, including at least 17 retractions (Retraction Watch 2018), charges of research misconduct, and eventual dismissal. Some say that Wansink has become a scapegoat for researchers trained in an era when these kinds of research practices were more common. The real problem was that he did not recognize the errors of his research practices and appeared to actively obstruct researchers who were interested in actively reproducing his work to find potential corrections for the published record. Although the case has revealed academic misconduct, the case was exacerbated by problems in record keeping. If Wansink had kept better records about how the data had been handled in each of his papers, further investigations may not have been necessary.

The literature on these data inflation practices (not just Wansink's) are consistent in their recommendations as to how to avoid these practices. Central is the need to educate researchers on good data practice. One key feature of the replication crisis is that good data practice can be difficult if the research environment rewards bad behavior (Guest 2016). Other recommendations include ways to limit the temptation to inflate results for novel outcomes and instead reward good data management. The first of

these is giving researchers credit for publishing their data and methods. The second is to give greater value to replication. We have discussed both of these herein. The third is preregistering methods and hypotheses, to prevent the temptation to diverge from the original course of study.

*Preregistration* is the submission of the methods and tools that are intended to be used, before the data collection is carried out (Mellor & Nosek 2018). This limits “researcher degrees of freedom,” by requiring the researcher to commit to a plan of analysis before the work begins (Wicherts et al. 2016). There is still scope for exploratory research, as Nelson, Simmons, and Simonsohn (2018:519) note, “preregistrations do not tie researchers’ hands, but merely uncover readers’ eyes.” Preregistration services such as AsPredicted allow researchers to formalize and lock in hypotheses and/or analysis plans before beginning data collection. Preregistration can be lodged using online systems that freeze and date-stamp the submission, and these can be placed on embargo until the authors have collected the data and are ready to publish. We have used the preregistration tool available through OSF to register our hypothesis and methods for research currently being conducted.

The use of preregistration website services is useful, but some researchers are attempting to bring the practice into the journal publishing process itself, with registered reports (Nosek & Lakens 2014). A registered report contains the hypothesis and methods for a research project and is submitted to a journal and undergoes peer review before the data are collected. If the report is successful in the review process, the journal agrees in principle to publish the final paper, regardless of whether the results are statistically significant. The preregistration of methods prevents researchers from redirecting their research, while also reducing the likelihood of papers with non-significant results falling victim to the file drawer (Chambers 2013). Negative results are as valuable as positive ones, as they move us toward a more complete understanding of the phenomena being studied (Matosin et al. 2014). While preregistration is particularly useful in quantitative research, it can also be used to articulate the intended scope and limits of the data and methods in qualitative investigations before they commence. It can also ensure that researchers who make their data open are not “scooped” in their analysis, as there is a time-stamped public record of the intended use of the data.

More transparent presentation of methods and data leads to more opportunity for replication and reproduction of research. Normalizing this practice can help neutralize what some, such as Wansink, see as the threat of criticisms of their research agenda. A survey by Fetterman and Sassenberg (2015) found that scientists overestimated the negative effect on reputation from a failed replication, and a researcher’s reputation was more likely to be harmed if they refused to engage with the findings of replications. While replication is a good aim, to really confirm an effect, multiple replications are needed (Maxwell, Lau, & Howard 2015), as well as more meta-analysis (Stanley & Spence 2014), and replication or reproduction studies also need to be transparent about these methodological features as well (Brandt et al. 2014).

There is also a lesson to be learned here about the need to be open to criticism as part of the research process. Ad hominem attacks on people’s character as researchers are never acceptable, but appraisal of data and outcomes, and the methods used to obtain these, should be part of a healthy science of linguistics.

It is important to remember the role you play in interpreting your data for your audience. Being clear about your research methods, both to yourself and your audience, can help mitigate selective interpretation of data. Allowing your readers to access your data can also help ensure that others can also follow your analysis.

