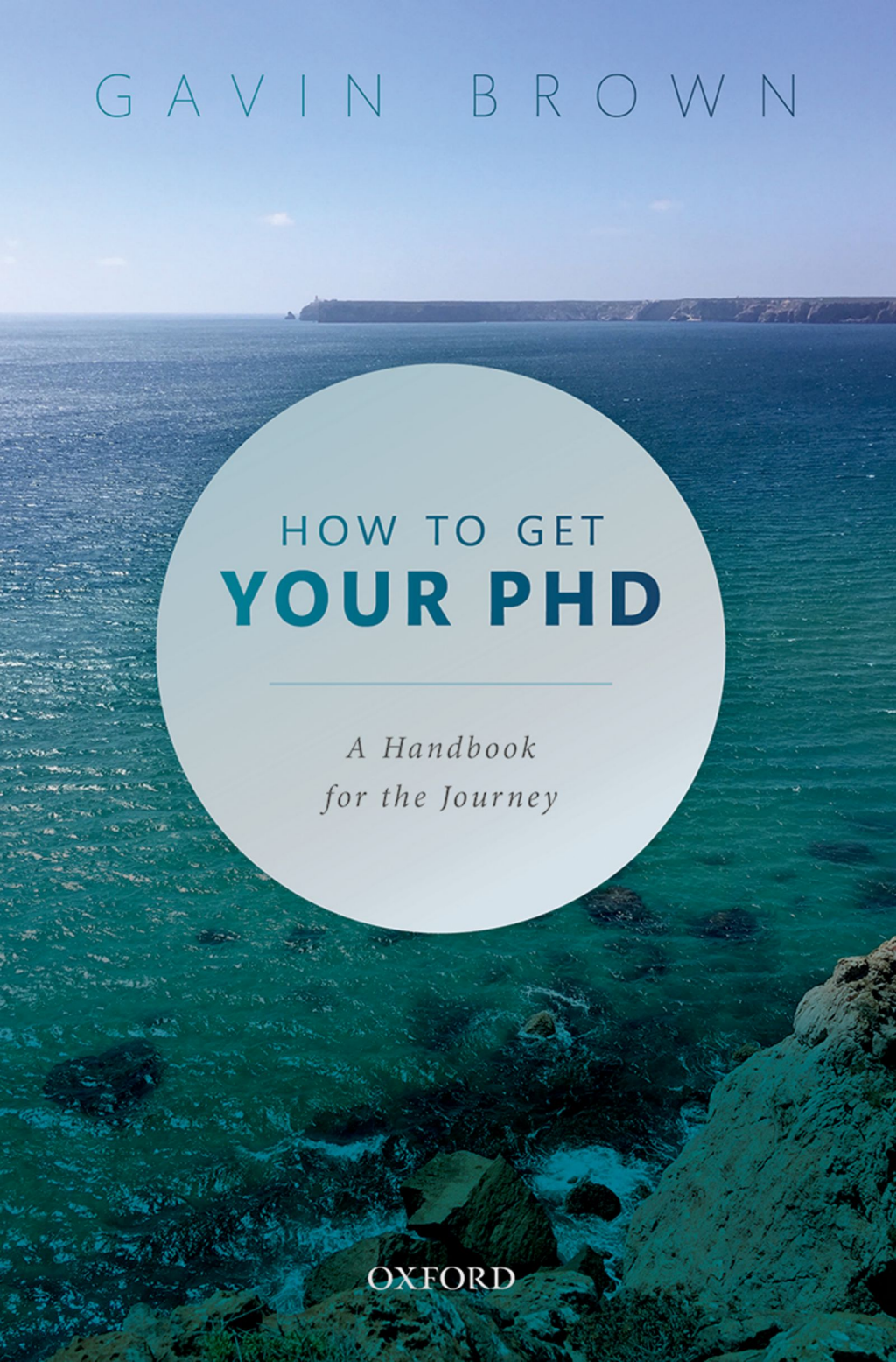

GAVIN BROWN
HOW TO GET
YOUR PHD
A Handbook
for the Journey
OXFORD

7

How to read

The more that you read, the more things you will know.
The more that you learn, the more places you will go.
—DR SEUSS

