

# A Framework for Really Reproducible Research

Scott Jackson

February 21, 2013



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Two motivations . . . . .	5
1.2	Three sides of the coin . . . . .	6
1.3	How to use this book . . . . .	6
1.4	Who this book is for . . . . .	7
1.5	Theses and structure of the book . . . . .	8
<b>2</b>	<b>Principles</b>	<b>11</b>
2.1	Replication vs. reproduction . . . . .	11
2.2	A sketchy history of reproducibility . . . . .	12
2.3	Expanded motivations . . . . .	13
2.3.1	Being a more successful scientist . . . . .	13
2.3.2	Doing better science . . . . .	16
2.3.3	Summary of motivations . . . . .	19
2.4	A dissenting opinion and a rejoinder . . . . .	19
2.5	Some core values . . . . .	23
2.6	Who is this book for? . . . . .	23
2.7	Goals . . . . .	24
<b>3</b>	<b>Dimensions of reproducibility</b>	<b>27</b>
3.1	Domain: the audience . . . . .	27
3.2	Range: the precision . . . . .	28
3.3	The minimum . . . . .	28
3.4	Publication standards . . . . .	28
3.5	Wide dissemination . . . . .	28
<b>4</b>	<b>The cycle of research</b>	<b>29</b>
<b>5</b>	<b>Scholarship</b>	<b>31</b>
<b>6</b>	<b>Theory</b>	<b>33</b>
<b>7</b>	<b>Data collection</b>	<b>35</b>
<b>8</b>	<b>Data analysis</b>	<b>37</b>

<b>9</b>	<b>Reporting</b>	<b>39</b>
<b>10</b>	<b>Conclusions</b>	<b>41</b>
<b>11</b>	<b>Implementation</b>	<b>43</b>
11.1	Guiding principles . . . . .	44
11.1.1	Free and open source . . . . .	44
11.1.2	Cross-platform . . . . .	44
11.1.3	Stable . . . . .	44
11.1.4	Well-documented . . . . .	44
11.1.5	Customizable . . . . .	44
11.2	Summary of tools . . . . .	44
11.2.1	Emacs . . . . .	44
11.2.2	Org-mode . . . . .	44
11.2.3	Git . . . . .	44
11.2.4	L <sup>A</sup> T <sub>E</sub> X . . . . .	44
11.2.5	Python . . . . .	44
11.2.6	R . . . . .	44
11.2.7	(Emacs) Lisp . . . . .	44
11.3	Scholarship . . . . .	44
11.4	Data collection . . . . .	44
11.5	Data analysis . . . . .	44
11.6	Sharing . . . . .	44
11.7	Collaboration . . . . .	44
11.8	Putting it all together . . . . .	44

# Chapter 1

## Introduction

### 1.1 Two motivations

This book is about *reproducible research*. A lot of the book is dedicated to fleshing out exactly what that term means, or perhaps what it *should* mean, but moreover, this book is about *really* reproducible research. In order to tell you what *that's* supposed to mean in a relatively concise way, I'm going to talk about the two different kinds of motivations that drive this book.

One type of motivation is the desire to do Good Science. It's my belief that scientific research should be conducted in a way that is maximally replicable, reproducible, transparent, and trustworthy. I think of this as a kind of ethical obligation for scientists, both to the scientific community and to the people that fund our research (usually taxpayers, in my line of work). I also believe that this is inherent to the success of the scientific method, but I am not an expert on the philosophy of science, so I won't go too deeply into that side of things, beyond expressing some of my personal opinions. From this ethical/philosophical point of view, reproducible research is in line with the value system of science, and *really* reproducible research is a set of real-world methods and techniques that do more than lip service to these ideas; they put them into action.

The second type of motivation is the desire to be a Successful Scientist. This is the more practical, pragmatic side of things. In short, being a successful scientist means doing work that gets published frequently in high-quality journals. Obviously for many academics, there are other sides to professional life like teaching, service, etc. Maybe some of the ideas here can apply some to those domains, too, but I'm going to focus on the research side. A successful researcher, from the most concrete point of view, is one with a good publishing record. From this angle, the motivation of reproducible research is to help the researcher be more efficient, effective, and productive, and to produce higher-quality work that is worthy of the most prestigious journals. And similarly, *really* reproducible research is meant to provide a concrete framework that's more than just recapitulating concepts about effective work habits.

Now hopefully, if you're engaged in research, you're interested in one of these

two motivations, maybe both. If you neither care about (a) producing good science worthy of the faith and trust of other scientists and your funding sources, nor (b) being a successful researcher with a great publication record, then maybe you're in the wrong field. The central aim of this book is to talk about how reproducibility — *real* reproducibility — can assist you in both of these goals.

## 1.2 Three sides of the coin

In order to address these two goals, there are three (count 'em!) sides of the coin we need to explore. One side is the conceptual side, trying to lay out a framework of principles for understanding what it means to be reproducible. The other side is a discussion of how these principles play out in the day-to-day activities of carrying out research. But while this second side is more concrete, it's still framed in terms of general principles that could be implemented in many different ways. The third side is an actual, working implementation. You can see the third side (the edge) of the coin as either the part that connects the two faces of theory and practice, or as I prefer to see it, as the edge of a wheel, where (to mix metaphors) the rubber really meets the road. It's fine to talk about good principles of reproducibility, but without a kind of “manual” for exactly how to put things into practice, there will be a high barrier to entry for most people, which means very little change. If we want *really* reproducible research we have to have an actual working system that enables our good intentions to be realized in a practical way.

This book is aimed at the first two, more conceptual sides. The companion to this book is a repository on GitHub that provides a set of tutorials and guides for using a set of (open source, free) software tools to actually implement the ideas in the book. This “companion” set of materials is ultimately what puts the “really” in the title of the book, but I just couldn't see how to effectively squeeze all those tutorials and things into a useful book format. What I will do is include lots of hyperlinks in this book, in order to make it easier to connect from the idea or motivation for a particular kind of habit, workflow, or tool, to a set of tutorials that will help you get up and running with a tool that actually works as quickly as possible. This brings me to the next section.

## 1.3 How to use this book

Here's how I imagine (and hope) this book could be used. Read through some of the general chapters in the beginning, to get a sense of the big issues. But then feel free to skip around to particular issues or problems that sound interesting. The material in this book will lay out some of my thoughts on the topic, which will hopefully lead you to arrive at your own thoughts and opinions, and if my ideas intrigue you (either because they sound good, or because they sound terrible), you can follow some links to the GitHub repo to see how I personally attempt to implement my ideas. Those links will teach you how to use some piece of software (or non-software solution, if appropriate) to implement things yourself. All the software and techniques I use are extremely customizable, so if you don't like exactly how I do it and want to tweak it

for your own use, or if you think you can improve on it more generally, my tutorials will hopefully give you a good start towards being able to do that as well.