## 5 Discussion

We have outlined some of the issues facing social psychology, which are a broader reflection of the issues facing all researchers in the social sciences, including linguistics. We also mentioned some of the solutions that have been proposed to counter these issues, most of which are centered on building a culture of open data and open research. In this section, we discuss three main themes linguists can take away from the social sciences. The first is that there are benefits that openness can bring; the second is that we can make the most of emerging tools and processes to enact openness; and the third is that we need to foster a positive cultural shift both in our own work and the field more generally. Some of these practices have been long established in some subfields of linguistics, and many are also raised in Berez-Kroeker et al. (2018) and throughout this Handbook.

### 5.1 Openness brings benefits

Researchers are rightly attached to the data they collect; conducting primary research takes a lot of our time and creative energy and drives the original contributions that we make to the collective understanding of language. Researchers can become caught in the sunk cost fallacy, where they feel that that they've invested so much in this work that it would be a loss for them to share it. However, another way of thinking about the effort of data collection is to ask whether there are more efficient ways to get returns on the investment you have already done. There are benefits to taking a more open approach. Open data also attracts attention to your specialist field—be it an underdocumented language or an uncommon grammatical construction—and ensures that if an individual researcher moves on from working on a particular set of data (increasingly common in a sector where we train far more PhDs than there are future jobs for), their data can continue to benefit the ongoing research process. There is evidence that research publications with open data attract higher citation rates (astronomy, Henneken & Accomazzi 2011; gene data, Piwowar & Vision 2013; social sciences, Pienta, Alter, & Lyle 2010). Open data also shows the greatest respect to your research participants, because their contributions go further than the bounds of your own research project.

It can feel like open data management involves a great deal more work on top of an already demanding set of research expectations. We can also decide how we wish to reward positive open data practices, in our field, but also in our institutions and professional organizations, particularly with regard to hiring and promotion (including tenure), awards, and research funding. The research sector is changing and beginning to acknowledge the publication of non-traditional outputs such as corpora and data sets, and funding bodies and publishers are encouraging open data, such as the European Research Council's Horizon 2020 Open Research Data pilot.<sup>4</sup> In our own work we can acknowledge that data and other documents produced at different stages of the research cycle are legitimate outputs of your research, and they can be counted as publications, including adding data as a specific publication type on academic curriculum vitae. See Alperin et al. (chapter 13, this volume) for discussion of valuation of data and data management as a research endeavor.

### 5.2 Using new tools and processes makes openness easier

Part of the response to the replication crisis in social psychology has been to develop tools and processes that facilitate good open data practice. Many of these tools are designed to be as easy to use as possible. OSF,<sup>5</sup> which we have been using in our own research practice, was created by the Center for Open Science,<sup>6</sup> the same group that have been running the large-scale replications. OSF benefits the work of Center for Open Science researchers, but is also useful for other disciplines. Linguists can harness these, as well as digital tools already created in our field, and institutional repositories and infrastructure, to facilitate their own open access plans. Many researchers feel uncomfortable about sharing data before they have completed their analysis in case another researcher beats them to publication, and the efforts they put into data collection do not result in the kind of publication they hoped for. However, it is important to realize that open data does not have to be open from the very moment it is created. Many online tools and archives provide the options for embargos, and private data sets allow you to protect your research goals, while making sure that your data also has a permanent home by the time of project completion. This does not mean you should leave the archiving of your data until after the project concludes; structuring data for sharing “as you go” and many repositories facilitate this.

Many newer tools and repositories are designed to be as user-friendly as possible. There are practical limits to how many tasks an individual researcher needs to become an expert in. Fortunately, universities are increasingly aware of the importance of open data, and most university libraries have research librarians and technical specialists who can provide support for archiving your materials, or providing training on the use of user-oriented repositories. The data-handling skills you need to make your work open access are the kind of skills that will make your data easier for you to use in your own research. The very best open data are well-organized, have clear metadata explaining what the data are and how they are structured, and are findable through a general web search. These kinds of information make your data easier to find and will make them easier for you to use in years to come.

### 5.3 A move to openness is about creating a positive cultural shift

We can have all the useful tools and processes in the world, but none of these will matter unless we foster a cultural shift in the discipline toward openness. There are several dimensions we see as key to this cultural shift, at the beginning of this discussion we briefly discussed incentives and there are others we discuss here: the first is the intergenerational shift; the second is openness even when data cannot be shared; or introspective data are used; and the third is reframing how we discuss data created by others, and how we respond to critique of our own analysis.

