bioRxiv preprint doi: https://doi.org/10.1101/088278; this version posted November 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ten simple rules for structuring papers

Konrad Kording and Brett Mensh

Abstract

Good scientific writing is essential to career development and to the progress of science. A well-structured manuscript allows readers and reviewers to get excited about the subject matter, to understand and verify the paper's contributions, and to integrate them into a broader context. However, many scientists struggle with producing high-quality manuscripts and are typically given little training in paper writing. Focusing on how readers consume information, we present a set of 10 simple rules to help you get across the main idea of your paper. These rules are designed to make your paper more influential and the process of writing more efficient and pleasurable.

Introduction

Writing and reading papers are key skills for scientists. Indeed, success at publishing has long been used to evaluate scientists [1] and can help predict their future success [2]. In the production and consumption of papers, multiple parties are involved, each having their own motivations and priorities. The editors want to make sure that the paper is significant and the reviewers want to determine whether the conclusions are justified by the results. The reader wants to quickly understand the conceptual conclusions of the paper before deciding whether to dig into the details. And the writer wants to convey the important contributions to the broadest audience possible. All of these goals can be facilitated by structuring the paper well at the sentence, paragraph, section, and document scales.

Clear communication is also crucial for the broader scientific enterprise because "concept transfer" is a rate-limiting step in scientific cross-pollination. This is particularly true in the biological sciences and other fields comprising a vast web of highly interconnected sub-disciplines. As scientists become increasingly specialized, it becomes more important (and difficult) to strengthen the conceptual links across specialties. Communication across disciplinary boundaries can only work when manuscripts are readable, credible, and memorable.

A successful paper communicates the connections between data, methods, and interpretations. The interpretations culminate in a claim that gives significance to the work. This claim has to be supported by data and a logic that gives it credibility. Without carefully planning the paper's logic, there will often be missing data or missing logical steps on the way to the conclusion. While this is beyond our scope, your scientific *logic* must be crystal clear to powerfully make your *claim*.

Here we present 10 simple rules for writing papers (see Table 1). The first four rules are *principles* that can be applied to all the parts of a paper and further, to other forms of communication such as talks and posters. The next four rules deal with the

bioRxiv preprint doi: https://doi.org/10.1101/088278; this version posted November 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

primary communicative goals of each of the main *parts* of papers. The final two rules deliver guidance on the *process* – heuristics for efficiently constructing manuscripts.

Principles (rules 1-4)

Writing is communication. Thus, the reader's experience is of primary importance, and all writing serves this goal. When you write, you should constantly have our reader in mind. These four rules help you avoid losing your reader.

Rule 1: Present only one central contribution per paper, which you communicate in the title.

Your communication efforts are successful if readers can still describe the main finding of your paper to their colleagues a year after reading it. The most common number of findings that your readers will remember a year later is zero. While it is clear that a paper often needs to communicate a number of innovations on its way to its final message, it does not pay to be greedy. Papers that try to simultaneously focus on multiple contributions tend to be less convincing about each and are therefore less memorable.

The most important element of a paper is the title-- think of the ratio of the number of titles you read to the number of papers you read. The title is typically the first element a reader encounters, so its quality[3] determines whether the reader will invest time in reading the abstract.

The title not only transmits the paper's central idea, but can also serve as a constant reminder (to you) to focus the text on transmitting that idea. Science is, after all, the abstraction of simple principles from complex data. The title is the ultimate refinement of the paper's contribution. Thinking about the title early--and regularly returning to hone it--can help not only the writing of the paper, but also the process of designing experiments or developing theories.

Rule 2: Write for flesh-and-blood human beings who do not know your work.

