

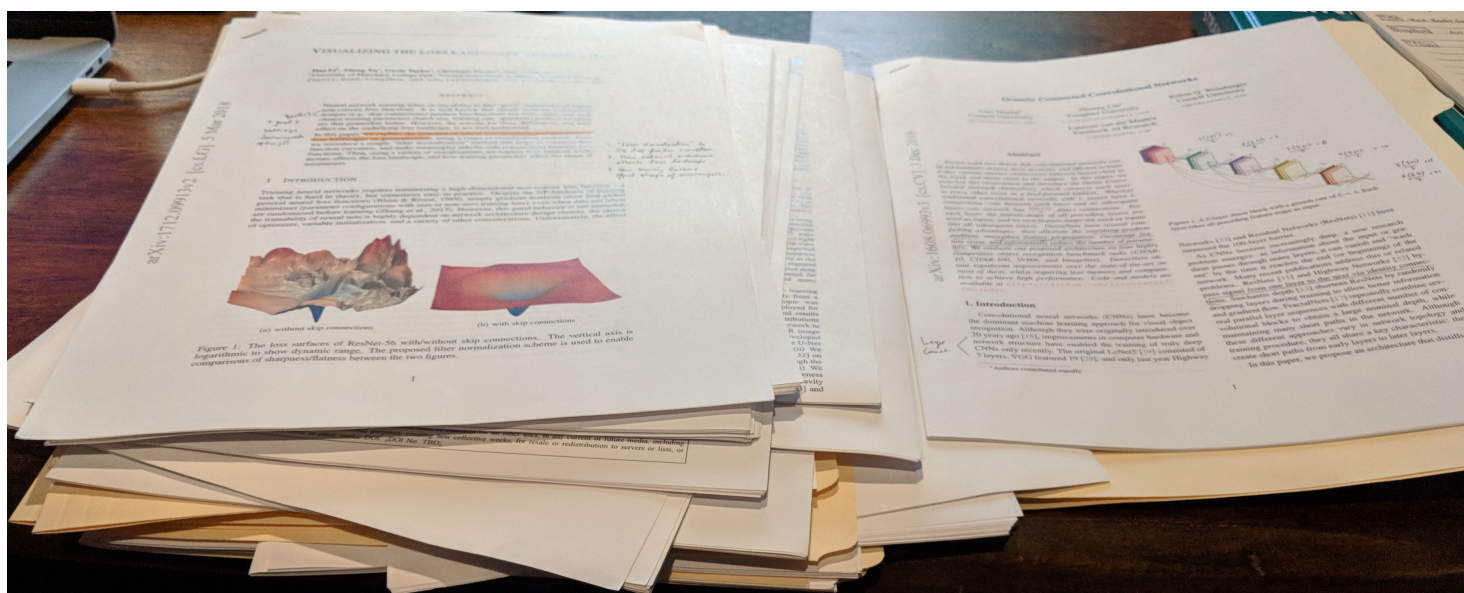
Guide to Reading Academic Research Papers

Learn to tackle this laborious process with a systematic approach!



Kyle M Shannon [Follow](#)

Jul 22, 2018 · 13 min read



My desk full of papers. Slightly shuffled around for dramatic purposes of course.

Working in data science and machine learning is an exciting and challenging field. New techniques and tools are constantly percolating and honestly, it can feel overwhelming. Many of these new developments are found and first revealed in academic research articles. Extracting knowledge from these articles is difficult because the intended audience of these papers tend to be other researchers. Yet in order to stay current reading papers is an essential skill — luckily one that can be improved with diligence and practice.

In graduate school, you get good (should get good...) at reading papers and ingesting research. Not everyone will get training in this skill, that doesn't mean you shouldn't benefit from the knowledge these papers hold. Public tax money is how most of this

research gets funded anyways! The goal here is to democratize academia, just a bit, and provide you with a scaffolding to apply when walking through a paper.

The way I read papers is not significantly unique, but it is effective and has served me well. Keep in mind it is not the only way, many techniques exist and as you read more and more I am sure you will find your own unique style.