Common across many of the stories about the replication crisis in social psychology is the role senior researchers had in encouraging junior researchers to participate in potentially questionable data collection and analysis processes, with Wansink's story being perhaps the most egregious for the way he publicly blogged about it as a positive experience for his junior colleague. In many areas of linguistics, students and junior researchers do not always work so directly with senior colleagues on research projects, but are still greatly influenced by them. Supervisorial relationships are built on unequal power relations, and if you are a student or early career researcher in a research lab or department where there is a culture against data sharing, you may have to wait for another project to begin good practices.

When working as part of a team, or in a supervisor-student relationship, working out who "owns" data and who can decide what happens to them can be a complicated problem—particularly if you are one of the junior members of a team. Check with your team members about whether there are structural reasons why the data can't be shared (for example, you don't have permission from the informants, or from the original creator of the data). In the absence of a structural barrier, it is possible that senior team members may not be aware of the advantages of open data. They may be concerned about you completing your research within a fixed timeline, and that focusing on data will prevent you from focusing on publication. They may be concerned that sharing the raw data will allow another researcher to scoop you and prevent you from publishing your work. Talking through these issues may help you to find some common ground. It is also worth remembering that even senior researchers who you admire can feel threatened

by the changing research landscape with regard to data. There are a number of good resources you can share with colleagues to help in these conversations, including the articles and books we reference in this chapter, the frequently asked questions from the Austin Principles of Data Citation in Linguistics web page<sup>7</sup> and resources shared by the ReproducibiliTea<sup>8</sup> global network of journal clubs (Orben 2019).

There are times where it is not appropriate to share data. The interests of the community you work with always take priority, and you should always be sensitive to those interests. However, you have an obligation to revisit this discussion from time to time, including when there are genres, or derivative data (such as modified transcripts) that people are more willing to share with a wider audience, or a step model of open access, where data are only shared with registered or invited individuals. You have an obligation to clearly express in publications that not sharing data is an expressed wish of that community. This helps other communities and researchers to see that this is a conscious choice by a community, or individuals, and not an omission or withholding of data on behalf of the researcher. See Holton, Leonard, and Pulsifer (chapter 4, this volume) for more on people, ethics, and data.

Several linguistic traditions are built on introspection, or other methods where there is not necessarily a clearly defined data collection process. Introspection is still a form of data collection, and while there may not be a set of recordings or texts, there is still an obligation to be transparent about the nature of introspective data and the precise methods used to generate it. Just as a research experiment requires information about participants, publications about introspection should make clear who provided the introspections, and how they were acquired. The concept of open data means openness about methods as well as the data themselves.

One final dimension of openness, and perhaps the most difficult to put into practice, is openness to discussion of publicly available data. The replication crisis in social psychology has seen a number of ways research has been queried and a variety of different responses from original authors. The move toward open data involves a new level of vulnerability for researchers, as they share an element of the research cycle that was previously not made public. This is not dissimilar to the evolution of the practice of sharing preprints of articles

intended for peer review. This practice, which started to become normalized in physics and is now moving into other areas of the physical and social sciences (such as through the SocArXiv<sup>9</sup> platform) requires researchers to rethink the research pipeline. The social cost to the individual of sharing research at the prepublication stage comes with a social benefit when the whole field participates (Tennant et al. 2019). In the same way that the practice around preprints is shifting and stabilizing in new fields, expectations about what constitutes good open data will continue to shift and grow; furthermore, open practices do not completely eliminate problems of fraud or even of honest unintentional misanalysis. Critique of open data cannot immediately be on the same level as that of research practices with longer traditions, such as the peer-reviewed journal article. We also need to ensure that we find a way to talk critically about data that does not come across as *ad hominem*. On the flipside, we also need to accept that open data mean another phase of our research cycle is open to scrutiny. In the flurry of activity in the social science replication crisis, key researchers whose work could not be replicated responded saying they felt personally attacked. None of this reframing of data is easy for any individual researcher, but as a whole we can approach data in a positive and respectful way, to help drive our discipline forward. We believe this is one of the most important lessons we can learn from the replication crisis in social psychology.

