

# Scientific Writing: The Art of Programming Brains

*Helmut Grubmüller\*, December 2020*

“On occasion of the stepwise descending motion downwards towards the ground, which was carried out by me, I herewith would like to add as a corresponding remark, particularly concerning the last step, which was performed most recently, that this step was a relatively minor effort by comparison to what this project and its successful completion (or near completion) potentially represents for a considerable fraction of the world’s population, by analogy.”

Neil Armstrong, transcribed into PhD thesis style.

“That’s one small step for [a] man, one giant leap for mankind.”

Neil Armstrong, July 20, 1969, Moon.

If you think writing is simple — after all, you learned it early at school! — you may be one of the many students who are baffled by your supervisor’s feedback to your first draft of a manuscript or thesis: lengthy, too wordy, and hard to read at best, and often just awkward and incomprehensible.

You will then realise that clear, concise, and coherent writing takes much more effort than expected, and it

always takes longer than you thought. So you invent an excuse: “I just don’t have writing talent; after all, I am a scientist, not a professional writer.”

However, effective writing is not a matter of talent. It results from sufficient efforts, the right mindset, and observing several rules. You will recall some of these from high school, but not all are adequate for scientific writing. For example, you may have been told to avoid repeating a word by all means, which can however compromise clarity. And there are other rules you may be unaware of.

To fill these gaps, many excellent textbooks on scientific writing have been published [1–5], which are a fine help for students. Nevertheless, I found several recurring issues in many drafts by masters, PhD, and postdoctoral students.

This is a collection of helpful rules to address these issues, distilled from the feedback I provided. While adherence to these rules does not guarantee an excellent text, ignoring too many of them will inevitably result in communication failure.

And because writing is one of *the* pivotal skills of a scientist, it is worth the effort.

---

\*Email: [hgrubmu@gwdg.de](mailto:hgrubmu@gwdg.de), Twitter: @CompBioPhys  
Max Planck Institute for Biophysical Chemistry  
Theoretical and Computational Biophysics Department  
Am Fassberg 11, 37077 Göttingen, Germany

## PART I: YOUR MINDSET

Let us agree on your main aim: Almost always it is much more vital for you than for the reader that your text is understood. Do not expect your reader to make every effort to understand what you possibly meant; it is *your* duty to make reading your text as easy as possible. You need to develop a plan of how to change the reader's mind, step by step, from its current state of knowledge to where you want it to be.

Here, most writing guides remain pretty vague, suggesting "put yourselves in your reader's shoes" or similar. But how, precisely, can you achieve this?

Let me explain.

Many think writing means to copy information from our brains onto paper. No wonder — that is what we were taught at high school: *First sort your thoughts, then write them down!* Of course, the first draft of this copy will be imperfect, incomplete, or even wrong. To refine, you ask: "Is my text consistent with what I meant to say?" Yes: ok, next sentence; no: improve until it is.

Sounds like a great approach, right? Unfortunately not. Just being correct ("But I *did* write that!", or even worse, "But I *did* mean that!") is not enough. Rather, it results in text that is hard to read and digest. As most teachers know very well, teaching is more than producing only a 'copy of my thoughts'. Throwing a

'core dump' of your mind in front of the reader's feet just will not do; it misses half — the essential half! — of what is actually needed to successfully communicate.

What the copy-approach fails to verify is whether your text, when read, actually creates a copy of the author's thoughts in the reader's mind. Sounds like a no-brainer? In fact, it's a two-brainer!

A simple example and Fig. 1 illustrate this idea. During writing, you often need to choose the proper word, expression, or description. Consider this text:

**Poor:** The figure compares the corresponding outcome with the Schrödinger equation, which yields larger values than our results.

Such vague descriptions (large blue regions in the figure) encompasses many different thoughts (grey dots), both in your mind and, crucially, in that of your reader. Of course *you* have (hopefully!) a much more precise meaning in your mind (red dot). However, your reader cannot read your mind and has access only to the large blue region. As a result, her brain will come up with a random guess (green dot), a markedly different thought. Communication has failed.