This guide is broken down as follows:

1. Learning this skill **will** help you! I promise
2. So I hear reading a paper is difficult. Why?
3. How are papers typically organized?
4. My “*bullet proof*” approach to reading papers
5. Tools to help you get the job done

Why Learn to Read Papers?

Reading papers certainly builds character because It often takes many hours and there is no guarantee you walk away with the whole story. This is not to disparage you, but merely to be open and transparent. Reading papers is difficult, there are no two ways about it. Advances in fields such as machine learning, deep learning, data science, databases, and data engineering often come in the form of academic research, whose language is that of academic papers. Think about some of the techniques you might use: Convolutional Neural Networks, PCA, and AdaBoost (even Deep Boosting!). These all came out of research, and yes they all have papers. Also, consider that there are many papers on the application and use of these techniques and when you are trying to solve a specific problem, these papers can be critical. Beyond staying current with research it is also worth traveling to the past and giving older papers a read. You will learn so much. I promise.

Looking at the field of deep learning it seems as though a new critical paper is coming out every few days or weeks. The only way to stay on top of it is to get a hold of the paper and give it a read.

Where the Difficulties Arise...

Here is a figure from a 2017 scientific paper¹ by Hubbard and Dunbar, about reading scientific papers. Scientific Paper inception!

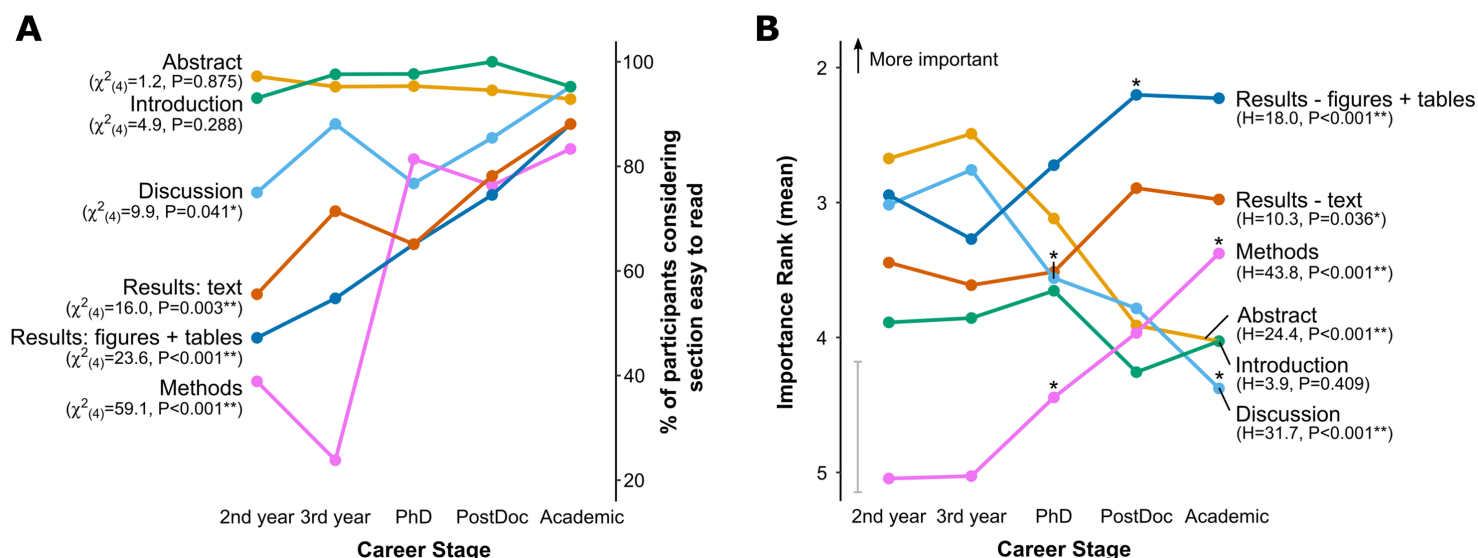


Fig 2. Different sections of scientific papers are considered easy to read and important at different stages of academic careers.