Because you are the world's leading expert at exactly what you are doing, you are also the world's least qualified person to judge your writing from the perspective of the naïve reader. The majority of writing mistakes stem from this predicament. Think like a designer—for each element, determine the impact you want to have on people and then strive to achieve that objective [4]. Try to think through the paper like a naïve reader and make it easy for this reader to grasp the message quickly and with minimal effort.

bioRxiv preprint doi: https://doi.org/10.1101/088278; this version posted November 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The vast knowledge base of human psychology is also useful in paper writing. For example, people have working-memory constraints: they can only remember a small number of items and are better at remembering the beginning and the end of a list than the middle [5]. Use your own knowledge of psychology, and an eye to the naive reader's mind, to guide the writing process. In rule 10 below, we expand on the benefits of getting feedback from naïve test-readers throughout the process.

Rule 3: Stick to the context-content-conclusion (CCC) scheme

The vast majority of popular stories have a structure with a discernible beginning, a well-defined body, and an end. The beginning sets up the context for the story, while the body (content) advances the story towards an ending where the problems find their conclusion. This structure reduces the chances that the reader will wonder "why was I told that?" (if the context is missing) or "so what?" (if the conclusion is missing).

There are many ways of telling a story. Mostly they differ in how well they serve a patient versus an impatient reader[6] . The impatient reader needs to be engaged quickly-- this can be accomplished by presenting the most exciting content first (e.g., in many newspaper article titles). The CCC scheme we advocate serves a more patient reader, who is willing to spend the time getting oriented with context. A disadvantage of CCC is thus that it may not optimally engage the impatient reader. This disadvantage is mitigated by the fact that the structure of scientific articles, specifically the primacy of the title and abstract, already forces the content to be revealed quickly. Thus, a reader who proceeds to the Introduction is likely engaged enough to be patient enough to absorb context. Further, one hazard of excessive "action first" story structures in science is that you may generate skepticism in the reader, since they may be missing an important piece of context that makes your claim more credible. For these reasons we advocate CCC as a 'default' scientific story structure.

The CCC scheme defines the structure of the paper on multiple scales. At the whole-paper scale, the introduction sets the context, the results are the content, and the discussion brings home the conclusion. Applying CCC at the paragraph scale, the first sentence defines the topic or context while the last sentence gives the conclusion to be remembered. In this paper, we focus on the CCC way of writing.

Deviating from the CCC structure often leads to papers that are hard to read, but writers often do so because of their own autobiographical context. During our everyday life as scientists, we spend a majority of our time producing content and the remainder among a flurry of other activities. We run experiments, develop the exposition of available literature, and combine thoughts using the ineffable magic of human cognition. It is natural to want to record these efforts on paper. For example, we might want to structure a paper chronologically. But for our readers, most details of our activities are extraneous. They do not care about the

bioRxiv preprint doi: https://doi.org/10.1101/088278; this version posted November 28, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

chronological path by which you reached a result; they just care about the ultimate claim and the logic supporting it. Thus, all our work must be reformatted to provide a context to make our material meaningful and a conclusion that helps the reader understand and remember.

Rule 4: Optimize your logical flow by avoiding zig-zag and using parallelism.

Avoiding zig-zag. Only the central idea of the paper should be touched upon multiple times. Otherwise each subject should be covered in only one place, to minimize the number of subject changes. Related sentences, or related paragraphs, should be strung together rather than bouncing back and forth in an A-B-A fashion. Ideas that are similar, such as two reasons why we should believe something, should come immediately after one another.

Using parallelism. Similarly, across consecutive paragraphs or sentences, parallel messages should be communicated with parallel form. Parallelism makes it easier to read the text because the reader is familiar with the structure. For example, if we have three independent reasons why we prefer one interpretation of a result over another, it is helpful to communicate them with the same syntax so that this syntax becomes transparent to the reader, allowing them to focus on the content. Only one word should be used to refer to each concept.

The components of a paper (Rules 5-8)

The individual parts of a paper—abstract, introduction, results, and discussion—have different objectives and thus slightly different overall structure and paragraph structure. These specialized structures will be the topic of this section, and are summarized in Figure 1.