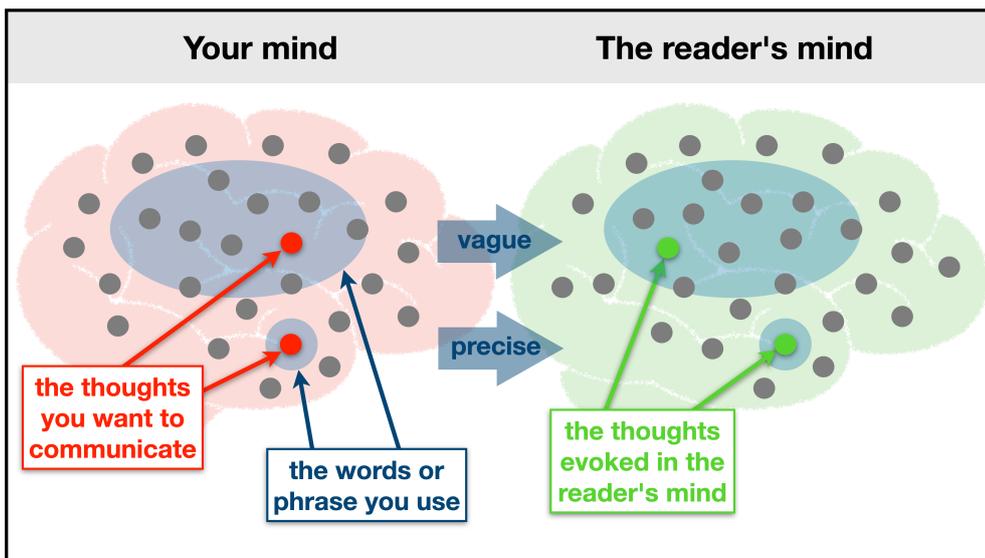


Fig. 1: Vague writing fails to communicate. In your (left) and the reader's mind (right), blue regions indicate 'clouds of thoughts (grey dots)' that are compatible with a word or phrase. The red dots are specific thoughts you want to communicate; the green dots are thoughts evoked in the reader's mind.

Contrast this with:

**Better:** We computed the ground state energies of the methane molecule (Fig. 42, solid line) using the SHRDLU software package. These energies are smaller than those calculated analytically (dashed line) via a spherical harmonics expansion using the Schrödinger equation.

This text (narrower blue regions in the Figure) is compatible with much fewer thoughts. As a result, the thoughts activated in the reader's mind are much more similar to yours. This is why textbooks tell us we should strive for precision.

Unfortunately, the vague terms are much easier to find, for two reasons. First, there are simply many more of them — 'result', 'finding', 'data', 'values' all qualify in most instances — whereas the more precise terms are much fewer. Second, they are harder to find and recall, simply because they are precise and therefore used rarely. As a result, the word or expression that first comes to your mind will most often be vague.

The problem is: Writing in copy-mode, you only check if your thought falls within the region of the word you wrote. And surely it does ("yep, 'corresponding outcome' is indeed what I had in mind; good, move on!"). And should you have attempted to find a more precise expression — 'ground state wave function calculated by SHRDLU' — chances are your first attempts will not pass the copy-test ("Well, strictly it is not a wave function ... and was it really calculated by SHRDLU or rather by me? — damn!"). And even if you happen to find the correct term ("Hurray! That's the word I was looking for!"), you might decide to play safe and nevertheless prefer the more general expression ("But wait ... am I really sure it's a ground state energy? What if I'm wrong? That would be embarrassing. Better call it a result, that can't be wrong.").

Unexpectedly, the effect of copy-mode is more detrimental. Not only does it fail to make sure the text creates an accurate copy of the author's thoughts in the reader's mind, it actually drives your text into the opposite direction! Whereas your first draft might do

reasonably well, it will deteriorate in your subsequent attempts to make it better agree with your thoughts. This is because your state of mind differs from that of the reader — at least in what you wish to communicate, typically in much more. By forcing the text to agree with *your current* knowledge, it will deviate from the *reader's initial* knowledge, and thus become increasingly hard to digest. Again, copy-mode lacks an accurate model of how the reader's mind changes during reading.

Contrast this approach with 'brain emulation' mode. Rather than checking if the word we chose is merely consistent with your thought, ask: "Which thoughts would this particular word most likely evoke in the reader's mind? Is there a fair chance those thoughts are close enough to what I actually want to say? Are there other thoughts nearby that I definitely do *not* want to evoke, but which are possibly also covered by my writing? And if so, how can I avoid that?" Make sure you only move on after you have found a word or phrase that passes all these tests.

But that's not all. We have a memory — which implies that what particular thought is evoked in the reader's mind depends on what she read before. As you write, you thus need to maintain and continuously update a precise model of the reader's mind. Continuously update a list of pieces of information the reader absorbs during reading — *starting with an empty list*. Make sure, during test-reading, that only what you actually wrote gets onto this list, not what you think or meant. Readers read your text, not your mind!

After each sentence, paragraph, or section, check if the model of the reader's mind is *precisely* where you want it to be. If not, change the text until it is — but not *after* the bug in the text, as many are tempted to do (again, because it is so much easier). Instead, analyse where *exactly* the reader's mind deviated and why, and fix the text *there*.

To see 'brain emulation' mode at work, I urge you to work through the stepwise improvement of a rather undigestible paragraph in the box on the next page. Do not read on before you are done!

## EXAMPLE:

**Poor:** We expected the angular uncertainty to scale as  $\Delta\Theta \propto N^{-1/2}$ . To express the spatial uncertainty  $\Delta x = R\Delta\Theta$  as a function of molecular mass, we assumed the following scaling of mass  $M \propto R^3$  and number of photons  $N \propto I_0M$ , where  $I_0$  is the incident beam intensity. Combining those with the functional dependency of  $\Delta\Theta(N)$  yields  $\Delta x \propto M^{-1/6}$ . Of the two competing trends: worsening of the spatial resolution with increasing radius and improvement of the angular resolution with growing molecular mass the latter one is dominant. And thus, for a given beam intensity structures with larger mass are better resolved.