A: The proportion of participants considering a section easy to read (presented as ‘Somewhat easy’, ‘easy’ ‘very easy’ combined) as a function of career stage. Results of Chi-square tests are indicated on the left hand side. **B:** The mean importance rank of sections as a function of career stage. Error bars are omitted from individual points for clarity, with the sole error bar in grey representing the largest 95% confidence interval for any of the data points. Asterisks above data points indicate significant differences in response compared with the previous career stage as determined by Mann-Whitney post-hoc tests.

One unsurprising result indicates the further an *academic* progresses into their career, the easier they find each section of a paper to read. An interesting point is how the various career stages view the importance of each section. Methods, Results and figures seem to be very important, ostensibly because as *academics* they have greater skill in their field, allowing them to be critical of a paper’s methods. It also means they know their field very well, therefore, the introduction and abstract have less importance. Early stage PhD students find the methods, results, and figures fairly difficult to understand. This makes perfect sense as those are the areas of a paper that

require the most knowledge of a field to get through. You are likely to have a similar experience.

What is it exactly that makes going through this process so difficult and time consuming?

- Authors tend to assume significant background knowledge from readers
- Academic syntax is dense and thus difficult for readers to parse
- Mathematical expressions are typically condensed and equations reordered for concision, often skipping steps in derivations
- Substantial knowledge gaps are filled if a reader has read cited papers (sort of like — you need experience to get a job, but need a job to get experience!)
- Not **all** conclusions drawn are correct. Small sample size and power, poor study design, researcher bias, and selective reporting ensures that you **must** be a critical reader!

Clearly there is a lot to consider when reading a paper. Scared? Time to lighten the mood. Here is a hilarious article written on the horrors of reading papers by Dr. Adam Ruben from Science. It shows even scientists can agree that papers are both difficult to read and given how dense they are, will keep you regular.

Think about this, the more papers you read, the more you will learn and the faster this process of reading becomes. Trends start cropping up into plain view, and you begin to gain insight into the scientific method, understand what certain authors and groups are working on, and form an appreciation for the field you are learning about. Over time all of this knowledge and skill builds into your ability to read papers quicker, more efficiently and with greater success. Learning to read papers is akin to learning to eat. It is messy at first, and your palette is not very well developed. But over time your eating experience enhances and you learn more about what you like and don't like and when a chef's meal is good and poor.

How Papers are Organized

Good news here. The overwhelming majority of papers follow, more or less, the same convention of organization:

1. Title: Hopefully catchy, possibly sexy! Includes additional info about the authors and their institutions
2. Abstract: High level summary
3. Introduction: Background info on the field and related research leading up to this paper
4. Methods: Highly detailed section on the study that was conducted, how it was set up, any instruments used, and finally, the process and workflow
5. Results: Authors talk about the data that was created or collected, it should read as an unbiased account of what occurred
6. Discussions: Here is where authors interpret the results, and convince the readers of their findings and hypothesis
7. References: Any other work that was cited in the body of the text will show up here
8. Appendix: More figures, additional treatments on related math, or extra items of interest can find their way in an appendix

Developing a Systematic Approach

When you sit down to read it's important to have a plan. Simply starting to read from page one to the end will probably do you no good. Beyond retaining limited information, you will be exhausted and have gained very little for the tremendous effort. This is where many people stop.

Do plan to spend anywhere from **3–6 hours** to really digest a paper, remember they are *very* dense! Be ready and willing to make **several passes** through the paper, each time looking to extract different information and understanding. And please, do yourself a favor and do not **read the paper front to end** on your first pass.

Below are two lists. (i.) the systematic approach I take, more or less, when reading a paper (ii.) a general list of questions I try to answer as I go through the paper. I

typically add more specific questions depending on the paper.

Let's get started!

