

The Scientific Workflow

John Paul Minda, PhD

September 4, 2019

1 Introduction

This is a guide for carrying out cognitive psychology and cognitive science research in my lab. Keep this workflow in mind when you are planning, conducting, analyzing, and interpreting scientific work. You may notice two themes that seem to run throughout the plan: **documenting** and **sharing**. That's the take home message: Document everything you do and share your work for feedback (share with the group, your peers, the field, and the public). Not every project will follow this outline, but most will.

2 Theory & Reading

The first step is theory development and understanding the relationship of our work to the relevant literature. We're involved in cognitive science and we develop and test theories about how the mind forms concepts and categories. We usually work from two primary theories. [Prototype / exemplar theory](#) both of which deal with category representations and the [multiple systems theory](#) (COVIS is an example) which addresses the category learning process and rule use. You should keep up with developments in the field using Google Scholar alerts and its recommendations. I check every week and I recommend that you do as well. We want to test the assumptions of these theories, understand what they predict, test their limitations and contrast with alternative accounts. We're going to design experiments that help understand the theory, the models, and make refinements and/or reject some aspects of our theories.

- Use Google Scholar to find updates that are important for your research.
- Save papers in [Paperpile](#) and annotate as needed.
- Document your work in Google Docs (or another note taking app).
- Share interesting papers and preprints with the whole lab group in the relevant channel(s) in Slack.

3 Hypotheses Generation

Hypotheses are generated to test assumptions and aspects of the theory and to test predictions of other theories. The hypothesis is a formal statement of something that can be tested experimentally and these often arise from more general **research questions** which are often broad statements about what you are interested in or trying to discover. You might arrive at a research question or an idea while reading a paper, at a conference, while thinking about an observation you made, or by brainstorming in an informal group or lab meeting. Notice that all of these assume you that put in some time and effort to understanding the theory and then allow some time to work over ideas in your mind, on paper, or in a computer simulation.

- Work on hypothesis generation in lab meetings, our advisory meetings, and on your own.
- Document your thoughts in Google Docs (or your own notes on paper, OneNote or Evernote).
- Share insights in lab meetings and in the relevant channel in Slack.

4 Design the study/experiment

Concurrent with hypothesis generation is [experimental design](#). In most case, we are designing experiments to test hypotheses about category representation and category learning and/or the predictions of computational models. We want to test hypothesis generated from theories and also carry out exploratory work to help refine our theories. *Avoid the temptation to put the cart before the horse and come up with experiments and studies that will produce an effect for its own sake.* We don't just want to generate effects.

The design comes first. Consider the logic of your experiment, what you plan to manipulate, and what you want to measure. Avoid the temptation to add in more measures than you need, just to see if there's an effect. For example, do you need to add in 2-3 measures of working memory, mood, or some demographic information just to see if there's an effect there? If it's not fully justified, it may hurt more than help because you have non-theoretically driven measures to contend with. I've been guilty of this in the past but it always comes back to haunt me.

- Work on experiment generation in lab meetings, advisory meetings, on your own.
- Document your work and ideas in Google Docs or a note taking app that you can share.
- Use [G*Power](#) to estimate correct sample size.
- Use [PsychoPy](#) or Qualtrics to build your experiment.
- Test these experiment protocols often.
- Develop a script for research assistants who will be helping you carry out the study.
- Share insights in lab meetings and in the relevant channel in Slack.
- Organize tasks and chores in the relevant Trello board for your project.

5 Analysis Plan & Ethics Protocol

This is where we start to formalize things. An **analysis plan** will link together the hypothesis and the experimental design with the dependent variables and/outcome measures. In this plan, we'll describe and document how the data will be collected, visualized, analyzed, stored and shared. This plan should describe how we will deal with outlier data, missing data, data from participants who did not complete the experiment correctly, experimenter error, malfunction, etc. This plan can include tentative predictions derived from a model and also a justification of how we intend to analyze and interpret the data. This plan can (and probably should) be pre-registered with [OSF](#), which is where we'll plan to share the data we collect with the scientific community.