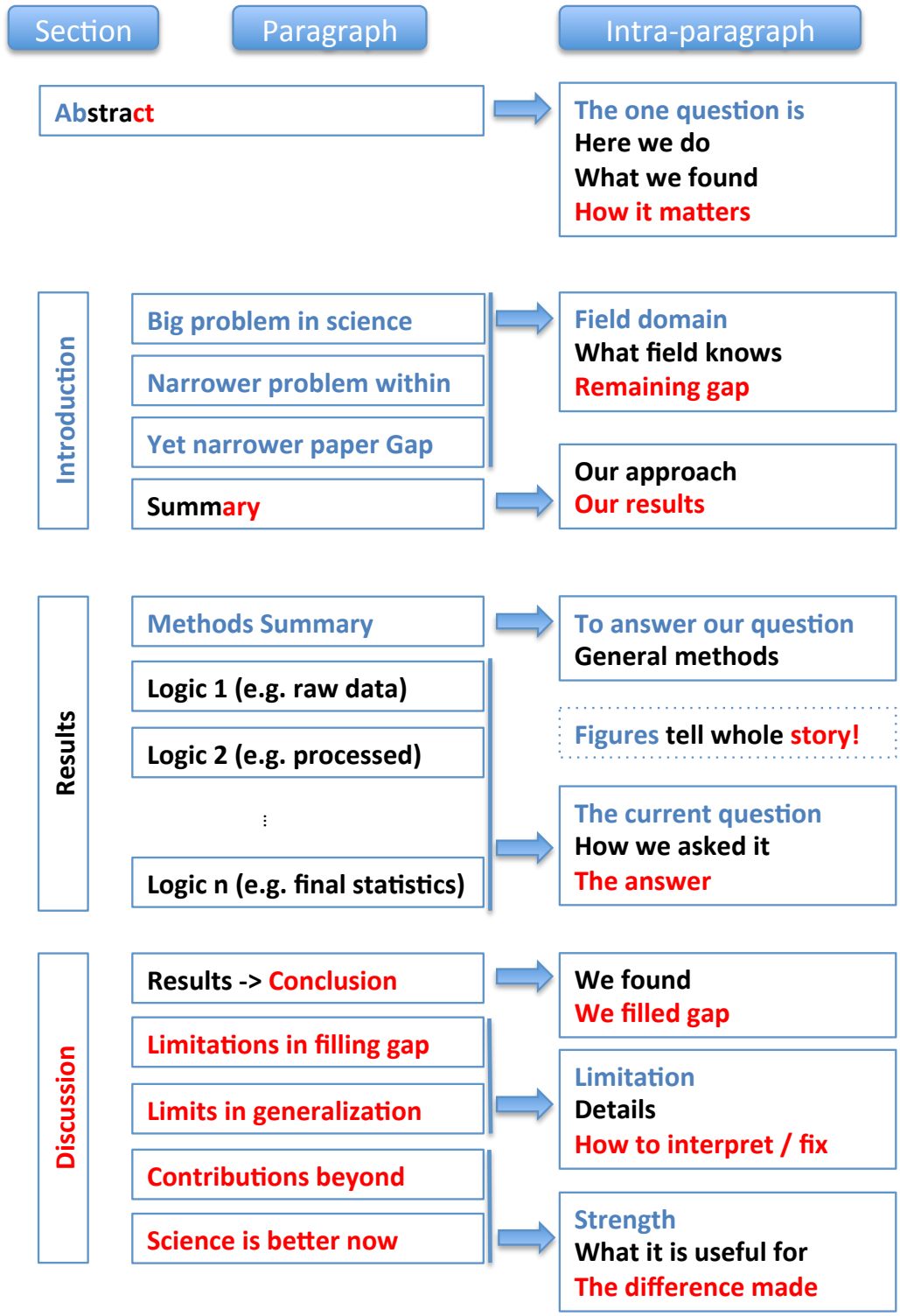

Figure 1: Context **Content** Conclusion make reading easy for the reader

Figure 1: A summary of all the typical structural elements in biomedical papers.