1. Try to find a quiet place for a few hours and grab your favorite beverage (could be coffee, tea, or anything really). Nowadays I often find myself working in splendid coffee shops².
2. Start by reading the **title** and **abstract**. Aiming to gain a high level overview of the paper. What are the main goal(s) of the author(s) and the high level results. The abstract typically provides some clues into the purpose of the paper. Think of the abstract as advertisement.
3. Spend about 15 minutes skimming the paper. Take a quick look at the figures and note any keywords to look out for when reading the text. Try to **get a sense for the layout of the paper** and where things are located. You will be referencing back and forth between the different sections and pages later on, it helps knowing where stuff is located. Try not to spend time taking any notes or highlighting/underlining anything just yet.
4. Turn your attention to the **introduction**. The more unfamiliar I am with the paper/field, the longer I spend in the intro. Authors tend to do a good job consolidating background info and providing copious amounts of references. This section is usually the easiest to read and it almost feels like you are reading from a textbook. Take notes of other references and background info you don't know or want to examine further.
5. This part is extremely critical. Carefully step through each **figure** and try to get a feeling for what they are telling you. When I was an undergrad, my neuroscience mentor gave me some good advice. Paraphrasing: *"Figures contain some of the most important information in a paper. Authors spend a lot of time creating them and deemed the information they contain to be important enough to communicate to the reader using a visual. Pay particular attention to them."* You will not understand all the figures very well the first time you step through them, but you will gain some idea of what the authors think is most important, and you will also reveal valuable information about what to pay attention to when you read the other sections.

6. So far you have probably spent about an hour. Take a break. Walk a bit, enjoy a croissant!
7. Now you are ready to make a first pass through the paper. This time you should start to take some high level notes. You will come upon words, and ideas that are foreign to you. Don't feel like you need to stop at every thing that does not make sense, you can simply mark it and move on. Aim to spend about an hour and a half. You don't want to get bogged down just yet in all the gory details. The goal of the first pass is to get acquainted with the paper. Like a first date. Your going to learn about the paper, ask some good questions, maybe make it laugh. But you don't want to get into every single little detail. That is rude. Begin again with the **abstract**, quickly skim through the **introduction**, give the **methods** section a diligent pass. Pay attention to the overall set up, the methods section includes a ton of detail you don't need to scrutinize every part at this point. Finally, read the **results** and **discussion** section aiming for some clarity on the key findings and how these findings were determined. Remember, the authors are trying to convince you, the reader, of the merit and findings of their work.
8. Saved by the bell. Take a break, do some jumping jacks and get the blood flowing. Unless you're in a coffee shop. Then don't do that.
9. Now that you have a good overview of the paper you are going to get into the nitty gritty of the **figures**. Having read the methods, results and discussion section, you should be able to extract out more gems from the figures. Find those gems. Aim to spend another 30 minutes to an hour on the figures.
10. You should feel confident in taking a second full pass through the paper. This time you will be reading with a very critical eye. This pass can take a long time an hour or two, you can also save this for later in the day, or the following day. Pay particular attention to the areas you marked as being difficult to understand. Leave no word undefined and make sure you understand each sentence. This pass you are trying to really learn the paper. Skim through areas you feel confident in (abstract, intros, results). The focus should be on shoring up what you did not understand previously, and gaining a command of the methods section and finally being a critical reader of the discussion section. The discussion section is where you can consider the authors' reasoning/rational and take what you learned from reading the paper and weigh it against evidence supplied in the paper. This section

should spark some interesting questions for you to ask your friends, or colleagues. You can even email the authors of the paper with an insightful question! It might take them a while to get back to you, but authors do enjoy having dialogue regarding their research and are typically more than happy to answer a question for a reader.

11. At this point, you should feel confident talking about the paper with colleagues, thinking critically about the results, and being able to compare the work to other research in the field (if you have read other papers). To retain and enforce what you have learned, I suggest you write about the paper. It can simply be a few paragraphs about what you learned and the significance of the results. You can reference the list of questions you were answering as you read through the paper.

As mentioned above, here is a general list of questions to help guide you. If you can answer these you have a solid understanding of the paper, at least to where you can communicate intelligently about it to others.