At the same time we also want to write an **ethics protocol**. This is a description of our experiment, the research question, and procedures for the University REB. This will also include standardized forms for information and consent, a policy for recruitment, subject safety, data storage and security. The REB has templates and examples, and our lab Slack channel on ethics can include examples as well. Use templates when ever possible.

Both of these documents, the analysis plan and the ethics protocol, should describe *exactly* what we are doing and why we are doing it. They should provide enough information that someone else would be able to reproduce our experiments in their own lab. These documents will also provide an outline for your eventual method section and your results section.

- Document your analysis plan and ethics protocol work in Google Docs.
- Link these documents to the project sheet or Trello board for your project.
- Share in the relevant channel in Slack.

6 Collect Data

Once the experiment is designed, the stimuli have been examined, we're ready to collect data or to obtain data from a third party (which might be appropriate for model testing). Before you run your first subject, however, there are some things to consider. Take some time to run yourself through every condition several times and ask other lab members to do the same. You can use this to make sure things are working exactly as you intend, to make sure the data are being saved on the computer, and to make sure the experiment takes as long as planned.

When you are ready to collect data for your experiment:

- Meet with all of your research volunteers to go over the procedure.
- Book the experiment rooms on the Google Calendar.
- Reserve a laptop or laptops on the Google Calendar.
- Recruit participants through SONA or flyers.
- Prepare the study for M-Turk or Prolific
- Use our lab email for recruitment.

After you have run through your experiment several times, documented all the steps, and ensured that everything is working exactly as you intended, you are ready to begin. While you are running your experiment:

- Document the study in Google Docs, Trello, and/or Slack (as appropriate)
- Make a note of anything unusual or out of the ordinary for every participant in a behavioural study.
- Collect signatures from participants if you are paying them.
- Data should be stored in text files that can be opened with Excel or Google sheets or imported directly into R. Be sure these are linked to the project sheet.
- Make sure the raw data are labelled consistently and are never altered.
- Be sure to follow the data storage procedures outlined in the ethics protocol.

7 Data Management

Your data plan should specify where and how to store your data. While you are collecting data you should be working on a script in R (or Python) to extract and summarize the raw data according to your plan. When

you reach the planned sample size, ensure that all of that data are secure and backed up and do an initial summary with your R script.

As you work on summarizing and managing your data:

- Make notes in the project sheet and/or Trello board about where the data are stored
- Document your steps in an [R Notebook](#) (or [Python Notebook](#)).

8 Plots & Stats

Remember the photo of [Dr. Katie Bouman](#), then a postdoc, when she first saw the rendering of the first photos of a black hole that her algorithms generated? That's the best part of science: seeing your data visualized for the first time. When you have completed your experiment and taken care of the data storage and basic processing, it's time to have fun and see what you discovered. The analysis plan is your guide and your analysis plan describes how you want to analyze the data, what your dependent variables are, and how to conduct statistical test with you data to test the hypothesis. But before you do any statistics, work on visualizing the data. Use your R notebook to document everything and generate boxplots, scatter plots, or violin plots to see the means, medians, and the distribution for the data.

Because you are using R Notebooks to do the analysis, you can write detailed descriptions of how you created the plot, what the plot is showing, and how we should interpret the plot. If you need to drop or eliminate a subject's data for any reason, exclude them in from data set in R, do not delete the data from the raw data file. Make a comment in the script of which subject was dropped and why. This will be clear and transparent.

You can also use R to conduct the tests that we proposed to use in the analysis plan. This might be straightforward ANOVA or t-test, LME models, regression, etc. Follow the plan you wrote, and if you deviate from the plan, justify and document that exploratory analysis.

If you are fitting a decision boundary model to your data, make sure you have the code for the model (these will be on my [GitHub](#)) and you should do your modeling separately from the behavioural analysis. The GLM models are saved as R scripts but you should copy or fork to your R-Notebooks for your analysis so you can document what you did. Make sure that you develop the version for your experiment and that the generic model is not modified.

If you are fitting a prototype or exemplar model, these have been coded in Python. Use Python 3 and a basic text editor or JupyterLab. JupyterLab might be better as it's able to generate markdown and reproducible code like R Notebooks.