Right at the first sentence, the reader will interrupt: "Wait. Where are we going? Which question does the author address?" So, for an improved version, let's provide this pivotal information first:

**Iteration #1:** How does the spatial resolution change with the mass of the molecule? We expected the angular uncertainty to scale as  $\Delta\Theta \propto N^{-1/2}$ .

But now the context of the second sentence has changed. Restarting our brain emulation *from the beginning* reveals that our reader will now wonder why the *angular* uncertainty  $\Delta\Theta$  is mentioned, when in fact the *spatial* resolution is of interest. So let's first address the spatial resolution:

**Iteration #2:** How does the spatial resolution change with the mass of the molecule? To express the spatial uncertainty  $\Delta x = R\Delta\Theta$  as a function of molecular mass, we assumed the following scaling of mass  $M \propto R^3$  and number of photons  $N \propto I_0M$ , where  $I_0$  is the incident beam intensity.

Next, our brain emulation shows that the reader expects  $R$  and  $\Theta$  to be discussed, and thus wonders why suddenly molecular mass is mentioned as a function of  $R$ . Further, the appearance of two further quantities  $N$  and  $I_0$  without obvious relation to the former triggers an overload alert. This is way too much for one sentence, so let's focus on  $M \propto R^3$ . This relation contains the expected  $R$ , but buried on the right side. Solve for  $R$ , and we can fulfil the reader's expectation:

**Iteration #3:** How does the spatial resolution change with the mass of the molecule? The spatial uncertainty  $\Delta x = R\Delta\Theta$  increases with the radius  $R \propto M^{1/3}$  of the molecule, which in turn increases with molecular mass  $M$ . We expected the angular uncertainty to scale as  $\Delta\Theta \propto N^{-1/2}$ .

That has also integrated  $M$  nicely in the right logical order. Because our re-emulation flags that the reader now expects  $\Delta\Theta$ , the first sentence of the original version finally found its proper place. Except, the logical connection has not yet been made, which triggers a search for the overarching structure of this paragraph.

Again, our brain emulation offers help. At "Of the two competing trends, [...]", the reader facepalms: "If only I had been told much earlier, that would have saved me so much guesswork!". So let's put this structure to work — but not at the point of the facepalm, but much earlier. Here is the result of a few more iterations:

**Iteration #4:** How does the spatial resolution change with the size of the molecule? Two opposing effects are expected. On the one hand, the spatial uncertainty  $\Delta x = R\Delta\Theta$  increases with the radius  $R \propto M^{1/3}$  of the molecule, which in turn increases with molecular mass  $M$ . On the other hand, the orientational uncertainty  $\Delta\Theta \propto N^{-1/2}$  decreases with increasing number  $N$  of recorded photons, which is proportional to the molecular mass  $M$  and to the beam intensity  $I_0$ . Therefore, the orientational uncertainty scales as  $\Delta\Theta \propto (I_0M)^{-1/2}$ .

Now we can harvest the fruits of our new structure, and 'close the bracket' that was opened by 'Two opposing effects':

Combining these two effects yields the counterintuitive result that the spatial uncertainty  $\Delta x \propto (I_0M)^{-1/6}$  decreases with molecular mass.

Note that 'counterintuitive result' has been added, because the reader would most likely have expected an *increase*, not a decrease with molecular mass.

The last sentence of the original version, properly adapted, provides the answer to the question that started the paragraph.

This scaling argument shows that better resolution is expected for larger molecules.

A final brain emulation of the whole paragraph shows us the reader's state of mind at this position: "So what? Why did the author tell me that?" — which is answered by the last sentence:

*We will therefore use small molecules as the most challenging test cases.*

Doesn't that read much better?

As you have seen from this example, writing text resembles writing a computer program. It is a sequence of tiny bits of information, crafted to change the reader's mind step by step until the desired result is achieved: Information has been reliably copied from your mind into the reader's mind.\*

Our step-by-step debugging has probably also illustrated one more similarity to computer programming: A quick local 'patch' of a flaw in the logical flow of your text will rarely do. Not only do you need to recall the complete context (as textbooks teach) but, more importantly, also the precise state of mind of the reader at the respective position within the text. Because nobody can memorise the reader's mind state everywhere in the text, you have to start from the most recent 'checkpoint' — often much earlier in the text — and emulate the reader's brain from there all the way to the passage in question. As a result, seemingly local text improvements often turn out to be far from local,

and more often than not require hard work and quite a bit of time. With some experience, one can tell a student — much to their surprise without even looking at the text — that their revision most likely won't do, just from the too short time they spent on it.

Very much like test-running your newly written code, emulating the reader's brain is key. Nobody would sell a computer program without extensive prior test runs. Why do so many authors?