## 6 Conclusion: It all comes down to transparency

This Handbook is designed to help you make choices about how you manage linguistic data. We hope that when you are making those choices you aim to be as transparent in your methodology and presentation of data as possible, to allow other researchers to engage with that data in a way that can help the field move forward positively. The chapters on archiving (Kung, chapter 8, this volume), developing research data management plans (Andreassen, chapter 7, this volume), and data copyright (Collister, chapter 9, this volume) will help you think in practical terms about how to implement transparency in your own research.

Berez-Kroeker et al. (2018:11) in linguistics and Asendorpf et al. (2013) in social psychology both make clear that good linguistic data management practice is not

just the responsibility of individual researchers, but of the field as a whole, including institutions, archives, publishers, and funding organizations. Institutions need to better value the development of corpora of data as a key output of research, with a positive valuation of more transparent data. See Alperin et al. (chapter 13, this volume) for more on the valuation of data and data management as a research endeavor. Individual researchers can also collaborate more, sharing the burden of validating key findings and driving research forward (Silberzahn & Uhlmann 2015). Archives need to continue to ensure they are providing a reliable repository of data, but are not difficult to use for both those depositing data and those accessing them. Publishers are increasingly seeing the value in encouraging or requiring researchers to make the data for a publication accessible (Nosek et al. 2015). Researchers in social psychology who are unwilling to wait for the publishers to drive change have started the Peer Reviewer's Openness Initiative (Morey et al. 2016), where people agree that they will use their role as peer reviewer to request a minimum level of data transparency for all papers they review. Funding organizations have an opportunity to set the agenda for how research is conducted through updates to funding rules. While we are not yet at the point where grant applications are essentially treated as preregistration of research methods (as per Bollen et al.'s recommendation to the National Science Foundation in 2015), greater focus should be given to how data will be managed, both in grant proposals and final reports. All of this requires a transformation in the way we approach research data.

## Acknowledgments

Thank you to all of our collaborators who make us think critically about data and how we use and share them. Thank you also to the participants in our work who are so often happy to have their contribution to research be as maximally useful as possible. This work was funded in part by La Trobe University, and The National Science Foundation (NSF SMA-1447886), Nanyang Technological University (NAP-SUG-M4081215 and Nanyang Researcher of the Year award to SJS), and the Singapore National Research Foundation (NRF2016-SOL002-011). Thanks also to Brent Roberts and Dan Simons for making the extensive reading list for their 2018 graduate course on Reproducibility and Replicability publicly available.

## Notes

1. <http://osf.io/wt95v/>.
2. <http://ice-corpora.net/ice/>.
3. <https://www.linguisticsociety.org/resource/statement-evaluation-language-documentation-hiring-tenure-and-promotion>.
4. <http://erc.europa.eu/funding-and-grants/managing-project/open-access>.
5. <http://osf.io/>.
6. <http://cos.io/>.
7. <http://site.uit.no/linguisticsdatacitation/faq/>.
8. <http://reproducibilitea.org/>.
9. <http://socopen.org>.