At some point, I expect this book to become rather static, in the sense that I will revise things as my thinking changes, and as other revisions/corrections need to be made, but if I ever start to change my mind in a big enough way, I may just need to write a different book. But the repository of tutorials and software is going to be a dynamic thing. It is ultimately a selfish endeavor, because I will base the tutorials around my own usage and needs. As those things change, so will the tutorials. However, since I will be using `git` for version control, you will be able to “roll back” any particular tool or tutorial to an earlier phase, if you thought those worked better for you. At least, you’ll be able to do this once you learn how to use `git`, which happens to be the base of what I’m going to recommend for a reproducible workflow! See, this stuff is already sounding useful, and we haven’t even really gotten started yet...

But the point is that the tutorial stuff will change a lot (assuming I can keep up with it), in an ongoing way. I would be very happy to have feedback on any aspect of this whole enterprise, whether you have suggestions or arguments with things in this book, or whether you have suggestions, problems, or alternatives to the implementation stuff on the GitHub site. Please direct all your love/hate mail to [shoestringpsycholing1@gmail.com](mailto:shoestringpsycholing1@gmail.com).

## 1.4 Who this book is for

By trade, I am a cognitive scientist, working in linguistics, psycholinguistics, and second language acquisition. Since this book is inherently a self-centered effort, I will be aiming at those audiences. However, I expect that the basic ideas and concepts should be applicable pretty generally across a lot of different domains of research. In particular, what I’m trying to do in this book is to take a look at some ideas that have been growing in popularity in computational and statistical sciences, and see whether they might apply more generally to research where developing computer code and new statistical/computational algorithms is not inherent to the field. In other words, by talking about how a (non-computational) linguist might be able to employ more reproducible methods, I expect the general principles will be pretty general, beyond the computation-heavy fields.

Thus, the book and the accompanying methods and tutorials are aimed at a pretty broad audience. I do not start out by assuming any particular knowledge of programming, statistics, or anything else. Of course, my implementations are all based around certain kinds of software packages that *do* involve learning at least a little programming, and I’ll argue that this is the most efficient way to implement a reproducible workflow, but the principles and ideas in this book are intended to be aimed at a much more general audience than the set of people who can easily put together a `makefile` (or even know what that is).

## 1.5 Theses and structure of the book

This is not the first thing that’s ever been written regarding reproducible research. It’s a rather hot topic in some circles, and lots of smart people have been thinking/writing/talking/blogging/tweeting about it. One goal of this book is to push a few new ideas into the discussion. In particular:

1. Whether something is *reproducible* is not an absolute, but is relative to the *range* (i.e., precision, extent) of reproducibility, the *domain* of reproducibility (i.e., what kind of activities are being reproduced?), and the *audience* of reproducibility (i.e., reproducible for whom?).
2. *Reproducible* has to be able to extend beyond just the domain of data analysis; it should apply to all aspects of the cycle of research.

I’ll focus on these two theses more or less in turn. I will start with a broad overview in chapter 2. I’ll review some of the previous work on the general topic, and further flesh out the broad concepts and motivations I’ve alluded to so far. This is more or less re-hashing and re-packaging a lot of things that other people have probably said better, though maybe not all in one place. Then in chapter 3 I’ll talk about the *dimensions* of reproducibility. I will argue that the notion of *reproducibility* is inherently gradient and scalar, and that any definition or standard will have to make some decisions about the target dimensions that qualify something as “in” or “out” with respect to that local definition of what counts as reproducible. In other words, *reproducible* is a relative goal, and while it may be possible to establish standards for a particular field/context, there is no such thing as a useful universal standard. I will then go on to suggest some possible starting points for where to position a reasonably useful definition along these dimensions. That will conclude all of this nosebleed-level discussion.

I’ll then turn to a slightly more concrete discussion of how to make the day-to-day tasks of research more reproducible. In chapter 4 I propose a general schema for the cycle of research, from reading other people’s work to producing your own work. Again, my purpose here is not to go too deep into the philosophy of science, so this is just intended as a way of breaking down the research process into some large chunks. I think this is useful for the present purpose, because these different kinds of activities will involve different issues, problems, and standards regarding what it means to be *reproducible*. But the big point is that this broad view of research goes far beyond the realm of what most people talk about when the term *reproducible research* gets thrown around.

In the chapters that follow (5 through 9), I will focus in one one of these domains of research, and discuss some of the special challenges for reproducibility, and at times grapple with the question of whether (and how) activities in this domain could ever be reproducible. In each of these chapters, I will discuss how the other dimensions (*range* and *audience*) play out for that particular domain. Each chapter will also contain a general description of some practices that could lead to better reproducibility, and I will link heavily to my repository of tutorials that will implement specific software (or non-software) solutions to the implementation problem.



The final chapter will conclude with some parting thoughts.

So let's get started!



## Chapter 2

# Principles

### 2.1 Replication vs. reproduction

Let's start with a terminological clarification. *Replication* and *reproduction* are terms that often get used to talk about similar yet distinct concepts. The basic idea is that *replication* refers to one of the cornerstones of science, which is that someone can perform an experiment, gets some results, and someone *else* can perform a *different* (but possibly similar) experiment, and get a set of results that is consistent with the first. We can say that these results have been *replicated*. *Reproducible* is often used to refer to the ability to get *exactly* the same results starting from the same set of data.

The meaningful distinction here, I propose, is that *replication* is about *results* and *reproduction* is about *methods*. These concepts are somewhat independent, because it may be possible to reproduce a set of methods almost exactly, but if you're collecting data from a different set of participants, you may not end up replicating the initial result. And conversely, it may be difficult or impossible to reproduce someone's methods exactly, or you may be intentionally running an experiment with a variation on the methods, but you still might get a set of results that replicate the earlier findings.

Throughout this book, I will assume that *replication* is a non-negotiable part of the scientific process (though see discussion in section 2.4). If no one can produce results that fit with an earlier set of results, then typically scientists assume there was something wrong with the initial results (all else being equal). I propose that *reproducibility* is a methodological approach (or set of approaches) that should make replication easier if a result is true, and should make aberrant or erroneous results more transparent. In other words, having reproducible methods does not guarantee replication, nor are reproducible methods strictly necessary for replication. But it's my contention (and the opinion of many others) that reproducible methods really facilitate the process in an important way.

With this distinction out of the way, let's talk a little about how the notion of reproducibility has developed recently.

## 2.2 A sketchy history of reproducibility

I will not go through a detailed, scholarly review of reproducible research, but I will give a brief overview, because to help orient you to how the content of this book fits with other discussions of the general topic.