Switching from 'copy-of-my-thoughts' mode to a 'create-a-copy-in-another-brain' mindset thus makes all the difference. It will transform your writing from a passive copy job into an active programming task. It will change your dry, boring, and incoherent sequence of lengthy sentences into a refreshing text that will be understood.

Your job is that of a brain programmer — *that* should be your mindset.

---

\*Of course, the 'programming brains' mindset is also very useful for communication beyond scientific writing.

## PART II: HELPFUL RULES

To program a computer, it helps to know how it works, and you must adhere to the rules of a programming language. Similarly, knowing how your reader's brain works will make it much easier for you to program it effectively. And although there is no rigid 'programming language' for brains, certain writing patterns have proven to achieve better communication than others.

Most rules for good scientific writing follow from these facts and thus help us achieve our main aim: to make it as easy as possible for the reader to understand your text.

### 1. BEFORE YOU BEGIN

**Separate the important from the unimportant.** You finally completed your project and are about to begin writing. Your mind is full of all the tricks, fixes, and technicalities of your work that kept you busy over the past weeks or months. In contrast, the more fundamental concepts and premises, as well as the main questions that initially triggered your work, have become so much second nature to you that you do not even think about them.

Ignoring this psychological effect will result in a random sequence of technicalities and mostly insignificant statements which at best belong to the Supplement, whereas the main conceptual steps are missing. The outcome is an incomprehensible and boring text, which will fail to convey the significance of your work.

It is therefore crucial that you make yourself aware of the main pillars of your story and that you separate primary from secondary and tertiary results. This essential mental effort requires more time and energy than you might expect. A carefully kept lab-book, listing not only the bare results but also the questions that led to them, as well as what you concluded, will turn out to be immensely useful at this point.

Next, go through the list and decide for every item whether or not it is important enough to survive. Only those items that are essential for an understanding of

your main conclusions go into your manuscript, as well as everything that is needed to reproduce your results. Discard all concepts or methods that are not relevant to your main results and conclusions — even though it may have taken considerable effort to learn those during your project and irrespective of how beautiful or exciting you may find them.\*

Scientific papers and theses are reports, not textbooks!

**Define new ideas and concepts.** All ideas and concepts unknown to your reader must be defined and explained. Which precisely those are, depends on the readership and, hence, on where you intend to publish. A higher impact journal with a broad readership will require more basic explanations than a very specialised one; for a Dissertation, a good rule of thumb is to recall what you knew and what you did not know before you started working on your first project.

More often than not, just throwing the formal definition at your reader will not suffice; it needs to be supplemented with examples, pictures, or suitable analogies. Also take into account that the human mind is not good at grasping abstract or general definitions; help the reader by providing a specific example *before* the abstract definition.

Some of the concepts you need to explain rest on others which, therefore, must have been defined before. Always put your definitions and explanations into the right order, with each new term building upon previous ones. Patchworks such as the phrase '*will be explained in Section X*' or the urge to repeat an explanation, point to a wrong order.

### 2. FIGURES

**Figures first.** Before you start writing, prepare *all* figures — at least as an accurate and detailed sketch — and determine their order. Imagine you give a short talk on the subject; you should be able to explain all the required concepts, methods, and results with the figures at hand. The same holds true for mathemat-

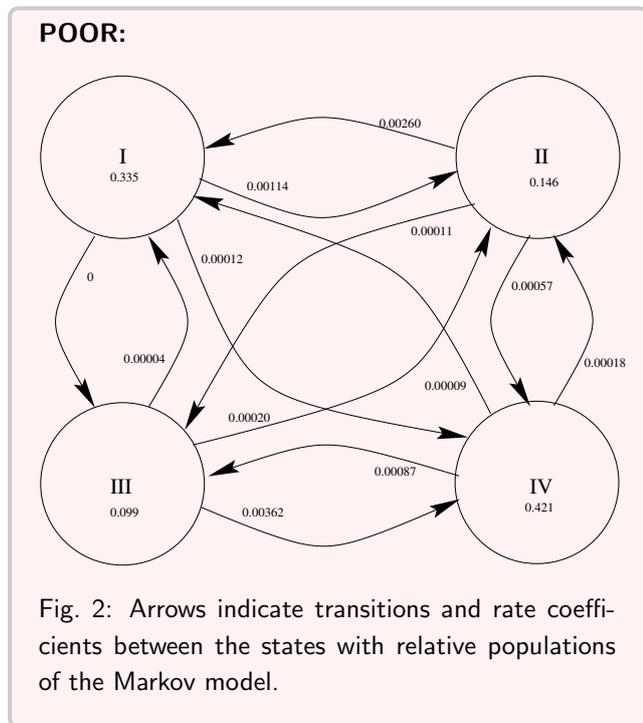
\*This is *not* a license for cherry picking; results that speak against your conclusions *are* relevant!

ical expressions. Very much like you would use the figures during your talk, pointing to specific items of each figure to convey your story, so should your text.

Most readers will first scan over the figures of your work and then decide whether or not to continue reading. Therefore, make the figures largely self-explanatory. All images and captions should be easily digestible, such that the reader absorbs the main message without effort.