Read it out of order, and read it many times

Reading a scientific paper is a very different process than reading a newspaper article about science, or a popular science magazine. First, we generally don't read a scientific paper sequentially, in the order it is presented. This may seem odd, but bear with me. The first step is to read the abstract. This should summarise the main elements of the paper, and help you decide whether it is really relevant to your work. Then, read the conclusions, and to try to figure out what the authors claim they have achieved, and hopefully what evidence they have for their claim. Finally, if the abstract and conclusion have still got your interest, read the main body of the paper. When you encounter bits you don't understand, maybe scribble some notes on the paper, but feel free to skip over them. Otherwise, you'll hit a wall and not get past it. Truly understanding a paper takes multiple passes, and can take many hours (or weeks, or months!) of dedicated time, so don't expect to read a paper in one sitting. You'll come back to those difficult bits later, armed with new knowledge, and understand more.

Don't just read

You shouldn't just *read* papers. You should be reading and, simultaneously, *evaluating*. **Critical thinking is the essence of research**, and involves asking yourself a series of questions as you read. As you get better at reading scientific papers, you will do this naturally but, at the beginning, you may need to be more systematic. Below I suggest three criteria you might consider, with a set of questions for each. If you can answer these questions, you will have a deeper understanding of the paper, as well as what it implies for your work. As time goes by, you will evolve your own set of questions, and your own strategy specific to your field.

Originality: What is really 'novel' in this paper? What do the authors claim they have achieved? Are they addressing a new or complex problem? Are they proposing a new approach to an old problem, or demonstrating a new capability or property that hasn't been seen before? Maybe they present new perspectives, arguments, or insights about an existing problem or solution? How does it relate to other papers you know? An extension? A special case? Does it support other work, or contradict it?

Significance: Is this *really* an important problem to solve, or an important proposed solution? What is the *evidence* that other people really care about it? How many subsequent papers have cited this one?[1] Are the results really that surprising compared to other work you've seen? Even if they are addressing an important problem—are they addressing it properly, without making too many simplifying assumptions?

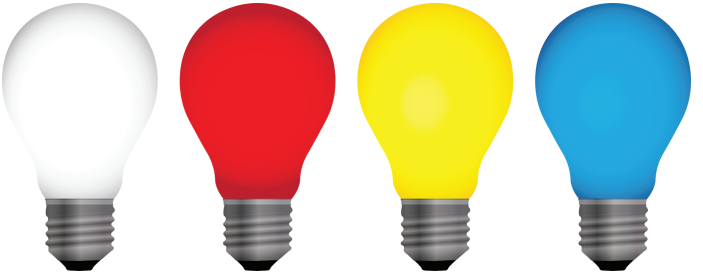
Rigour: How thorough have the authors been in their investigation? Have they considered all the relevant literature—are there papers that they should have cited, or compared their

[1] Various online tools can tell you this. Get to know the tool appropriate to your field—perhaps Google Scholar, SciVal, or Scopus. How many citations count as 'good' is very field dependent, so be sure to ask your supervisor.

ideas/results to? Are the experiments or results reproducible by others? Do they use well-established scientific methodologies and theories? Is their chain of reasoning solid—their assumptions justifiable? Are there unwritten assumptions that they may be glossing over, hoping you won't notice? Do they present sufficient evidence to really convince you of each of their claims? In simple language, how easy is it to pull a hole in their conclusions?

Consider how you could do better

Once you have answers to some of the questions above, you will be closer to your own new research ideas, which improve on this work. So, can you think of alternative approaches to the problem these papers solve, ones which would be simpler, or more rigourous? What is a good argument against the approach they use? Can their solution apply in another area? Have they raised some future work that you might be well positioned to pursue?