At its roots, the discussion of reproducible research in the modern context comes from computer science, statistics, and more recently, other disciplines (like genetic biology) that involve increasingly computation-heavy analysis. One of the seminal works is Knuth (1984), which coined the eponymous term “literate programming.” the specific idea was that good programs should be written with the *human* reader in mind, not (just) the computer that runs the programs. This idea has lead to further ideas and developments in methods of documenting code, methods of writing code to be more legible (including developing computer languages themselves to be more legible), and methods of intertwining (or to use the Knuth’s term, “weaving”) human language that explains the code with the code language itself. But the broader idea is the foundation of reproducibility in the computer sciences, which is that in order for a program (and its results) to be shared effectively with the broader community, some care needs to be taken in how it is created. A program that is opaque to people other than its creator will be far less useful than a program that can be understood easily, because the latter can be improved, expanded, adapted, modified, and debugged far more easily.

Jon Claerbout is the next big name on the list, as the coiner of the term “reproducible research” (and, according to Buckheit and Donoho 1995, Claerbout preceded me in using the term “really reproducible” as well), and as an influential early adopter and developer of methods for reproducible computational research. One of Claerbout’s earliest forays was a paper given at the 1992 annual meeting of the Society of Exploration Geophysics,<sup>1</sup> but more recent commentaries are given in Schwab, Karrenbach, and Claerbout (2000) and Fomel and Claerbout (2009), and there are various places (like the Madagascar project, at <http://www.ahay.org/>) where colleagues have implemented complete systems for reproducible research within particular domains.

Other people that have been especially visible in various ways include Robert Gentleman<sup>2</sup> (e.g., Gentleman and Temple Lang 2004; Gentleman et al. 2005), Roger Peng (e.g., Peng 2009; Peng 2008), and Victoria Stodden (e.g., Stodden 2009; Stodden 2010; Stodden 2011; Stodden 2012).

The common thread with all of these major movers and shakers is that they are primarily in the computational and statistical fields. There is a reason for this. The most common current notion of reproducible research, as forwarded by the people above and others, comes down to sharing the data and code that generate a particular analysis, set of results, figures, etc. This has become a bigger deal in these fields, as the analyses themselves have become more complex and sometimes very computation-intensive. In computer science, it seems obvious that sharing code should be part of the publication process, because in many cases, the code *is* the

<sup>1</sup>An “extended abstract” is available here:

<http://sepwww.stanford.edu/doku.php?id=sep:research:reproducible:seg92>

<sup>2</sup>Yes, [that Robert Gentleman](#).

research! For example, someone writes a piece of software to accomplish X, and then publishes a paper that describes the overall ideas, and maybe some benchmarks or other evidence that it does actually accomplish X. But the primary interest from other people in the field may be in *how* the software actually does what it does. In proprietary contexts, this may be kept secret, so that the owners of the software can sell it and beat their competitors, who are unable to do X. But in scientific contexts, I think secrecy is unacceptable, because it makes it impossible for the scientific community to review, accept, and synthesize the work. Black boxes are not good for science, though they may be good for warfare or profit (*maybe*).

Beyond sharing your code because your code *is* the research, the argument is that sharing your code allows for scientific openness because other people can inspect your code for mistakes, intentional or otherwise. It's a "show your work" kind of argument. In fields where the computations are non-trivial, and *how* you get a results is as much an innovation as *what* the results are, showing your work is an important part of dissemination and communication across the field. So it's no wonder that fields like Peng's (biostatistics) and Stodden's (statistics and computation) are at the forefront of pushing the ideas of reproducible research.

What about fields in the cognitive and social sciences? What about the processes involved in science before you have the data in hand, or after the analyses have been performed? While I think the current movement in reproducible research is great, it doesn't go nearly far enough. That's why I'm writing this book.

## 2.3 Expanded motivations

If you've gotten this far, I've given you a few teasers in terms of what this is all about, and why you might care. Or maybe you've skipped to this part to decide whether reading this is worth your time. So here I'm going to expand a little on the opening section and enumerate what I think the main "selling points" are for putting in the effort to make your research methods more reproducible.

### 2.3.1 Being a more successful scientist

I will talk about the "selfish" interests first. I'm talking about producing more and better research, getting more published, and generally being better at your job (if that involves research).

Here are some common issues and questions within research/academia, which can be enormously time-consuming, energy-consuming, and even (if things go poorly) career-threatening:

- You have completed months of data formatting, coding, and analysis, and have spent hours and hours formatting tables of results, creating, formatting, and labeling plots, and inserting statistics into a paper. You (or a colleague, grad student, or especially sharp reviewer) realize that some part of your data was coded wrong, or should be excluded. How long will it take to reproduce everything you've done after fixing your data?

- You are working on a complex data set, and in the process you try several different analyses, some of them very similar, varying only by a single parameter in your statistics software. You decide to report what you think is the best analysis in a paper. Six months later, someone asks you how you did that analysis, because they are trying something similar, and not getting the same results. How do you find the exact sequence of steps and set of options that produced the precise results in your paper? How long will this take?
- Could you pick up one of your own papers of from two years ago (or more) and remember exactly how you analyzed the data? Often the time between doing an analysis and it appearing in print can easily be two years or more. Someone reads your paper “hot off the presses,” (i.e., after a couple of years have passed), and has some good questions. Will you be able to easily remember the details to address those questions?
- You are interested in a particular line of research and want to start your own research by replicating the same pattern of effects someone else has shown. How much effort will this take? If you can’t replicate, how do you compare what you did to what they did? If you ask them to provide details (maybe five or more years after they did the initial work), how likely do you think they will be able to provide sufficient detail to help you? How easy would it be for you to do the same if someone wanted to replicate *your* work?
- How can you be sure that a result/statistic you report in your paper is correct? Detecting typos in prose is relatively easy, since the copyeditor (presumably) knows the language you’re writing in. But how could a copyeditor, or anyone else, catch a typo in your results/stats?
- How quickly can you access a complete bibliography of research that has influenced your thoughts on a particular topic?
- Can you quickly and accurately recall which important theoretical points were developed in which particular papers by a particular author?
- Can you quickly compare notes with someone about papers they’ve read to see if there are any holes in your own reading of the literature?
- Making notes about papers as you’re reading them is often very helpful, and can lead to new insights and new research questions. How easy is it to find those notes when you need them (or even remind yourself that you have notes)? How easy is it to lose those notes, and forget the points you thought about when reading?
- How much time do you spend searching for bibliographic references, even of papers you’ve cited before (or even papers you’ve written)?
- How much time do you spend formatting references, and making sure that the references section of your paper has all and only the references you cite?