Figure 3 was improved from Fig. 2 by obeying these rules:

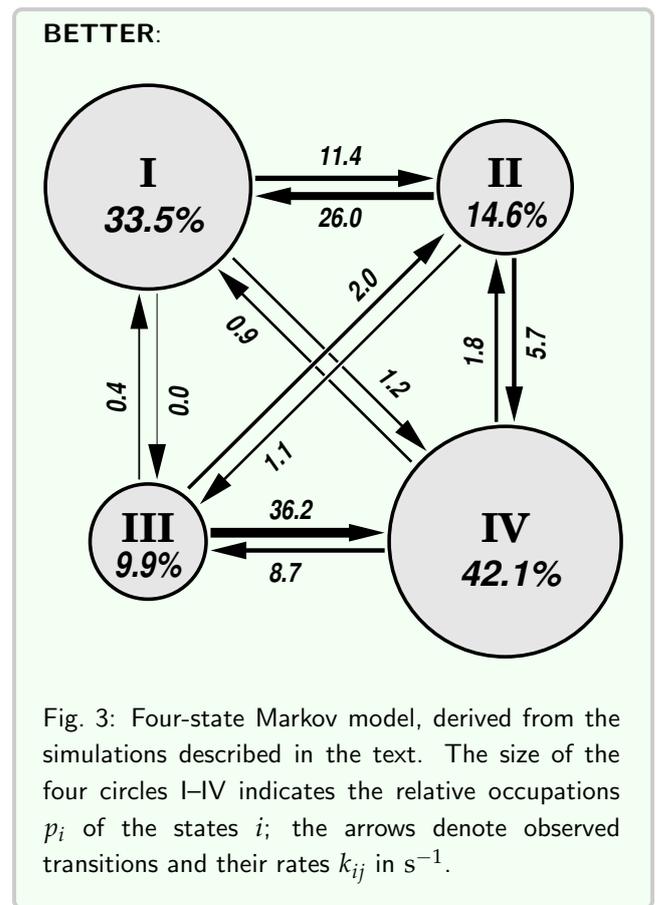
**Similar items — similar graphics; different items — different graphics.** Group similar things by using similar graphical elements; distinguish different things from each other by using different graphical elements. Also, group things that logically belong together by using a similar style or spatial proximity. Parts of the figure that need to be distinguished from each other should be represented differently, e.g. by different line thickness, hatching, colour, saturation, borders, font, or size. It is very helpful for the reader to graphically reflect the values of quantities, e.g. via line thickness or symbol size.



**Placement of annotations.** Make sure that the annotations of graphical elements (e.g. the rates in Fig. 3) are clearly and uniquely associated with the graphical elements they annotate (the arrows). In most cases, careful positioning will do; if in trouble, use lines or small arrows to indicate where the annotation belongs to. In any case, similar colour etc. helps.

**Use proper fonts.** For annotations, sans serif is the best choice. Multiline text is best displayed with serifs to better guide the reader's eye along the rows of text — or, even better, avoided.

**Avoid clutter.** Particularly if many lines, arrows, and symbols are required, their careful arrangement is key, as Figs. 2 and 3 illustrate. Be prepared for many iterations, and don't be satisfied too soon! Note the small trick in Fig. 3 to avoid the intersecting arrows from merging with each other.



**Specify units.** All physical quantities require units; either in the figure (preferred) or in the caption. Choosing the proper unit avoids too many zeroes and makes it easier to compare the displayed numbers. For example, it is easier to compare 5.4 mm to 45.0 mm than 0.0000054 km to 0.000045 km. Similarly,  $5.4 \times 10^{-3}$  m may mistakenly be considered larger than  $4.5 \times 10^{-2}$  m at first sight. Never use the lazy 5.4E-3 m, and don't tell your supervisor that your favourite plot software cannot handle superscripts!

**Caption.** The caption answers two questions, and only those: (1) What is shown? (2) What do the graphical elements represent? The caption must explain *every* element of the figure. Should an element represent a physical quantity, it may be useful to relate it to the symbol used in the text, as in '*Arrows indicate transition rates  $k_{ij}$ .*' The phrase '*For an explanation see text.*' should only be used in an emergency. Further explanations, such as how the data were obtained or what they mean, do not belong to the caption.

Always consider the **final size of your figure** in print. Most Journals layout figures single column (8 cm) or double column (16 cm), sometimes 1.5 column. Whereas your super-detailed plot may look terrific on your large screen, lines and fonts may be too tiny in the final print. Therefore, *always* print your figures at their final size and make sure every element of the plot can be conveniently captured by the reader. Fonts in the figure should not be (much) smaller than in the main text. Lines are rarely too thick, and very often too thin — 1 pt (ca. 1/3 mm) should be the absolute minimum. You may also use your figures for slides, in which case your super high-resolution screen may be projected at  $1024 \times 768$  resolution or even worse. A one-pixel thick line on your screen will at best be anti-aliased to a dim grey shadow; at worst, it just vanishes.