## References

- Asendorpf, J. B., M. Conner, F. De Fruyt, J. De Houwer, J. J. A. Denissen, K. Fiedler, S. Fiedler, et al. 2013. Recommendations for increasing replicability in psychology. *European Journal of Personality* 27 (2): 108–119.
- Baker, M. 2016. Is there a reproducibility crisis? *Nature* 533 (7604): 3–5.
- Barreda, S. 2016. Investigating the use of formant frequencies in listener judgments of speaker size. *Journal of Phonetics* 55:1–18.
- Bem, D. J. 2011. Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology* 100 (3): 407.
- Berez-Kroeker, A. L., L. Gawne, B. F. Kelly, and T. Heston. 2017. Survey of current reproducibility practices in linguistics journals, 2003–2012. <http://sites.google.com/a/hawaii.edu/data-citation/survey>. Accessed October 5, 2018.
- Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, et al. 2018. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics* 56 (1): 1–18. <http://doi:10.1515/ling-2017-0032>.
- Bergh, D. D., B. M. Sharp, H. Aguinis, and M. Li. 2017. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization* 15 (3): 423–436.
- Boersma, P., and D. Weenink. 2013. Praat: Doing phonetics by computer (Version 5.3.39) (computer program). <http://www.praat.org>. Accessed December 31, 2018.
- Bollen, K., J. T. Cacioppo, R. M. Kaplan, J. A. Krosnick, J. L. Olds, and H. Dean. 2015. *Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science*. Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences, 3.
- Brandt, M. J., H. IJzerman, A. Dijksterhuis, F. J. Farach, J. Geller, R. Giner-Sorolla, J. A. Grange, et al. 2014. The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology* 50:217–224.
- Brown, Sacha D., David Furrow, Daniel F. Hill, Jonathon C. Gable, Liam P. Porter, and W. Jake Jacobs. 2014. A duty to describe: Better the devil you know than the devil you don't. *Perspectives on Psychological Science* 9 (6): 626–640.
- Caballero, Gabriela. 2017. Choguita Rarámuri (Tarahumara) language description and documentation: A guide to the deposited collection and associated materials. *Language Documentation and Conservation* 11:224–255.
- Chambers, C. D. 2013. Registered reports: A new publishing initiative at Cortex. *Cortex* 49 (3): 609–610.
- Chambers, Chris. 2017. *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*. Princeton: Princeton University Press.
- Cohen, J. 1962. The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology* 65:145–153.
- Elms, A. C. 1975. The crisis of confidence in social psychology. *American Psychologist* 30:967–976.
- Fanelli, D. 2011. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90:891–904.
- Fecher, B., and S. Friesike. 2014. Open science: One term, five schools of thought. In *Opening Science: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing*, ed. S. Bartling and S. Friesike, 17–47. Springer Open. <http://doi.org/10.1007/978-3-319-00026-8>.
- Ferguson, C. J., and M. Heene. 2012. A vast graveyard of undead theories publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science* 7 (6): 555–561.
- Fetterman, A. K., and K. Sassenberg. 2015. The reputational consequences of failed replications and wrongness admission among scientists. *PLoS One* 10 (12): e0143723. <http://doi.org/10.1371/journal.pone.0143723>.
- Feynman, Richard P. 1974. Cargo cult science. *Engineering and Science* 37 (7): 10–13.
- Fowler, J. 1986. The social stratification of (r) in New York City department stores, 24 years after Labov. Unpublished manuscript.
- Franco, A., N. Malhotra, and G. Simonovits. 2014. Publication bias in the social sciences: Unlocking the file drawer. *Science* 345 (6203): 1502–1505.