Rule 5: Tell a complete story in the abstract.

The abstract is, for most readers, the only part of the paper that will be read. This means that the abstract must effectively convey the entire message of the paper. To serve this purpose, the abstract's structure is highly conserved. Each of the C-C-C elements has two pieces to it, detailed below.

The *context* must get across the gap that the paper will fill. The first sentence orients the reader by introducing the broader field in which the particular research is situated. Then this context is narrowed until it lands on the open question that the research answered. A successful context section sets the stage for distinguishing the paper's contributions from the current state of the art by communicating what is missing in the literature (i.e., the specific gap) and why that matters (i.e., the connection between the specific gap and the broader context that the paper opened with).

The *content* ("Here we") first describes the novel method or approach you used that empowered you to answer the gap/question. Then you present the meat—your executive summary of the results. Avoid words with highly technical meanings (jargon—these words lose readers), and use as few words with subtle meanings as possible.

Finally, the *conclusion* interprets the results to answer the question that was posed at the end of the context section. Then, especially for more 'general' journals with a broad readership, there is a 2nd part of the conclusion section which highlights how this conclusion moves the broader field forward (i.e. "broader significance").

This structure helps you avoid the most common abstract mistake: talking about results before the reader is ready to understand them. Good abstracts usually take many iterations of refinement to make sure the results fill the gap like a lock and key. The broad-narrow-broad structure allows you to communicate with a wider readership (through breadth) while maintaining the credibility of your claim (which is always based on a finite/narrow set of results).

Rule 6: Get across why the paper matters in the introduction.

The introduction highlights the gap that exists in current knowledge or methods and why it is important. This is usually done by a set of progressively more specific paragraphs that culminate in a clear exposition of what is lacking in the literature, followed by a paragraph summarizing what the paper does to fill that gap.

As an example of the progression of gaps, a first paragraph may explain why understanding cell differentiation is an important topic and that the field has not yet solved what triggers it (a field gap). A second paragraph may explain what is unknown about the differentiation of a specific cell type such as astrocytes (a subfield gap). A third may provide clues that a particular gene might drive astrocytic differentiation, then state that this hypothesis is untested (the gap within

the subfield that you will fill). The gap statement sets the reader's expectation for what the paper will deliver.

The structure of each Introduction paragraph (except the last) subserves the goal of developing the gap. Each paragraph first orients the reader to the topic (a context) and then explains the 'knowns' in the relevant literature (content) before landing on the critical 'unknown' (conclusion) that makes the paper matter at the relevant scale. Along the path, there are often clues given about the mystery behind the gaps that lead to the untested hypothesis or undeveloped method of the paper; we learn why closing the gap is promising. The introduction should not contain a literature review beyond the motivation of the paper. This gap-focused structure makes it easy for experienced readers to evaluate the potential importance of a paper – they just need to assess the importance of the claimed gap. And these gap statements are what the paper's contribution will be measured against; excitement, expectations, and the high-level context are set by the paper's gap definition.

The last paragraph of the introduction is special – it compactly summarizes the results, which fill the gap you just established. It differs from the abstract by not needing to present context (which has been given just above), by being somewhat more specific about the results, and by only briefly previewing the conclusion of the paper, if at all.

Rule 7: Communicate the results as a sequence of statements that build on one another to support the central contribution.

The results section needs to convince the reader that the central claim is supported by data and logic. Every scientific argument has its own particular logical structure, which dictates the sequence in which its elements should be presented.

For example, one paper may set up a hypothesis, verify that a method for measurement is valid in the system under study, and then use the measurement to disprove the hypothesis. Alternatively, a paper may set up multiple alternative (and mutually exclusive) hypotheses, and disprove all but one to provide evidence for the remaining interpretation. The fabric of the argument will contain controls and methods where they are needed for the overall logic.