### 3. GLOBAL LEVEL: WHAT BELONGS WHERE?

The **Introduction** starts from the reader's state of knowledge, which is often the state of knowledge the author had at the beginning of the project. It is definitely *not* the state of knowledge the author has now!

From that initial state of knowledge, the introduction guides the reader, step-by-step, to the central question of the project; anything that does not serve this purpose does not belong to the introduction. Do include contributions by others to the main question, as well as concepts and terms that are essential for understanding the main question and its relevance in the larger scientific context. *Take the reader by the hand!*

Some readers find it helpful to learn about the main finding of the paper already at the end of the introduction, as it helps them to assess the facts and pieces of evidence presented later in the text in light of the main conclusions at the end of the text. Don't be afraid to spoil the tension by revealing the 'murderer' early on. You are writing a report, not a crime story.

The **Methods Section** describes all methods (surprise!) used in the project, as well as why a particular method was preferred. Describe everything to sufficient detail such that others can repeat your experiment or calculation. Published methods do not need to be described in detail (or at all), a proper reference suffices. It may help, though, to summarise the essence of the method.

Results do not belong to the Methods Section! Make clear which of the methods have been developed by others and which are your own. Note that if you describe a method without reference, the reader assumes it is yours.

The **Results Section** presents the results — not methods, interpretations, conclusions, or opinions. This rule is at the core of science, as it strictly separates facts from the author's opinion. *All* results that the author wants to present belong to the Results Section. Depending on the Journal, a discussion part may follow a results part, provided you make it clear which part is the results and which is the discussion. To convey the logical flow of arguments, it often helps to combine several results/discussion blocks one after another, in which case the Section is denoted 'Results and Discussion'.

The **Discussion and Conclusions Section** describes what the author thinks about the results and how he

interprets them *in light of the initial questions*. It is therefore essential that every question raised in the Introduction finds an answer in the Conclusion Section; conversely, every answer in the Conclusion Section requires a preceding question in the Introduction, preferably in the same order. It hence will often be necessary to modify the Introduction to fit the Conclusion Section.

You may find it difficult to tell the difference between Discussion and Conclusion. As a general rule, the Discussion relates mainly to the results, whereas the Conclusion Section chiefly answers the central question(s) raised at the end of the Introduction. Typical questions answered in the Discussion Section are: How accurate are the results? Are the observed differences statistically significant? Which of the results are unexpected? Which hypotheses are supported, which are not? How strongly do the results support these answers? Which control experiments support which main result, and what are the remaining weak points?

At the end of the Discussion Section, the reader should have been given all pieces of evidence and their weight/reliability, which in the Conclusion Section serve to support the central answers. One or two paragraphs on more speculative aspects, such as possible generalisations, and new questions that arise as a result of your findings are often welcome.

On proofreading, it is essential that you carefully check that *all* conclusions are supported by sufficient evidence. This criterium is central to the scientific endeavor, and thus is also key to acceptance of your manuscript.

#### 4. INTERMEDIATE LEVEL: LOGICAL FLOW OF INFORMATION

**Write purposefully and motivate the reader.** Your text is not just a sequence of true statements; its purpose is to communicate. It must, therefore, guide the reader, such that at any point in the text the reader is aware of *why* he is reading what you have written, and what you are aiming at.

**POOR:** Molecular dynamics simulations are computationally quite expensive, and limited to microseconds. Many biochemical processes occur on milliseconds to seconds time scales. Here we describe methods to increase the efficiency of molecular dynamics simulations.

**BETTER:** Molecular dynamics simulations are computationally quite demanding; even the fastest currently available supercomputers can simulate only several microseconds of protein dynamics. However, many biochemical processes — particularly the most relevant ones — are much slower and occur on time scales ranging from milliseconds to seconds. Computer power increases at an amazing pace, of course, such that in 20 years many of these processes will become accessible. To study them already now, clever approximations need to be developed to drastically reduce the computational cost. This is the main aim of this work.

#### Sections and Chapters are not isolated short stories.

Not only sentences, paragraphs, but also whole sections and chapters need to be written in a coherent sequence. Therefore, whenever you start writing a new section or chapter, you need to have in mind the precise state of knowledge of your reader *at this point*. Scanning the previous pages is thus a *must* and avoids starting with ‘Adam and Eve’ over and over again, just repeating parts of the Introduction or, even worse, supplying information that should have been provided already in the Introduction. Neither the Methods nor the Results Section starts with the motivation of your work. Only if really necessary, just briefly remind the reader of the main aim at the start of the Results Section, and point him to the current subgoal from time to time. By all means, avoid the ugly ‘*As already mentioned in Sec. 3.4, ...*’; should your supervisor ask you to better connect this part of the manuscript to a previous one, don’t even think of this phrase!