- Franco, A., N. Malhotra, and G. Simonovits. 2016. Under-reporting in psychology experiments: Evidence from a study registry. *Social Psychological and Personality Science* 7:8–12.
- Galak, J., R. A. LeBoeuf, L. D. Nelson, and J. P. Simmons. 2012. Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology* 103 (6): 933.
- Gawne, L., B. F. Kelly, A. L. Berez-Kroeker, and T. Heston. 2017. Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation and Conservation* 11:157–189.
- Goldacre, Ben. 2010. *Bad Science: Quacks, Hacks, and Big Pharma Flacks*. London: McClelland and Stewart.
- Goldfarb, B., and A. A. King. 2016. Scientific apophenia in strategic management research: Significance tests and mistaken inference. *Strategic Management Journal* 37 (1): 167–176.
- Goodman, S. N., D. Fanelli, and J. P. Ioannidis. 2016. What does research reproducibility mean? *Science Translational Medicine* 8 (341): 341ps12.
- Grabitz, C. R., K. S. Button, M. R. Munafò, D. F. Newbury, C. R. Pernet, P. A. Thompson, and D. V. Bishop. 2018. Logical and methodological issues affecting genetic studies of humans reported in top neuroscience journals. *Journal of Cognitive Neuroscience* 30 (1): 25–41. [http://doi.org/10.1162/jocn\\_a\\_01192](http://doi.org/10.1162/jocn_a_01192).
- Guest, Olivia. 2016. Crisis in what exactly? *The Winnower* 6:e146590. doi:10.15200/winn.146590.01538.
- H2020 Programme. 2016. *Guidelines on FAIR Data Management in Horizon 2020*. Version 3.0, July 26, 2016. European Commission Directorate-General for Research and Innovation. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf). Accessed November 1, 2018.
- Head, M. L., L. Holman, R. Lanfear, A. T. Kahn, and M. D. Jennions. 2015. The extent and consequences of p-hacking in science. *PLoS Biology* 13 (3): e1002106. <http://doi.org/10.1371/journal.pbio.1002106>.
- Henneken, Edwin A., and Alberto Accomazzi. 2011. Linking to data: Effect on citation rates in astronomy. *arXiv arXiv:1111.3618*.
- Himmelman, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 6:161–195.
- Ioannidis, J. P. 2005. Why most published research findings are false. *PLoS Med* 2:e124.
- Ioannidis, J. P. A. 2012. Why science isn't necessarily self-correcting. *Perspectives on Psychological Science* 7 (6): 645–654. <http://doi:10.1177/1745691612464056>
- Jennions, M. D., and A. P. Møller. 2003. A survey of the statistical power of research in behavioral ecology and animal behavior. *Behavioral Ecology* 14 (3): 438–445.
- Johnson, Valen E., Richard D. Payne, Tianying Wang, Alex Asher, and Soutrik Mandal. 2017. On the reproducibility of psychological science. *Journal of the American Statistical Association* 112 (517): 1–10. <http://doi:10.1080/01621459.2016.1240079>.
- Joseph, Heather. 2013. The open access movement grows up: Taking stock of a revolution. *PLoS Biology* 11 (10): e1001686. <http://doi:10.1371/journal.pbio.1001686>.
- Just, D. R., Ö. Siğirci, and B. Wansink. 2014. Lower buffet prices lead to less taste satisfaction. *Journal of Sensory Studies* 29 (5): 362–370.
- Just, D. R., Ö. Siğirci, and B. Wansink. 2015. Peak-end pizza: Prices delay evaluations of quality. *Journal of Product and Brand Management* 24 (7): 770–778.
- Kitchin, Rob. 2014. *The Data Revolution*. London: Sage.
- Klein, R. A., M. Vianello, F. Hasselman, B. G. Adams, R. B. Adams, Jr., S. Alper, M. Aveyard, et al. 2018. Many Labs 2: Investigating variation in replicability across sample and setting. *PsyArXiv* November 19. <https://doi.org/10.31234/osf.io/9654g>.
- Kniffin, K. M., Ö. Siğirci, and B. Wansink. 2016. Eating heavily: Men eat more in the company of women. *Evolutionary Psychological Science* 2 (1): 38–46.
- Koole, S. L., and D. Lakens. 2012. Rewarding replications: A sure and simple way to improve psychological science. *Perspectives on Psychological Science* 7 (6): 608–614. doi:10.1177/1745691612462586.
- Kunert, R. 2016. Internal conceptual replications do not increase independent replication success. *Psychonomic Bulletin and Review* 23 (5): 1631–1638.
- Labov, W. (1966) 2006. *The Social Stratification of English in New York City*. Cambridge: Cambridge University Press.
- LeBel, E. P., and K. R. Peters. 2011. Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study in deficiencies in modal research practice. *Review of General Psychology* 15:371–379. <http://doi10.1037/a0025172>.
- Lilienfeld, S. O. 2017. Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science* 12 (4): 660–664.
- Loken, E., and A. Gelman. 2017. Measurement error and the replication crisis. *Science* 10:584–585. <http://doi:10.1126/science.aal3618>.
- Lykken, D. T. 1991. What's wrong with psychology anyway? In *Thinking Clearly about Psychology*, ed. D. Cicchetti and W. Grove, 3–39. Minneapolis: University of Minnesota Press.
- MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates.
- Makel, M. C., and J. A. Plucker. 2014. Facts are more important than novelty: Replication in the education sciences. *Educational Researcher* 43 (6): 304–316.