Be sceptical

One of the first things to understand in your research training is this: just because a piece of work is 'published' does not necessarily mean it can be trusted. The academic publishing industry is driven by much more than just the pursuit of science. Publishing companies want to make money, and individuals want to advance

their careers. As such, corners can be cut, and details can be glossed over, meaning you will doubtless encounter work that is less than trustworthy. You will find articles (and books) with typos, poor-quality writing, unwritten assumptions, and unjustified or even false claims. The problem you have is how to distinguish this paper from the papers you can trust, given your limited time. The only way to know for sure is to read it, but there are a lot of papers to read. Fortunately, there are some indicators which you can use as a proxy.

Look at where the article is published, and who is publishing it. Is it a respectable publication venue? Is it a respectable academic institution? Do the authors have a strong track record in this field? Ask your supervisor and other senior PhD students for advice on this. These questions should indicate the degree of credibility behind the paper, which may, in turn, indicate how much you can trust the work. Of course, it's certainly not the case that smaller institutions cannot publish truly great, world-leading research and, equivalently, it's not the case that world-leading institutions only publish great work. But it's a strong indicator—in the end, you may just have to read the thing.

How many papers?

Occasionally, a student will ask me, *'How many papers should I read in a week?'* They have heard some fellow student boasting that they read five papers in a day, or some other ridiculous figure. My response is always the same. If I were just to literally **read** the text, I could *read* it in twenty to thirty minutes. I'd be able to tell you what the authors claim to have done, and remember the name of the techniques they use, and I'd probably remember a few of the papers they cited in their literature review. But that's about it. I wouldn't truly **understand** the content of the paper. I wouldn't be able to guide my research in relation to theirs. I wouldn't be able to reliably criticise their methodologies,

techniques, or claims. So, if someone asks me how long it should take to read a paper, I'd say that is depends on which paper, and what you mean by 'read'. Some students will 'read' one paper in the morning, and another in the afternoon, and convince themselves they have finished. But they're most likely wrong. Admittedly, with some papers that are *exactly* in my area of expertise, I can indeed understand most of them in an hour. But, with others, I've had them on my desk for years, and I still don't *truly understand them.* So, don't measure your progress by how many papers you read. **Measure it by how many you truly** *understand***.**

How do you know when you *truly* understand something?

Let's say there is some complex **Thing** described in a paper—a method, a physical artefact, a mathematical proof, whatever. You'll have read the paper, and maybe used the **Thing** once or twice, and be able to state a few facts about it. But, even having done all that, *truly understanding* it is a very different experience.

A test of how much I understand something is in its **connections**. In research, everything is *connected*. Everything came from somewhere, building on previous work. So, ask yourself some hard questions on the *connections* and *context* of your research **Thing**. It will be made up of several components—which ones are *absolutely* necessary? What other **Things** share the same properties? Is your **Thing** a special case of another? Or a generalisation? How does it relate, *exactly*?

Another test is that of **explanations**. Albert Einstein said once, *'If you can't explain something simply, you don't understand it well enough.'* Try *explaining* the **Thing** to someone else. Depending on who it is, you'll have to change the level of detail, and pick out qualities that will matter the most to them. If you can do this effortlessly with several people, you can be more confident that you understand.

In general, it's easy to *deceive* yourself into *thinking* you understand, and find out some time later that you really do not. I prefer

not to have such surprises, so I try to hold a very high bar for uttering the words *'Yes, I understand that.'*

Read ferociously, but selectively

Whatever field you're in, there are probably a **lot** of research papers out there. The internet has made them all distressingly easy to find and, as a consequence, it can feel like a never-ending supply that you need to read. You have to *prioritise* and *organise* your papers.