**Every paragraph describes one step in your line of thought**, no more, no less. A good way to test if this is the case is to summarise every paragraph in one sentence that captures its content. As an additional bonus, reading the resulting summary sentences in a

row is a powerful way to check if your line of thought is consistent and complete. Every paragraph typically contains a key statement, possibly preceded by a topical sentence, a transition from the previous paragraph, or a question leading to the core statement. The paragraph may also contain evidence, examples, or a logical deduction of the core statement, plus a possible transition to the next paragraph.

**Avoid abstract-deductive writing.** Abstract definitions may be correct and even mathematically elegant, but most often they are equally indigestible. Better first illustrate the definition or concept with an example or analogy, remaining short of 100% precision, and add the exact or more general formal definition *subsequently*. Pictures are particularly helpful here. A common pitfall is to ‘translate’ the formal definition into verbal computer code, instead of explaining the concept as if you were talking to a human being.

**Poor:** The shells  $S_i$  were chosen as follows:

$$S_i := \{ A \in H_2O \mid \exists X_p \exists Y \in \text{same molecule:}$$

$$d(X_p, Y) \leq i\Delta r \wedge Y \notin S_j, j < i \},$$

i.e. shell  $S_i$  contains all atoms  $A$  of the solvent, for which there is a protein atom  $X_p$  and an atom  $Y$  of the same molecule, the distance between which is smaller than  $i\Delta r$ , and for which atom  $Y$  is not already part of a smaller shell. This procedure yields the concentric shells shown in Figs. 3.1 and 3.2., each of thickness  $\Delta r$ , which wrap around the protein like onion skins.

**POOR:** To test whether observed differences in helicity are significant, we proceeded as follows: Given two samples of size  $n_1$  and  $n_2$  from two random experiments in which an event  $A$  (e.g. “residue helical”) may occur  $k_1$  and  $k_2$  times. The respective probability estimates read  $p_i = k_i/n_i, i = 1, 2$ . Defining the joint probability  $p := (k_1 + k_2)/(n_1 + n_2)$ , generate  $B$  synthetic sample pairs  $(\mathbf{x}^{*b}, \mathbf{y}^{*b})$ , with  $\mathbf{x}^{*b} = (x_1^{*b}, \dots, x_n^{*b})$  and  $\mathbf{y}^{*b} = (y_1^{*b}, \dots, y_n^{*b})$ , each  $x_i^{*b}$  and  $y_i^{*b}$  be assigned 1 with a probability  $p$ . Set  $k_1^{*b} := \#\{x_i^{*b} \mid x_i^{*b} = 1\}$ ,  $k_2^{*b}$  similar, and calculate the respective value of the test quantity  $t = |p_1 - p_2|$ . Estimate the significance according to  $\alpha_b = \#\{t \mid t \geq t_{obs}\}$ , where  $t_{obs}$  is the value of the test quantity for the original sample.

**BETTER:** To test whether observed differences in helicity are significant, we proceeded as follows: For two structure samples of sizes  $n_1$  and  $n_2$ , let the studied residue be in helical conformation in  $k_1$  and  $k_2$  cases, respectively, yielding probability estimates  $p_i = k_i/n_i, i = 1, 2$ . With the null hypothesis that the underlying probabilities are equal,  $p = (k_1 + k_2)/(n_1 + n_2)$  yields the best probability estimate. This estimate was subsequently used to generate bootstrap samples, yielding empirical distributions for  $k_1$  and  $k_2$  and, therefore, also for the test quantity  $t = |p_1 - p_2|$ . The fraction of  $t$  that exceeds the observed difference yields the significance.

**Beware of the ‘Golden-Gate effect’.** Imagine you see the Golden Gate Bridge for the first time. You are excited and describe it to your friend: “Wow, it’s huge and tall, with two super-elegant pillars that carry the whole bridge, all six lanes, via super-massive steel cables. And the colour! This beautiful reddish colour against the sunset, and ... oh, by the way, here’s a photo I took!” — I very much guess your friend would think: “Damn, why didn’t this jerk show me the picture right away???”. I call this the *Golden Gate effect*:

**POOR:** The energy landscape as a function of a conformational coordinate is hierarchically structured. There are low energy barriers which are crossed rapidly, and higher barriers which are crossed more slowly. Figure X shows a sketch of this scenario.

**BETTER:** Figure X illustrates the hierarchical structure of the energy landscape. It shows the energy of the system as a function of a conformational coordinate. Low barriers (A, B, C) are crossed rapidly (large arrows), whereas higher barriers (D, E, F) are crossed more slowly (small arrows).

**Provide information at the point where it is required,** not thereafter. If you realise upon proofreading that an important piece of information is missing, inserting it at the right position — where the reader expects it — often will require you to re-assess and possibly re-organise the text. You may be tempted, instead, to supplement the piece *after* the respective thought step, which is much easier — for you, of course, but not for

the reader. Don't give in to this temptation. Good indicators that you did are *'in other words'* and *'i.e.'*, as well as exuberant chains of subclauses.

**POOR:** We introduced a committor regression scheme as a principled means to find useful reaction coordinates, i.e. good committor correlates, within the pool of collective variables.