- Makel, M. C., J. A. Plucker, and B. Hegarty. 2012. Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science* 7:537–542.
- Mather, P. A. 2011. The social stratification of /t/ in New York City: Labov's department store study revisited. *Journal of English Linguistics* 40 (4): 338–356.
- Matosin, N. E. Frank, M. Engel, J. S. Lum, and K. A. Newell. 2014. Negativity towards negative results: A discussion of the disconnect between scientific worth and scientific culture. *Disease Models and Mechanisms* 7:171–173. doi:10.1242/dmm.015123.
- Maxwell, S. E., M. Y. Lau, and G. S. Howard. 2015. Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist* 70 (6): 487–498. <http://doi.org/10.1037/a0039400>.
- Mellor, D. T., and B. A. Nosek. 2018. Easy pre-registration will benefit any research. *Nature Human Behavior*. doi:10.1038/s41562-018-0294-7.
- Meyer, Michelle N., and Christopher Chabris. 2014. Why psychologists' food fight matters. *Slate*. [http://www.slate.com/articles/health\\_and\\_science/science/2014/07/replication\\_controversy\\_in\\_psychology\\_bullying\\_file\\_drawer\\_effect\\_blog\\_posts.html](http://www.slate.com/articles/health_and_science/science/2014/07/replication_controversy_in_psychology_bullying_file_drawer_effect_blog_posts.html). Accessed October 4, 2018.
- Morey, R. D., C. D. Chambers, P. J. Etchells, C. R. Harris, R. Hoekstra, D. Lakens, S. Lewandowsky et al. 2016. The peer reviewers' openness initiative: Incentivizing open research practices through peer review. *Royal Society Open Science* 3 (1): 150547.
- Murphy, K. R., and H. Aguinis. 2017. HARKing: How badly can cherry-picking and question trolling produce bias in published results? *Journal of Business and Psychology* 34 (1): 1–17.
- Nelson, L. D., J. Simmons, and U. Simonsohn. 2018. Psychology's renaissance. *Annual Review of Psychology* 69:511–534.
- Neuliep, J. W., and R. Crandall. 1990. Editorial bias against replication research. *Journal of Social Behavior and Personality* 5 (4): 85–90.
- Neuliep, J. W., and R. Crandall. 1993. Reviewer bias against replication research. *Journal of Social Behavior and Personality* 8 (6): 21–29.
- Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck et al. 2015. Promoting an open research culture. *Science* 348 (6242): 1422–1425.
- Nosek, B. A., and D. Lakens. 2014. Registered reports: A method to increase the credibility of published results. *Social Psychology* 45 (3): 137–141.
- Nuijten, M. B., C. H. Hartgerink, M. A. van Assen, S. Epskamp, and J. M. Wicherts. 2016. The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods* 48 (4): 1205–1226.
- Orben, A. 2019. A journal club to fix science: Reproducibility can build up open science without top-down initiatives. *Nature* 573:465. doi:10.1038/d41586-019-02842-8.
- Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349 (6251): aac4716.
- Peterson, G. E., and H. L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24 (2): 175–184.
- Pienta, Amy M., George C. Alter, and Jared A. Lyle. 2010. The enduring value of social science research: The use and reuse of primary research data. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. <http://hdl.handle.net/2027.42/78307>.
- Piwowar, Heather A., and Todd J. Vision. 2013. Data reuse and the open data citation advantage. *PeerJ* 1:e175.
- Retraction Watch. 2018. The *Joy of Cooking*, vindicated: Journal retracts two more Brian Wansink papers. <http://retractionwatch.com/2018/12/05/the-joy-of-cooking-vindicated-journal-retracts-two-more-brian-wansink-papers>. Accessed December 31, 2018.
- Rosenthal, Robert. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86 (3): 638–641.
- Salfner, Sophie. 2015. A guide to the Ikaan language and culture documentation. *Language Documentation and Conservation* 9:237–267.
- Schmidt, S. 2009. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology* 13 (2): 90–100.
- Schweizer, G., and P. Furley. 2016. Reproducible research in sport and exercise psychology: The role of sample sizes. *Psychology of Sport and Exercise* 23:114–122.
- Siğirci, Ö., and B. Wansink. 2015. Low prices and high regret: How pricing influences regret at all-you-can-eat buffets [retracted article]. *BMC Nutrition* 1 (1): 36.
- Silberzahn, R., and E. L. Uhlmann. 2015. Crowdsourced research: Many hands make tight work. *Nature* 526:189–191. <http://www.nature.com/news/crowdsourced-research-many-hands-make-tight-work-1.18508>.
- Simonsohn, Uri. 2013. Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science* 24 (10): 1875–1888.
- SPARC. n.d. Open Data Factsheet (11.10–2). <http://sparcopen.org/our-work/sparc-fact-sheets/>. Accessed October 5, 2018.
- Stanley, D. J., and J. R. Spence. 2014. Expectations for replications: Are yours realistic? *Perspectives on Psychological Science* 9 (3): 305–318.
- Sterling, T. D. 1959. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association* 54 (285): 30–34.