- How many times have you made a typo on someone’s name in your references?
- Have you ever wanted to be able to “rewind the tape” on an analysis or draft of a paper, either to revert back to an earlier stage, or recall something that you had removed?
- Have you ever sent dozens of paper drafts back and forth with colleagues via email? How easy is it to find the most updated version, and be confident that it really is the most updated version? Do you have half a dozen different drafts labeled “FINAL”? How easy is it to find a particular draft of the paper, where some critical change was made? How easy is it to attribute certain sections to particular authors?

This is just an initial sampling. I contend that *really* reproducible methods provide you with apparently super-human abilities to address these issues and many more. What if I told you that the answer to the first bullet would be something like “dozens of hours” for non-reproducible methods, and “maybe an hour or two” for really reproducible methods? I also propose that if you have a decent career, you will run into these issues over and over. Therefore, the investment of time and energy into reproducible methods has enormous payback potential.

The above points are about efficiency. But I also claim that making your work more reproducible will also lead to greater fame (and possibly fortune, depending on your field, I suppose) and higher quality work. By employing more reproducible methods, you will catch mistakes more often, and when you do catch mistakes, it will be easier to correct them. You will be able to explore more lines of thought and experimentation with an analysis or argumentation, and easily “switch back” if the experiment doesn’t work out. The effort you put into making your work reproducible will result in cleaner, more careful work, which should result in a better publication rate, faster publication, quicker resolution of questions by reviewers, and publication in better journals. You will be able to collaborate more effectively, and amplify your output by engaging in more collaborative work. By making your work more reproducible, it will make it easier for other people to build on what you’ve done. This will increase your citations, and lead to a bigger impact on the field than you would have had otherwise.

I’m making some pretty big claims here. Is it snake oil? I don’t think so, but I also don’t have much in the way of empirical results to give you to back it up, either. My personal experience thus far is that when I’ve taken the time to make my work more reproducible, that has *always* paid off. Every time. No exceptions. I have found that I *always* need to recall certain details or update an analysis after some change, or compare an alternative, etc., etc., and more reproducible methods make all of those things much less painful and less time-consuming.

Of course, I have also spent quite a bit of time fiddling around with different software packages, trying out methods that don’t work all that well, and generally sinking a lot of time into learning a bunch of different tools. One of my hopes for this book and the accompanying tutorials is to greatly reduce this cost of entry for you, gentle reader. It’s also a good way for me to iron things out for myself, so this isn’t entirely altruistic. But the other benefit is that the more people use these

methods, the easier collaboration becomes. I'm a huge fan of the  $\text{\LaTeX}$  system, but it can make collaboration difficult when other people don't use it. The more people use it and other tools that allow more reproducibility, the better, but the cost of entry needs to be sufficiently low.

In the end, though, I'm doing all this myself because I have a firm belief that it will pay off for me personally. I have already seen evidence of this. Whether it works for you is up to you to find out.

### 2.3.2 Doing better science

As I claimed in the beginning section, reproducible methods should also lead to better science, and reproducibility is partly an ethical responsibility for scientists. We live in an exciting but perilous time for science. The rise of the internet into a mature infrastructure, the continuing advances in personal and large-scale computing, great strides in terms of data collection and analysis of types we could barely contemplate 10 or 20 years ago, and so on, provide an exciting new global realm of scientific discovery and collaboration. On the other hand, because of various domains of economic and social change, science is also under attack. Basic scientific education (for both “hard” and “soft” sciences) is in jeopardy in many spheres of public education. Funding for basic science<sup>3</sup> is becoming harder to come by, and more competitive. Several severe cases of academic fraud have received a lot of exposure in the popular press. So while the potential for advances in sciences — including cognitive and social sciences — has never been greater, issues of risk, accountability, and demonstrating value to society loom heavy over the academic landscape.

An absolutely critical piece for both sides of this picture is the issue of trust. On the one hand, as possibilities for new data sources and analyses and collaborations explode, reproducibility is key, in order to maintain trust and order within the scientific community. For example, some advances in statistical methods that have recently become much more accessible (e.g., mixed-effects models, Bayesian analysis) are still not fully integrated or fully understood by researchers in many fields. This means both that some researchers may be employing methods they do not (yet) fully understand, and that some journal reviewers may be resistant to new methods, even if they do not have good reason to be, simply because they are unfamiliar. More transparent, reproducible methods of employing these and other more novel analyses would greatly facilitate the ability to share, evaluate, and critique these methods. In many cases, the publication standards that journals require simply cannot keep up with innovations in analysis, and so a journal article may simply not contain the information that someone would need to know enough about what the authors had

---

<sup>3</sup>“Basic” science is typically described as “science for sciences’ sake.” In other words, science for the sake of increasing understanding. This is contrasted with “applied” science, which is science with an aim of addressing some real-world problem with direct social, economic, or military applications. So for example, theoretical particle physics (e.g., the search for the Higgs boson) or theoretical linguistics (e.g., the search for abstract linguistic universals) are basic science, and developing a particular branch of particle physics to explore some new energy weapon, or applying a theory of universal grammar to problems in machine translation, are examples of applied science. In reality, there's a large continuum. The point here is that in all sorts of domains, it's harder and harder to do research without some kind of applied angle to justify funding.



done in order to evaluate the work. Reproducible methods fill that gap, and make it easier for researchers to share findings and confidence in those findings, whether or not journals require certain things.

More broadly, reproducible methods promote transparency and trust. When people outside the academic scientific community can pick up and replicate analyses and results, it can help break down the “ivory tower” metaphor. It increases accountability and decreases the possibilities for fraud and scandal. If anyone with a computer can re-run and inspect for themselves some important analysis of (e.g.) climate change, voter fraud, economic disparity, health issues, etc., then there is far more opportunity to bring discourse of such topics into the realm of facts and better decisions, and out of the realm of hearsay and fact-twisting partisanship. There is a growing practice of people circulating various plots and graphs through social media like Facebook or Twitter, showing things like debt growth under Democrats vs. Republicans, relationships between gun laws and gun violence, etc., etc. But without an ability to replicate the methods (and directly inspect the data) that went into creating such graphs, there is no real reason to trust any of them. A bar graph can lie just as easily as anything else, especially if you can’t see how it was made.

Within academia, there has been a growing recognition of and dissatisfaction with the problem of replication. The standards for publications in most fields reward studies (by publication and dissemination) for showing effects, while “null results” or failed replications of the same studies may have an extremely difficult time getting published, even though the existence of many well-done failed replications should cast significant doubt on the initial published effects. To make matters worse, replication is more difficult and resource-consuming if the original study is not very thoroughly described. By making replication easier, we can save time and money by reducing the amount of resources wasted on failed replication attempts. There are some interesting current projects trying to address the so-called “file drawer” problem of unreported failed replications, but increased reproducibility is a critical piece of making such efforts successful.

