Thoughts on thesis structure

I have now graduated around 30 Master's and PhD students, and for most of them, I have discussed my thoughts on the structure of a thesis and its importance. I thought it was an opportune moment to save myself some time and write this down, so future students can read it instead of listening to me banging on about a topic that I think is very important, but others may not (yet) understand why.

In high school and university, I had to write reports for science projects. The structure was generally Introduction, Methods, Results, Discussion, and Conclusions, which seemed reasonable at the time, but the flow often felt linear and disjointed. By the time of my PhD thesis, I knew there had to be a literature review, which made sense, but I didn't deeply appreciate it beyond ensuring the reader understood my work. Now, I see the literature review in a quite different light. Here, I will go through the different sections that most (though not all) scientific theses should have, why they are important, and how they link together.

1. **Title**. Clarity, specificity, and brevity are key here. If you put this online, it may be the first thing that indicates to someone whether they want to read further. Although you may start with a working title, as you progress, the specific focus may change. Don't be afraid to alter the title at the end to ensure it accurately captures your focus.
2. **Abstract**. This is another point where people decide whether to read further. Essentially, it expands on the title with a brief description of the problem statement and the most important findings. Include key methodologies and implications of the findings. Again, brevity is key, as long as you do not omit anything important. Write the abstract last.
3. **Acknowledgements**. This is the one part of the thesis where you can be emotional. Theses are hard and take a lot of effort, and generally, others have supported you, be it your supervisor(s), peers, family, or friends. Acknowledge how important they have been in the process. Don't forget to thank any funders who may have supported the research.
4. **Introduction** (this can be relatively short—often down to just 4 or 5 pages). Here, things get more detailed. By the end of the introduction, you should have laid out your research questions and hypotheses. The bulk of the introduction is a narrative explaining why these questions make sense. It quickly zooms in from the field as a whole to the specific portion you are examining. By the end of the introduction, it should be clear why your questions are interesting and significant and why your hypotheses seem reasonable. The rest of the thesis will be about answering these questions, so it is fine if your hypotheses are wrong. However, you should provide reasons in the introduction for your hypotheses. There may be several research questions, which can be split into various parts.

   Some factors to consider when coming up with your research questions:

   1. They should be relevant and novel. It should be clear why these are interesting questions, given the introduction.
   2. The questions should not be ambiguous, and it should be clear by the end of the thesis whether you have answered them.
   3. They should be answerable within the time and resources available. While there may be huge questions in the field, you are unlikely to answer them in a graduate thesis without some extraordinary means. The questions should seem manageable and bite-sized, giving you a realistic chance of answering them.

Note that you should have some research questions at the beginning, but these may change, subtly or significantly, throughout the research. While it is important to have research questions at the start, the introduction is often written after completing the investigation to create a clearer narrative. You may realize halfway through that a subtly or substantially different question is more appropriate or interesting, which should be reflected in the introduction.

5. **Literature review**. This is not just a generic review of the literature. It should make clear where your research questions sit in the field, what has been done around them, and include all relevant literature showcasing the methods you will use to answer the questions. The literature review bridges the introduction and the methodology. Everything in the methodology should be included for a reason based on the literature. Imagine the literature as a landscape. Your research question sits at some unlit point in the landscape. You need to write about the work that surrounds the unknown, so that it is clear what is known and that there is a gap in the field, and then you need to write about any tools that you will need to try and light this point. There is a tendency to write about everything connected with the field. However, while the literature review can be a substantial part of the thesis, ensure everything has a purpose and isn't just peripherally interesting.

   Think about your audience. Decide how basic to get. If you are writing a thesis about Offline Multiagent Reinforcement Learning, should you start with an introduction to basic Reinforcement Learning? It might be worth summarizing briefly, but don't give it too much space, as you can presume readers know the basics.

   Regarding the ordering of literature, I often see students discuss one paper and then move on to another, potentially in chronological order. However, this approach tends to read jarringly. It is much better to organize thematically and potentially bring in the same literature under multiple themes. A single piece of literature may highlight multiple aspects, and it is preferable to read a thesis where themes build the narrative through the literature review.

   Note also that this is an analysis of the literature too, and not just a repetition of it. It should be clear that you understand it, and aren't just parroting it.

6. **Methodology**. This is where you show the methods you will use to answer your research questions. The methodology should be firmly based on the literature review and reference it wherever possible. Every methodological choice should have a well-founded basis in the literature, and if you develop new methods, there should be good reasons for doing so, based on the literature. By now, you should see the vital interaction between the research questions, the literature, and the methodology. They should all work together coherently. Make sure that your methodology has enough detail for someone else to reproduce your results. This should include the research design, data collection methods, computations, etc. If you are writing code, put it on Github if you can (commented well!) and put a link in the methods section.
7. **Results**. This is, in a sense, the most stand-alone section. It should be as clear as possible, and choosing the right format to display your results is key. Consider statistical significance, robustness, and clarity. Every result should stem from the methodology section, and everything in the methodology should have an associated outcome (a chain of methods can lead to a single result, which is fine). Clear captioning of figures and tables is key. Many good books explain how to display data in the most digestible ways. Study these methods!

8. **Discussion**. This is where I initially had the least idea about the real purpose of the discussion and conclusion, but they are importantly distinct. This section discusses the results in light of the literature review. How do your results fit within the field? How do they agree with or contradict prior results? Here, you should again bridge two sections, heavily referencing the literature review and the results. Note that you should not discuss the research questions yet. The discussion is crucial for analyzing your results in light of existing knowledge in the field.
9. **Conclusion**. The conclusion again bridges two sections: it discusses the research questions (as set out in the introduction) in light of the results. Have you answered your original questions? Rephrase them at the beginning of the section. If your hypotheses were incorrect, in what way? If you haven't fully answered the questions, why not, and what is missing? What were the limitations of the work that prevented you from fully answering the questions? What would make the answers more robust? In addition, it is important to emphasise the broader implications of your findings. Finally, you might propose new research questions that naturally follow from your work.
10. **Bibliography**. I recommend using a bibliography manager, like Zotero, early on. I never used such a tool for my PhD thesis, and that was not a good choice. There are many good tools available, but choose one and stick with it for the coming years! Build a comprehensive set of papers to draw on for future research. You should be able to change the format and order of the bibliography and citations within each section at the click of a button.

OK, so that's my take on it. Let me know if you disagree, or indeed find this useful, if there's anything that you would add, or take away.