Styles, Suzy J. 2018. Coin-flipping cowboys: The wild west of data science, lucky cowboys and picking your team after the rodeo is over. *DR-NTU (Data)*, V1. <http://doi.org/10.21979/N9/8Z8CRQ>.

Styles, Suzy J., and Lauren Gawne. 2017. When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception* 8 (4): 2041669517724807. <http://doi.org/10.1177/2041669517724807>.

Tennant, Jonathan P., Harry Crane, Tom Crick, Jacinto Davila, Asura Enkhbayar, Johanna Havemann, Bianca Kramer, et al. 2019. Ten hot topics around scholarly publishing. *Publications* 7 (2): 34.

Thieberger, N. 2017. LD&C possibilities for the next decade. *Language Documentation and Conservation* 11:1–4.

Thomason, S. 1994. The editor's department. *Language* 70:409–423.

van der Zee, Tim, Jordan Anaya, and Nicholas J. L. Brown. 2017. Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab. *PeerJ Preprints* 5:e2748v1. doi:10.7287/peerj.preprints.2748v1.

Wagenmakers, E.–J., R. Wetzels, D. Borsboom, D., and H. L. J. van der Maas. 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.

Walster, G. W., and T. A. Cleary. 1970. A proposal for a new editorial policy in the social sciences. *American Statistician* 24 (2): 16–19.

Watrous, R. L. 1991. Current status of Peterson–Barney vowel formant data. *Journal of the Acoustical Society of America* 89:2459–2460.

Wicherts, J. M., D. Borsboom, J. Kats, and D. Molenaar. 2006. The poor availability of psychological research data for reanalysis. *American Psychologist* 61 (7): 726–728.

Wicherts, J. M., C. L. Veldkamp, H. E. Augusteijn, M. Bakker, R. C. Van Aert, and M. A. Van Assen. 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid *p*-hacking. *Frontiers in Psychology* 7:1832.

Zwaan, R. A., A. Etz, R. E. Lucas, and M. B. Donnellan. 2018. Making replication mainstream. *Behavior and Brain Sciences* 41:e120. doi:10.1017/S0140525X17001972.



© 2021 The Massachusetts Institute of Technology

This work is subject to a Creative Commons CC-BY-NC license. Subject to such license, all rights are reserved.



This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Berez-Kroeker, Andrea L., editor. | McDonnell, Bradley James, editor. | Koller, Eve, editor. | Collister, Lauren B., editor.

Title: The open handbook of linguistic data management / edited by Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller and Lauren B. Collister.

Description: Cambridge, Massachusetts : The MIT Press, [2021] | Series: Open handbooks in linguistics series | Includes bibliographical references and index.

Identifiers: LCCN 2020044363 | ISBN 9780262045261 (hardcover)

Subjects: LCSH: Computational linguistics. | Natural language processing (Computer science) | Data mining.

Classification: LCC P98 .O64 2021 | DDC 410.285—dc23

LC record available at <https://lcn.loc.gov/2020044363>