Another hot-button issue that relates to reproducibility is the notion of open access. The current fact of the matter is that a great deal of research is funded by taxpayers, but very few taxpayers are able to access the products of the research (i.e., published papers) without having to pay additional fees to a publisher. Universities (also largely funded by taxpayers) pay often exorbitant fees to access the journals that their faculty are publishing in or need access to for their research. While open access and reproducibility are not necessarily related, they would both thrive in world where open, transparent, reproducible methods were the norm. If researchers could easily share their work in formats that other researchers could more easily inspect, reconstruct, and critique, then peer-review should be facilitated, alleviating some of the pressure for journals to be “gatekeepers” of quality.

The argument, therefore, is two-fold. First, science is facilitated by openness. If your methods provide a barrier to openness, that is an impediment to science. In some sense, this impediment is inherent to the task of communicating our thoughts in a way that other people can interpret them and expand upon them, but the more reproducible your methods are, the more this natural impediment is removed. Second, science exists not in a vacuum but as part of a social contract between the

people that enable scientists (taxpayers, employers, etc.) and the scientists themselves. Reproducible methods should help promote trust and accountability. On efficiency grounds alone, if reproducible methods help researchers do more research and waste less time, then such methods are part of a responsible use of the resources (time and money) given to researchers.

While I do not personally believe that intentional waste and fraud is widespread in academia, there are certain areas that are more vulnerable than others, like research on pharmaceuticals or other big medical issues, and those are inevitably the areas where there is the most to gain or lose by fraud. Would reproducible methods have prevented the terrible debacle of the [Duke cancer trials](#)? Maybe not, but one of the major people involved in uncovering the faults of the study ([Keith Baggerly](#)) believes pretty strongly that reproducible methods would have made it much easier (thus much less time-intensive and much less expensive) to catch the mistakes.<sup>4</sup> And if you think this kind of thing is an isolated case, spend some time reading or listening to Ben Goldacre.<sup>5</sup> Goldacre focuses on the deception and bad science behind many medical studies. He doesn't focus explicitly on reproducible methods, but one of his big themes is the selective withholding of data. If a pharmaceutical company withholds the data from 75% of the trials done on their drug, how confident can we be that the effects reported are real? A fully open, reproducible set of methods is a key piece in making all kinds of findings open to other academics, to policymakers and decision-makers, and to the public.

But what about low-stakes (i.e., most) research in cognitive science, or other disciplines? Do I really think that linguists are insidiously withholding data in order for them to defend their pet theory within Minimalist syntax? Not exactly, but the effects of publication bias are well-known and discussed in many places. In essence, since there is a bias to publish positive findings (i.e., “statistically significant” results), many of the studies that have attempted a replication but found no results will not be published, meaning that the literature will present a skewed, and perhaps completely false picture of the actual pattern of findings. Worse, Uri Simonsohn and colleagues (Simmons, Nelson, and Simonsohn 2011) have shown that many of the typical practices of researchers in psychology and other cognitive sciences result in the ability to find nearly any effect to be “statistically significant,” but *only when these practices go unreported*.

There are two points here. First, flaws and inaccuracies in the ways that research findings are presented are not limited to high-stakes, big-money areas. Even very well-meaning and experienced researchers can inadvertently fall into some of the traps. Research is *difficult*, after all! Second, while reproducible methods are not a panacea, they go a long way towards improving the situation. Back to the issue of ethics, what if your “slightly fudged” results end up sparking a lot of interest, and several people get grants to pursue issues based on those findings? That could be millions of dollars and years of work wasted. Transparent, reproducible methods don't ensure that this could never happen, but they discourage it.

<sup>4</sup>See here for a video and slides of a talk he's given:

[http://videlectures.net/cancerbioinformatics2010\\_baggerly\\_irrh/](http://videlectures.net/cancerbioinformatics2010_baggerly_irrh/)

<sup>5</sup>Who is actually very entertaining! See this talk for example:

[http://www.ted.com/talks/ben\\_goldacre\\_battling\\_bad\\_science.html](http://www.ted.com/talks/ben_goldacre_battling_bad_science.html)

### 2.3.3 Summary of motivations

I've argued for two different kinds of reasons to at least consider reproducible research methods. One is that it will help you *personally* to be more productive, to publish better papers, and to have a bigger impact in your field, the more reproducible your methods are. The other is that reproducible methods help promote and enforce the trust inherent in the social contract of research.

## 2.4 A dissenting opinion and a rejoinder

I would be remiss if I didn't recognize that not everyone is convinced. In a recent unpublished manuscript, Drummond (2012) offers a "dissenting opinion." He outlines four points that he interprets to be the main points of the reproducible research "movement" that he sees growing, and he offers responses to each. The following are lifted verbatim from Drummond (2012, p.2–3):

1. **Claim for reproducible research** It is, and always has been, an essential part of science; not doing so is simply bad science.

**Drummond's rebuttal** Reproducibility, at least in the form proposed, is not now, nor has it ever been, an essential part of science.

2. **Claim for reproducible research** It is an important step in the "Scientific Method" allowing science to progress by building on previous work; without it progress slows.

**Drummond's rebuttal** The idea of a single well defined scientific method resulting in an incremental, and cumulative, scientific process is highly debatable.

3. **Claim for reproducible research** It requires the submission of the data and computational tools used to generate the results; without it results cannot be verified and built upon.

**Drummond's rebuttal** Requiring the submission of data and code will encourage a level of distrust among researchers and promote the acceptance of papers based on narrow technical criteria.

4. **Claim for reproducible research** It is necessary to prevent scientific misconduct; the increasing number of cases is causing a crisis of confidence in science.

**Drummond's rebuttal** Misconduct has always been part of science with surprisingly little consequence. The public's distrust is likely more to with [sic] the apparent variability of scientific conclusions.

I'll address these points in turn. First, Drummond starts by articulating a position pretty similar to the distinction I draw in section 2.1. What I call "replication" — finding results that are consistent with a pattern of results found in a different study — he calls "Scientific Replication," and he seems to admit that this is pretty

important. His point is that the *exact* reproduction of an analysis from the same data set is a pretty novel development in the history of science, and not at all a cornerstone of scientific progress. I don't disagree with his point, but I think it's missing the bigger issue. That is, I agree that strictly speaking, reproducible methods are not *necessary* for the kind of replication that advances scientific knowledge. As Drummond points out, getting the same pattern of results with a near-identical experiment is less impressive (for the purposes of generalization) than getting the same results under a different set of circumstances. But I think this strict reading is missing the benefit that reproducible methods can have. Are they *necessary* or even *sufficient* for scientific progress. No, of course not. Will they *facilitate* scientific progress? I believe so. Drummond's third point argues "no," but I'll get to that shortly.