In the outlining phase of paper-preparation (see Rule 9), sketch out the logical structure of how your results support your claim and convert this into a sequence of declarative statements that become the headers of subsections within the Results section—most journals allow this type of formatting. These headers keep the reader oriented to the big picture.

Figures and legends are particularly important. Indeed, some readers will only read the abstract and look at the figures. Ideally, the figures should graphically tell the story without the need to read the legend or text.

The first results paragraph is special in that it typically summarizes the overall approach to the problem outlined in the Introduction, along with any key innovative methods that were developed. Some readers do not read the Methods, so this paragraph at least gives them the gist of the methods that were used.

Each subsequent paragraph in the results section starts with a sentence or two setting up the question that the paragraph answers. For example "To verify that there are no artifacts,...", "What is the test-retest reliability of our measure?", or "We next tested whether Ca^{2+} flux through L-type Ca^{2+} channels was involved". The middle of the paragraph presents data and logic that pertain to the question, and the paragraph ends with a sentence that answers the question. For example, it may conclude that none of the potential artifacts was detected. This structure makes it easy for experienced readers to fact-check a paper. Each paragraph convinces the reader of the answer given in its last sentence. This makes it easy to find the paragraph where a suspicious conclusion is drawn and check the logic of that paragraph. The result of each paragraph is a logical statement, and paragraphs farther down in the text rely on the logical conclusions of previous paragraphs, much as theorems are built in the mathematical literature.

Rule 8: Discuss how the gap was filled, the limitations of the interpretation, and the relevance to the field.

The discussion section explains how the results have filled the gap identified in the introduction, provides caveats to the interpretation, and describes how the paper advances the field by opening new opportunities. This is typically done by recapitulating the results, discussing the limitations, and then revealing how the central contribution may catalyze future progress.The first discussion paragraph is special in that it generally summarizes the important findings from the results section. Some readers skip over substantial parts of the results, so this paragraph at least gives them the gist of that section.

Each following paragraph in the discussion section starts by describing an area of weakness or strength of the paper. It then evaluates the strength or weakness by extensively linking into the relevant literature. Lastly, it concludes about the authors' way of perceiving the contribution, discussing potential ways for going forward.

For example, the first paragraph may summarize the results, focusing on their meaning. The second through fourth paragraphs may deal with potential weaknesses and how the literature alleviates those concerns or how future experiments can deal with these weaknesses. The fifth and the globally most exciting sixth paragraph may then culminate in a description of how the paper moves the field forward. Step by step, the reader thus learns to put the paper's conclusions into the right context.

Process (Rules 9-10)

To produce a good paper, authors can utilize helpful processes and habits. Some aspects of a paper are more important to its impact than others, suggesting that your investment of time should be weighted towards the issues that matter most. Moreover, iterating using feedback from colleagues allows the authors to improve the story at all the levels to produce a powerful manuscript. Choosing the right process makes writing papers easier and more effective.

Rule 9: Allocate time where it matters: Title, abstract, figures, and outlining.

The central logic underlying a scientific claim is paramount. It is also the bridge connecting the experimental phase of a research effort with the paper-writing phase. Hence it is important that the logical structure of the paper is planned in concert with experiments being performed. Thus, it is useful to formalize the logic of ongoing experimental efforts into an evolving document of some sort that ultimately becomes the blueprint for the outline of the paper.

You should also allocate your time according to the importance of each section for communicating with the reader. The title is read by far more people than the abstract, which is read by far more people than the bulk of the paper, and the rest of the paper receives far more attention than the methods. Budget accordingly.

The time that we do spend on each section can be used efficiently by planning text before producing it. Make an outline. We like to write one informal sentence for each planned paragraph. It is often useful to start the process around descriptions of the obtained results – these may become the section headers in the results section. As the story has an overall arc, each paragraph should have a defined role in advancing this story. This role is best verified at the outline stage.