I have a trick using what I call the **inner circle**, the **middle circle**, and the **outer circle**.[2] For each new paper you pick up, follow the strategy I outlined earlier: read the abstract, and the conclusions. But then, before you go any further, try to put it into one of the following three 'circles', which should tell you how much effort you need to put in:

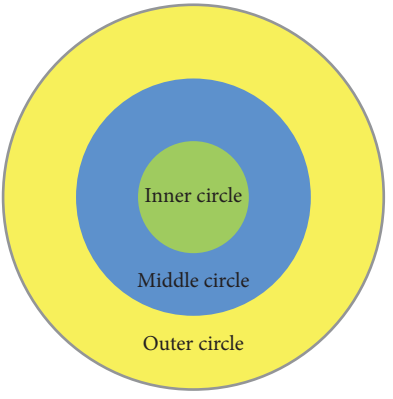
> **Inner circle:** This circle contains the papers that form the *core research challenge* you are dealing with. It probably has only one to five papers, but you should know these papers really deeply, inside out. You will have probably multiple printed copies of each, which you will have scribbled all over: in red ink and highlighter pens, annotating them while you're trying to understand, and you may well have contacted the authors to ask questions.[3] These are papers that you will have read so many times that you could almost reconstruct the text from memory, able to recite phrases they use in the text word-for-word. These are papers from which you will build the core of your own thesis.

[2] Alan Bundy, *The Researcher's Bible* (Edinburgh: The University of Edinburgh, 1985).

[3] I encourage you to do this—it really is common practice amongst academics. I do appreciate, it is scary the first time you send an email to some unknown author who has written your favourite paper. But, to give you some reassurance, it's likely the authors will be flattered that anyone is interested in their work. Give it a go, if only once.

Middle circle: This circle contains probably five to twenty papers that will certainly be related to your core idea but are more likely be competitors, or use alternative techniques/ideas, compared to the main one you are pursuing. These are papers you will certainly have read in detail, but perhaps didn't go into every single detail, apart from that necessary to compare/contrast your own idea. These are papers that will form the basis of your literature review.

Outer circle: This circle probably contains 100+ papers that are right on the edge of your area. You will have read the abstracts, and know what they claim, but you won't have gone into detail, and wouldn't claim yourself to understand them properly. But that's fine. These are in the periphery, only relevant as they attack problems that are vaguely similar to your own. If they're lucky, you might cite them in your literature review.



How do you know when you've found all the relevant papers?

You can't. You never will. And by this I mean you will never have *all* the papers, because new ones are always being written. And, even of the ones that are already written, there's a good chance there are still one or two papers that you've missed. But, you can

give yourself some reassurance. I advise simply this—use the strategy of **making Google work for you**. Write a very brief summary of your research, using five to ten words—just the key words. Type it into Google, and then Google Scholar (do both). *If you are not aware of EVERY single paper on the first search page, then you need to read more*. Now, I don't mean you have to **understand** every paper—you just need to be aware that they exist, and know the 'elevator pitch' (see chapter 9) of each one. Do this every month, or more frequently—and try small variations in the keywords, using synonyms. For example, instead of the search term 'data-set', why not try 'data'? And, instead of 'variation', why not try 'variance' or 'variability'? This can make a difference in what Google pops up for you.

In short, if you tell me you work on topic X, I shouldn't be able to surprise you by typing 'topic X' into Google and showing you the first page of search results!

You will learn more from reading than you realise

When you read a paper, your hope is that you'll learn the *technical material* it contains. However, there is more that you can learn from a truly great paper than just the technical material.

One of the things you'll learn from reading a *lot* of papers are the conventions of your area. What are the typical sections expected in a paper? How much (and what kind) of evidence is considered 'enough' to publish with? What sort of tone of language should you use? What proportion of the paper needs to be given over to literature review, and evaluation? How is an introduction structured, and what sort of things are typically in it? You can even learn some useful phrases—ways of defining key concepts in formal language, such that the research community will understand it. But, here's the hard part—you can also learn *bad* habits and conventions, if you read weaker papers. Just because a

paper uses particular shortcuts, or contractions of language, doesn't mean you should, too.

So, there's your challenge—learn the conventions of your research community, but try to have a filter: learn from the best and most successful authors.