The second point is a case of pedantic nitpicking, in my opinion. He points out that not everyone is convinced that science proceeds by incremental steps. This is exactly why I also admitted early on that I am not an expert in the philosophy of science. Because I do not think the usefulness of reproducible methods hinges on some acceptance of what the philosophically "correct" view of science is. I don't think there necessarily needs to be a single "correct" view. All I claim is that transparency of methods and increased ability for people to understand and reproduce each other's work will only grease the skids, whatever the actual trajectory of scientific progress is. I don't think any of the benefits of reproducible research depend on a notion of science as incremental and cumulative. Maybe some proponents believe otherwise, but at least I propose that one can find value in reproducible methods without adhering to this (or any other) narrow view of what "science" is.

The third point is what I think is really at the heart of Drummond's complaint, and I think the most legitimate concern. In short, he seems worried that if journals start to impose arbitrary constraints on what the authors need to provide, then the system of peer-review will get clogged with data and software, and there will be a much greater burden on authors, reviewers, and editors. It just sounds like a big hassle to him, and not worth it. He also thinks it will foster distrust, because it would treat everyone as if they had something to hide, by forcing them to put data and code out into the open.

There are several issues here. One of the more disturbing comments Drummond makes is that submitting code will "simply result in an accumulation of questionable software (p. 5)." This is disturbing to me, because if one really believes that the software that generated the results of a paper is questionable, why would you ever trust the results reported in the paper? I think what he means is that most pieces of code that researchers throw together to do their analyses are not the cleanest, most efficient, most generally useful pieces of software. I think this is probably fair, given that I would think that most researchers are not also expert software developers, and are thus putting together scripts that are designed to get the answers they need, not produce general-purpose software that could be used in a "production-level" context. However, I think if researchers took the reproducible research goals seriously, they would be producing code that more cleanly replicates and produces the results of the paper. If you still have doubts about the code, at least you are able to actually inspect the code and confirm or disconfirm your doubts. If you think most people's

code is crap, and you don't ask them to provide their code, then you really have no reason to read any of the papers in the first place.

This segues into Drummond's point about reviewers. I think he imagines a situation where reviewers are called upon not only to review the paper, but to review the (messy, "questionable") code. I think this is only one way it could play out. If we imagine a really reproducible paper like the kind I will discuss in this book, the paper *cannot exist* without the code. If you take the code away, then there will be no tables of results, no figures, no reporting statistics at all. In this kind of set-up, as long as the reviewers believe the results, then there's no reason to question the code, because by definition, the authors have produced code that results in the paper. Only in the case where a reviewer is familiar with the type of analysis and wishes to know about a detail not reported in the main body of the paper (which happens *very* frequently, in my limited experience), those details can be obtained by inspecting the code. Maybe the authors would be asked to point out where in the code such-and-such can be found, or why they chose to implement something in a particular way, but if the research is *really* reproducible, the reviewer should be able to verify or falsify any particular concern they have. But if the reviewer doesn't care about the code, I don't see how it would be any different from the current situation if they merely trust that the authors' code does what they say it does, and trust the results as reported in the paper. All it would change is provide a much more thorough and expedient way for reviewers' questions and concerns to be addressed. But I think Drummond's point is well-taken that journals may not always come up with the best policies regarding reproducible research, so some effort should be taken to get things right, if a journal decides to enforce some degree of reproducibility.

Another technical downside is that if code and data are shared in addition to papers, this could represent an exponential increase in the burden on journals to host such material. This is a legitimate concern, since it starts to put a very real price tag on reproducible methods. But since publishers currently charge for access to an article *in perpetuity*, it doesn't seem unfair to me that they would therefore have the responsibility of hosting the materials in perpetuity. Maybe some more obscure articles' materials get "retired" to archives that don't need to be kept in on-demand storage on a disk somewhere. But I think if this became the norm, there would be some effort in figuring out how to host things in a cost-effective way. And again, this is only a downside towards the mandatory *publishing* of reproducible methods. All of my arguments above are based on the idea that people would use reproducible methods first for their own private benefit (regardless of journal requirements), and second in order to facilitate sharing their work (again, regardless of whether that sharing is done through a journal publisher or not). So this concern about clogging publisher's servers is not a general problem.

The bigger issue, from my perspective, is the burden that reproducible methods put on *researchers*. If it were automatic and easy to implement reproducible methods, then everyone would be doing it already. It's my belief that implementing reproducible methods, and learning the techniques/software that enable the implementation are all worth the time and energy expended. But it's hard to say if it's really worth it to everyone, if some people would benefit more than others, or what. And it's hard to put a precise price tag on it. Most people were not initially trained

on the tools that I will recommend for an implementation of reproducible measures. That means that virtually any research in the fields I work it will have some learning curve in order to start doing more and more reproducible research. And changing work habits is difficult!

To this objection, I can only say that my personal experience has been that it's more than worth it. Some people have naturally gravitated towards these kinds of techniques, and they will naturally appeal more to some people than to others. This book and the accompanying tutorials are a big experiment. Maybe they will help enable people to implement reproducible research, where they never would have otherwise. Or maybe it will just help a subset of people, who may have eventually stumbled on some similar ideas anyway, to find these techniques faster. Or maybe it will only help me! This is still an empirical question, so I'm writing the book to find out the answer.

Drummond's final point seems to be that misconduct in science is not a new phenomenon, and in fact scientific progress is remarkably robust against the bad effects of misconduct. Somewhat paradoxically, he seems to suggest (referring to the Duke cancer trials) that perhaps the solution is that we should be "more skeptical about the results of scientific experiments (p. 7)." This is really perplexing to me, because he seems to be suggesting that requiring reproducible research would increase distrust in the scientific community, but then suggests that instead of reproducible methods, we should simply distrust more science? The biggest point about the Duke trials is not that reproducible methods would have prevented the problem (which was apparently largely due to some very basic problems in data formatting: having one column displaced by a row with respect to the values in the other columns), but if we are to believe Baggerly, *thousands* of man-hours could have been saved if they had not had to go through the excruciating work of reconstructing methods that were not provided. Skepticism is natural (or acquired) tendency of all good scientists; reproducible methods merely make it easier to satisfy one's skepticism.

To put this another way, it *might* be the case that the scientific process is robust against the "base rate" of misconduct (though one has to wonder when you hear Ben Goldacre talk about the pervasiveness of it in some medical sciences). But even in the examples that Drummond gives, reproducible methods would have helped enormously. He brings up the claim of "cold fusion" by Pons and Fleischmann in 1989. And he mentions that the impact of this study were "short lived." But in doing so, he mentions that "many scientists did attempt to reproduce the result and failed (p. 7)." Is the implication that the time and money spent in all those failed replications was negligible? Those scientists had nothing better to do? That's hard for me to believe. Of course, even with reproducible methods, some time and expense would be lost in trying to replicate false finding like this, but the point is that (almost by definition) the more reproducible the methods are, the less time and money would be wasted in the process.