Rule 10: Get feedback to reduce, reuse, and recycle the story.

Writing can be considered an optimization problem in which you simultaneously improve the story, the outline, and all the component sentences. In this context, it is important not to get too attached to one's writing. In many cases, trashing entire paragraphs and rewriting is a faster way to produce good text than incremental editing.

There are multiple signs that further work is necessary on a manuscript (see Table 1). For example, if you, as the writer, cannot describe the entire outline of a paper to a colleague in a few minutes, then clearly a reader will not be able to. You need to further distill your story. Finding such violations of good writing helps in improving the paper at all levels.

Successfully writing a paper typically requires input from multiple people. Test readers are necessary to make sure that the overall story works. They can also give valuable input on where the story appears to move too quickly or too slowly. They can clarify when it is best to go back to the drawing board and retell the entire

story. Reviewers are also extremely useful. Non-specific feedback and apparently bored reviews usually imply that the reviewers did not "get" the big-picture storyline. Very specific feedback usually points out places where the logic within a paragraph was not sufficient. It is vital to accept this feedback in a positive way. As input from others is essential, a network of helpful colleagues is fundamental to making a story memorable. And to keep this network working, make sure to pay back and read their manuscripts.

Table 1: A summary of rules and their uses

Rule	**Sign it is violated**
1: One big idea	Readers cannot give one-sentence summary
2: Humans as audience	Readers do not "get" the paper
3: Context, Contents, Conclusion	Readers ask why something matters or what it means
4: Optimize logical flow	Readers stumble on a small section of the text
5: Abstract: Compact summary of paper	Readers cannot give the 'elevator pitch' of your work after reading it
6: Introduction: Why the paper matters	Readers show little interest in the paper
7: Results: Why the conclusion is justified	Readers do not agree with your conclusion
8: Discussion: Preempt criticism, give future impact	Readers are left with unanswered criticisms/questions on their mind
9: Time allocation	Readers find data clearer than the ultimate message
10: Iterate the story	The paper's contribution is rejected by test readers, editors or reviewers.

Discussion

This paper focused on the structure or "anatomy" of manuscripts. We had to gloss over many finer points of writing, including word choice and grammar, the creative process, and collaboration. A paper about writing can never be complete; as such there is a large literature dealing with issues of scientific writing [6,7,8,9,10,11,12,13].

Personal style often leads writers to deviate from a rigid, conserved structure and it can be a delight to read a paper that creatively bends the rules. However, as with

many other things in life, a thorough mastery of the standard rules is necessary to successfully bend them.

References

1. Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 102: 16569-16572.
2. Acuna DE, Allesina S, Kording KP (2012) Future impact: Predicting scientific success. Nature 489: 201-202.
3. Paiva CE, Lima JPSN, Paiva BSR (2012) Articles with short titles describing the results are cited more often. Clinics 67: 509-513.
4. Carter M (2012) Designing Science Presentations: A Visual Guide to Figures, Papers, Slides, Posters, and More: Academic Press.
5. Murdock BB, Jr. (1968) Serial order effects in short-term memory. J Exp Psychol 76: Suppl:1-15.
6. Lisberger SG (2011) From Science to Citation: How to Publish a Successful Scientific Paper: Stephen Lisberger.
7. Simons D (2012) Dan's writing and revising guide.
8. Sørensen C. This is Not an Article — Just Some Thoughts on How to Write One; 1994; Oulu University. Syöte, Finland. pp. 46-59.
9. Day R (1988) How to write and publish a scientific paper: Phoenix: Oryx.
10. Lester JD, Lester J (1967) Writing research papers: Scott, Foresman.
11. Dumont J-L (2009) Trees, Maps, and Theorems: Principiae.
12. Pinker S (2014) The Sense of Style: The Thinking Person’s Guide to Writing in the 21st Century: Viking Adult.
13. Bern D (1987) Writing the empirical journal. The compleat academic: A practical guide for the beginning social scientist 171.