1. What previous research and ideas were cited that this paper is building off of? (this info tends to live in the introduction)
2. Was there reasoning for performing this research, if so what was it? (introduction section)
3. Clearly list out the objectives of the study
4. Was any equipment/software used? (methods section)
5. What variables were measured during experimentation? (methods)
6. Were any statistical tests used? What were their results? (methods/results section)
7. What are the main findings? (results section)
8. How do these results fit into the context of other research and their 'field'? (discussion section)
9. Explain each figure and discuss their significance.
10. Can the results be reproduced and is there any code available?
11. Name the authors, year, and title of the paper!
12. Are any of the authors familiar, do you know their previous work?
13. What key terms and concepts do I not know and need to look up in a dictionary, textbook, or ask someone?
14. What are your thoughts on the results? Do they seem valid?

I recommend finding people either in person or online to discuss the paper. Start a journal club with a goal of getting through 1–2 papers a month. The amount of extra insight I have gained from discussing a paper with a friend is immense. Remember..

the only thing better than suffering through a paper alone, is suffering through it with friends!

On another note there was a good article written by Keshav³ on how to read a paper. He introduces and explores a three phrase approach that might be of some interest to you. Give it a read as well!

Tools to Help You Get the Job Done

You can find papers primarily from several sources:

- **arXiv:** is an open-access repository (maintained at Cornell) where you can freely download and read pre-print research papers from many quantitative fields. Here is some more general info about arXiv. Many papers you find on the web will link back to the arXiv paper.
- **PubMed:** They say it best: *“PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health’s National Library of Medicine (NIH/NLM).”* PubMed has a robust search feature if you are looking for medical or life science related papers.
- **Google Scholar:** I use google scholar just as I would use google. Simply search for a topic, author or paper and google gets to work, on your behalf. As Google puts it *“Google Scholar provides a simple way to broadly search for scholarly literature. From one place, you can search across many disciplines and sources: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites. Google Scholar helps you find relevant work across the world of scholarly research.”*
- **Social media:** I find out about a lot of new papers simply by following and keeping up with several people who actively publish. Added bonus.. they typically push other papers they find interesting and which you might want to know about or read.
- **Friends and colleagues:** Find people interested in the same stuff as you, read papers with them and learn from each other. I get recommendations for good papers from friends. They act as good filters.

- University: going to your local college or university (if there is one close by) gives you access to libraries, librarians (very helpful search wizards!) and many journals where you can find and read articles that are typically behind online paywalls.

As you begin to read more papers you are going to want to store them somewhere. Tossing PDFs into a folder on your drive is all well and good, but there are creature comforts missing. Most researchers and grad students use a reference manager. Zotero and Mendeley are very popular, I like Zotero. Recently, I have been using PaperPile. I like PaperPile because it is lightweight, lives in my browser, and uses Google Drive to back up and store all my PDFs. It has a simple, refreshing user interface, and it has a really good tagging and folder hierarchy system. I can also annotate PDFs in my browser and build citation lists when I write. You get a lot of these features with almost any reference manager, but I happen to like PaperPile best.

A reference manager will quickly become your best friend as you collect and read more and more papers.

Thanks for reading through this. I hope you found it helpful and it gave you some good ideas when tackling your next paper. Most people have their own unique process when reading a paper. I am sure you will develop your own tweaks in time, hopefully this is a good template for you to get started.

For now just *trust the process*.

I am also hoping that we will get some good feedback and comments with other tips and tricks from readers.

Cheers,

Kyle

Reach me at: datasci@kmshannon.com

[linkedin.com/in/kmshannon/](https://www.linkedin.com/in/kmshannon/)

twitter.com/ButFirstData

[1] Hubbard, K. E., & Dunbar, S. D. (2017). Perceptions of scientific research literature and strategies for reading papers depend on academic career stage. *PloS one*, 12(12), e0189753.

[2] Shout out to Chris at CoffeeCycle! Simply the best coffee in San Diego.

[3] Keshav, S. (2007). How to read a paper. *ACM SIGCOMM Computer Communication Review*, 37(3), 83–84.

[Data Science](#)[Machine Learning](#)[Reading Research](#)[Reading](#)[Deep Learning](#)

Medium

[About](#) [Help](#) [Legal](#)