In the end, I think Drummond's points mostly stack in *favor* of reproducible research. While reproducibility is not a *necessity* of science, it facilitates what he calls "Scientific Replication" (and what I call just "replication" in section 2.1) by providing more transparency and facilitating the process of replication. The fact that not everyone has the same view of what the "scientific method" is also comes

out favoring reproducible research, because perhaps an author is making some set of judgment calls as to which of his results are “important” or not. If this is not transparent in a way that allows others to examine the impact of alternatives, then the work would be of little use, or even misleading to someone with different views. In other words, greater transparency and reproducibility benefits more viewpoints, not fewer. Finally, the consistent presence of misconduct is not likely to disappear from science, no more than from any other sphere of human endeavor. However, reproducible measures should both discourage “accidental” misconduct, and make all kinds of misconduct and Bad Science (to use [Goldacre’s term](#)) easier to catch and less onerous to correct.

## 2.5 Some core values

We’ve already discussed replication, but what else is at stake here? Why should we care about reproducibility?

## 2.6 Who is this book for?

Full disclosure: this book is primarily a selfish effort. I am putting this system together to improve my own methods and productivity, and having it all written down in an organized way is a helpful way for me to put my thoughts together into a system that’s documented and clear. But this point also gets to the heart of it: I believe that adopting more reproducible methods will lead not only to Better Science, as alluded to by the intro above, but a more efficient and productive workflow. In other words, I believe that Really Reproducible Research (more on what I mean by that in section ??) is not just The Right Thing for Science, but The Way to Get More Done and Published.

Because of this last point, I intend my primary audience (after myself) to be academic scientists. It also follows from my self-centered goals that the specifics will be geared towards academics working in linguistics, and psycholinguistics. However, I expect that people from other fields could find the discussion and implementation helpful, and easily adapted. To frame it another way, ask yourself the following questions:

- Have you ever picked up an old paper of yours and wished you could remember some detail of how the data was collected/analyzed?
- Have you ever needed to update a figure/table/statistical analysis after a change in the data or analytic procedure?
- Have you ever read someone else’s paper and wished you could see exactly how they did their analysis?
- Have you ever been frustrated in how much time you spend chasing down and re-typing/re-formatting the same set of references across multiple papers?

- Have you ever had a request for your data/analysis/other details from a paper and shuddered at the effort needed to share it in an accessible way?
- Have you ever had a problem with inconsistency in a paper, where the stats are from one data set, but the figures (or summary tables, or stats in another section) are taken from a different data set (e.g., after some additional data, or some additional data-cleaning, or something)?
- Have you ever lost track of what kinds of data-cleaning (outlier trimming, transformations, missing data, etc.) have been performed on a data set, and which ones were applied to results in a given paper or presentation?
- Have you ever gone through some laborious data-organization or analysis process (e.g., sorting/labeling/tweaking/cleaning things in Excel by hand), only to have to do it over and over when you discover mistakes or when the data changes in some way?
- Have you ever taken hours to carefully construct some kind of complex figure or diagram by hand (e.g., graph, flowchart, theoretical model, syntactic tree), only to have to re-format it for a journal submission, or a talk handout, or a PowerPoint presentation, or some other formatting issue?

If you are still in the early stages of your career, and are unsure about whether anything like this may happen to you, just do a quick poll of your advisor, more advanced students, etc. If none of these things apply to you, you are likely either (a) not an academic, (b) an academic in a non-scientific field, or (c) already doing a fantastic job doing reproducible research. But if any of these things apply, and you like the idea of doing something about it, then my hope is that this book will help.

Finally, this book does *not* assume you already have facility with programming, etc. Many of the implementation tools I'll discuss in Part II involve some level of savvy in programming, using command-line tools, and other things normally associated with steep learning curves. My intention is to present arguments for why these tools are worth the effort to learn and use, but I will start out assuming that the reader is a user of commercial products like the Windows or Mac operating systems, programs like Microsoft Word and maybe a little of Excel, and a graphical stats package like SPSS or JMP, if anything. My hope is that this book could be picked up by people early in their academic careers and applied as they go. My greater hope is that the ideas will be appealing enough and the implementation easy and effective enough that even experienced, established academics could find some utility in improving some of their habits and/or tools. People tend to get entrenched, though, so I'm not holding my breath on the latter group. But one can hope...

## 2.7 Goals

So what exactly do I hope to accomplish with this book? What exactly should you, the reader, expect to be able to get out of it? To return to my motivations and audience, I would like to enable linguists and psycholinguists (and others, perhaps)



to produce Really Reproducible Research, from soup to nuts. Currently, there are bits and pieces of resources and ideas spread around multiple fields and websites and repositories. My purpose here is to collect what I think are the best of the best, and assemble them into a system of principles, tools, and methods that will work well together for a “complete” system of Really Reproducible Research. Additionally, many resources on reproducible research (including the website of that name) are geared primarily towards computational or statistical work, and their principles can be summed up as “share your data, include your code, and make your code legible to others.” These principles are certainly relevant, but they don’t capture the whole messy system of producing scientific research that is truly reproducible.

Therefore, on the one hand I aim to present a more general discussion and system for carrying out reproducible scientific research beyond “include your code,” and on the other hand, I aim to provide a very specific configuration of tools geared towards carrying out reproducible research in linguistics and psycholinguistics. The book is organized with these goals in mind. In the rest of this first part of the book, the discussion will remain tool agnostic in general, although the principles discussed will end up favoring some kinds of tools over others. The goal of this part of the book is to lay out the principles and concepts for what it means to carry out Really Reproducible Research, and what the benefits and drawbacks might be.

The second part of the book makes this more concrete by spelling out a particular implementation. The implementation is partly a set of software recommendations and partly a set of workflows, procedures, and methods for doing typical research tasks in a way to support Really Reproducible Research. There will be plenty of room for customization, because I don’t expect that any two researchers will want to do things in exactly the same way, but the goal is to be as specific and concrete as possible, so that you are not left wondering about how to connect the dots. Some suggestions for alternatives will be included, but I will focus on tools that I use and that I think are best for the job, and I will not go through an exhaustive review of tools I’m less familiar with.