- Follow your analysis plan.
- Consult with me or your peers if you notice any unusual patterns with anything.
- Make notes in the project sheet and/or Trello board about what analyses you've completed.
- Document your steps in an [R Notebook](#) (or [Python Notebook](#)).
- If you drop a participant for any reason, indicate in the comments of your R script (or other notes). We want this information to recorded and transparent.

9 Present and Explain Your Work

While you working on your analysis, you should present the interim work often in lab meetings for the rest of the group and we can discuss the work when we meet individually. The reason to present and discuss often is to keep the ideas and work fresh in your mind by reviewing manageable pieces of it. If you try to do too much at once, you may miss something or forget to document a step. Go over your work, make sure its documented, and then work on the new analyses, and repeat. You should be familiar with your data and your analysis so that you can explain it to yourself, to me, to your peers, and eventually anyone who reads your paper.

Use the following guidelines for developing your work:

- Make your best plots and figures.
- Present these to the lab on a regular basis.
- Use [R Pubs](#) to share summary work instantly.
- Keep improving the analysis after each iteration.
- You should always have 8-10 slides that you can present to the group.
- Document your work in R Notebooks, Google Docs, Trello, and Google Slides.

10 Write papers around this flow

The final step is to write a paper that describes your research question, your experimental design, your analysis and your interpretation of what the analysis is. A scientific paper, in my opinion has two important features:

1. **The paper should be clear and complete.** That means it describes exactly what you wanted to find out, how and why you designed your experiment, how you collected your data, how you analyzed your data, what you discovered, and what that means. Clear and complete also means that it can be used by you or by others to reproduce your experiments.
2. **The paper should be interesting.** A scientific paper should be interesting to read. It needs to connect to a testable theory, some problem in the literature, an unexplained observation. It is just as long as it needs to be.

I think the best way to generate a good paper is to **make good figures**. Try to tell the story of your theory, experiment, and results with figures. The paper is really just writing how you made the figures. You might have a theory or model that you can use a figure to explain. You can create clear figures for the experimental design, the task, and the stimuli. Your data figures, that you made according to your analysis plan, will frame the results section and a lot of what you write is telling the reader what they show, how you made them, and what they mean figures. *Writing a scientific paper is writing a narrative for your figures.*

Good writing requires good thinking and good planning. But if you've been working on your experiment according to this plan, you've already done a lot of the thinking and planning work that you need to do to write things out. You've already made notes about the literature and prior work for your **introduction**. You have notes from your experimental design phase to frame the experiment. You have an ethics protocol for your **methods** section and an analysis plan for your **results**. You'll need to write the **discussion** section after you understand the results, but if you've been presenting your 8-10 slides in lab meeting and talking about them you will have some good ideas and the writing should flow. Finally, if you've been keeping track of the papers in Paperpile, your **reference** section should be easy.

11 Submit the paper

The final paper may have several experiments, each around the theme set out in the introduction. It's a record of what we did, why we did it, and how. The peer reviewed journal article is the final stage, but before we submit the paper we have a few other steps to ensure that our work roughly conforms to the principles of [Open Science](#), each of which should be straightforward if we've followed this plan.

- Create a publication quality preprint using the lab template. We'll host this on [PsyArXiv](#) (unless submitting a blind ms.)
- Create a file for all the stimuli or materials that we used and upload to OSF.
- Create a data archive with all the raw, de-identified data and upload to OSF.
- Upload a clean version of your R Notebook that describe your analyses and upload to OSF.

The final steps are organized around the requirements of each journal. Depending on where we decide to submit our paper, some of these may change. Some journals will insist on a word doc file, others will allow for PDF. In both cases, assume that the Google Doc is the *real* version, and the PDF or the .doc files are just for the journal submission. Common steps include:

- Download the Google Doc as a MS Word Doc or PDF.
- Create a *blind* manuscript if required.
- Embed the figures if possible otherwise place at the end.
- Write a cover letter that summarizes paper and why we are submitting.
- Identify possible reviewers.
- Write additional summaries as required and generate keywords.
- Check and verify the names, affiliations, and contact information for all authors.
- **Submit and wait for 8-12 weeks!**