Finally, the third part of the book is a set of tutorials designed to enable you to use the tools in the implementation. For example, I discuss Emacs and org-mode as major tools in the implementation. Most people are not Emacs or org-mode users. Both Emacs and org-mode have extensive documentation, including books, tutorials, and tons of articles spread across the web. However, existing documentation is both more and less than what you would need to implement the system I describe in Part II. They are *more* in the sense that there are *tons* of functions in both Emacs and org-mode that may be great features and very useful, but not relevant or necessary for the system I outline here. Existing tutorials and documentation also provides *less* than what I do here, in the sense that using these tools in the specific way I describe in Part II may not be obvious, even if you worked your way through the general manuals or tutorials already available. In other words, my goal is not to teach you Emacs for all general purposes, but rather to teach you how to use Emacs in the system of Really Reproducible Research. Even if you know Emacs, there may be something useful for you in my tutorials, but there will also be lots more to learn about Emacs after you’ve mastered my tutorials. And as I mentioned in the previous section, I will assume that you have experience with Word, and that’s about it, so

you should approach the tutorials with minimal anxiety.

With these three parts, my ultimate goals are to (1) describe what I think are the critical elements of reproducible research and convince you that these are worthy and useful goals, (2) describe a concrete system for achieving reproducible research in the real world of working academia, and (3) enable readers with no knowledge of the tools I describe to learn and apply these tools in their own personal approach to reproducible research. This way, I hope that the end result of this book is not just a series of suggestions, but the actual means to implement and improve upon my idea in your own work.

## Chapter 3

# Dimensions of reproducibility

Thus far, I have only hinted vaguely at what *reproducible* really means. I have frequently used the phrase “Really Reproducible,” implying that some values of “reproducible” may be less than desired. In this chapter, I will tackle the definition of “reproducible” in a more systematic way. I argue that “reproducible” is a continuum, and even more so, a two-dimensional continuum. With this understanding, we are in a better position to zero in on appropriate principles and standards in defining a target for what Really Reproducible means.

The first approximation of reproducible comes from the general idea in the scientific method that results should be able to be replicated. That is, I can present some data and an analysis, and in order for it to qualify as “good science,” it should be possible for someone else to also collect similar data and perform a similar analysis and get (generally) the same result. Put another way, if no other scientist/lab in the world can get the same results you can, that’s a big problem.

But when you start thinking about this seriously, it’s apparent that this raises two questions. Reproducible by who? How “similar” must the data, analysis, and results be for it to qualify as “reproducible”? These are what I call the *domain* and *range* of reproducibility.

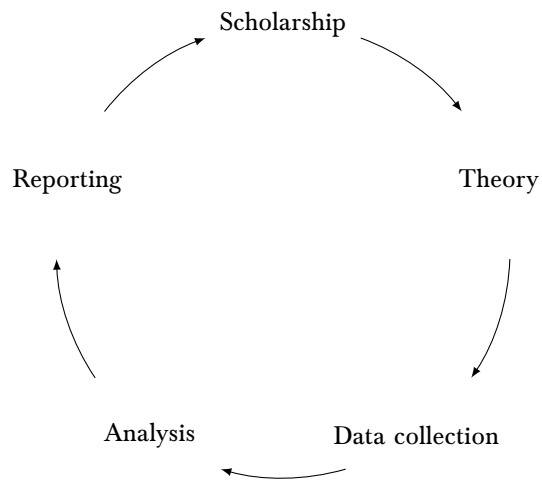
### 3.1 Domain: the audience

The first dimension of reproducibility is the *domain* or the *audience*. In short, *who* do you expect to be able to reproduce your work? On one end of the continuum is yourself. If you cannot reproduce your own work, how could you expect anyone else to? On the other end of the continuum is virtually anyone in the world, which is probably almost always impossible. Figure 3.1 illustrates a few important values for the domain.

- 
- 3.2 Range: the precision**
  - 3.3 The minimum**
  - 3.4 Publication standards**
  - 3.5 Wide dissemination**

## Chapter 4

# The cycle of research





## **Chapter 5**

# **Scholarship**





## **Chapter 6**

# **Theory**



## **Chapter 7**

### **Data collection**



## **Chapter 8**

# **Data analysis**



## **Chapter 9**

# **Reporting**





## **Chapter 10**

## **Conclusions**





# Chapter 11

## Implementation

### 11.1 Guiding principles

#### 11.1.1 Free and open source

#### 11.1.2 Cross-platform

#### 11.1.3 Stable

#### 11.1.4 Well-documented

#### 11.1.5 Customizable

### 11.2 Summary of tools

#### 11.2.1 Emacs

#### 11.2.2 Org-mode

#### 11.2.3 Git

#### 11.2.4 L<sup>A</sup>T<sub>E</sub>X

#### 11.2.5 Python

#### 11.2.6 R

#### 11.2.7 (Emacs) Lisp

### 11.3 Scholarship

### 11.4 Data collection

### 11.5 Data analysis

### 11.6 Sharing

### 11.7 Collaboration

### 11.8 Putting it all together

# Bibliography

- Buckheit, Jonathan B. and David L. Donoho (1995). *WaveLab and reproducible research*. Tech. rep. 474. Department of Statistics, Stanford University.
- Drummond, Chris (2012). “Reproducible Research: a Dissenting Opinion”. posted at <http://cogprints.org/8675/>.
- Fomel, Sergey and Jon F Claerbout (2009). “Reproducible Research”. In: *Computing in Science & Engineering* 11.1, pp. 5–7.
- Gentleman, Robert and Duncan Temple Lang (2004). “Statistical analyses and reproducible research”. In: *Bioconductor Project Working Papers*.
- Gentleman, Robert et al. (2005). “Reproducible research: A bioinformatics case study”. In: *Statistical Applications in Genetics and Molecular Biology* 4.1, p. 1034.
- Knuth, Donald E. (1984). “Literate programming”. In: *The Computer Journal* 27.2, pp. 97–111.
- Peng, Roger D. (2008). “Caching and distributing statistical analyses in R”. In: *Journal of Statistical Software* 26.7, pp. 1–24.
- Peng, Roger D (2009). “Reproducible research and Biostatistics”. In: *Biostatistics* 10.3, pp. 405–408.
- Schwab, Matthias, N Karrenbach, and Jon Claerbout (2000). “Making scientific computations reproducible”. In: *Computing in Science & Engineering* 2.6, pp. 61–67.
- Simmons, Joseph P, Leif D Nelson, and Uri Simonsohn (2011). “False-Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant”. In: *Psychological Science* 22.11, pp. 1359–1366.
- Stodden, Victoria C (2009). “Enabling reproducible research: licensing for scientific innovation”. In: *International Journal of Communications Law and Policy* 13.
- (2010). “Reproducible Research: Addressing the Need for Data and Code Sharing in Computational Science”. In: *Computing in Science & Engineering* 12.5, p. 8.
- (2011). “Trust your science? Open your data and code”. In: *Amstat News* 409, pp. 21–22.
- (2012). “Reproducible Research: Tools and Strategies for Scientific Computing”. In: *Computing in Science & Engineering*, pp. 11–12.