**BETTER:** We introduced a committor regression scheme as a principled means to find reaction coordinates that best reflect the reaction path within the pool of collective variables.

**POOR:** The aim of the simulations was to test the hypothesis, derived from experiments, that at these forces the deformation of bond or torsion angles becomes relevant for the elasticity of the polymer, because the measured force/extension curve could not exclusively be explained in terms of entropic forces, which are required to unfold a polymer from its folded state, and which can be accurately described by the 'wormlike chain model', such that in addition to the entropic effects one more component seemed to be involved.

[I did not make up these examples. All of them are from real manuscript drafts, written with the sincere intention of communicating.]

**BETTER:** At low extension forces, the elasticity of a polymer is a purely entropic effect, as evidenced by the excellent agreement between the 'wormlike chain model' and force/extension curves that have been measured in AFM experiments. At larger forces, however, deviations were observed, which suggested that additional forces become relevant for the elasticity of the polymer. One idea was that these are due to the deformation of bond or torsion angles. The simulations described below served to test this hypothesis.

**Avoid Textbook Style** (except in textbooks). Don't brag about everything you have learned, don't preach! You are not writing a textbook to educate others, you report about your work. Just tell the reader everything that is important for her to know to understand

what you did and how you did it. No less, no more. Watch out for phrases such as *'can be'*, *'may be'*, *'one can'*, *'it is practical'*, *'usually'*, etc.! These indicate that you slipped into textbook mode.

**POOR:** For open shells, one usually uses [...]. For closed shells, one can [...]. It is further practical to [...]. This approximation is justified when [...].

[Reader thinks: Nice to know, but what did the authors *actually do*?]

**BETTER:** For open shells, we used [...]. For the present case of closed shells, this approximation is not sufficiently accurate, and therefore we resorted to [...]. Whenever [...], we took advantage of [...]. We note that this approximation is justified in our case because [...].

**Provide evidence** for whatever you state.

**POOR:** It is widely known that Eq. (2) holds.

**BETTER:** Eq. (2) holds [23].

**POOR:** Obviously, Eq. (2) holds.

**BETTER:** Equation (2) follows from generalising Eq. (1) to non-commuting groups.

All **abbreviations, technical terms, and mathematical symbols** must be defined at their first appearance, usually only once.

## 5. LOCAL LEVEL: SENTENCES AND WORDS

### 5.1. Precision and Clarity

'Brain emulation' mode dictates to use as **accurate and specific terms** as possible and to avoid blurry and general terms. To this end, always double-check the range of meanings and usage context of the word or phrase you chose (non-native speakers: use a dictionary!), and do not stop your search until (a) this range

is as narrow as possible, (b) it does not evoke any unwanted connotations, and (c) what you actually want to communicate is right at its center.

It may also turn out in the process that you do not have as clear a picture in your mind of what you want to say as you may have thought. Don't give up here — clarify your thoughts and keep searching for the best terms.

The harder it is for you, the easier it will be for your reader.

Put **known or related information before new information**. Many sentences contain two pieces of information, one that relates to what has been said before, and a second one that adds new information. Always put the former first, the latter last!

This is a very important rule, violation of which is likely to produce text that reads clumsily and is hard to comprehend, often without obvious reason. The diagnosis is that new information was given *before* the familiar one, and the cure is a simple swap. More often than not, you will be pleased to find that the improved sequence reveals redundancies and thus enables substantial shortenings.

**POOR:** An average distance of 5 nm was determined from the trajectories. Via bootstrapping, the uncertainty of this distance was estimated. A significance level of 5 % was determined using this estimate.

**BETTER:** From the trajectories, an average distance of 5 nm was determined. The uncertainty of this distance was estimated via bootstrapping. Using this estimate, a significance level of 5 % was determined.

**EVEN BETTER:** From the trajectories, an average distance of 5 nm was determined. Its uncertainty was estimated via bootstrapping, yielding a significance level of 5 %.

**Put the main statements at an emphasis position.**

Make sure that the important statements in a sentence are where the emphasis is: at the beginning or the end of the sentence, and in any case in the main clause, which is preferably short. Subordinate clauses, expressions in dashes or brackets, or the middle of a long sentence are no emphasis positions.

**EXAMPLE:** On July 20<sup>th</sup>, 1969, Buzz Aldrin set foot on the moon, shortly after Neil Armstrong, who was first.

Meticulously check for **correct semantics**. Particularly those sentences that read especially elegant often turn out to be semantically plain wrong. This short sentence actually contains four(!) semantic errors:

**POOR:** Thereby, the free energy landscape turns into a highly complex hyperplane, within which dynamical statements are hard to make.

[Homework: Find the four semantic errors!]

**POOR:** To separate the electronic degrees of freedom from the nuclear motion, I first introduce the Born-Oppenheimer approximation.

**BETTER:** The Born-Oppenheimer approximation serves to separate the electronic degrees of freedom from the nuclear motion.

**Same word for the same thing; different words for different things.** Your English teacher may have told you that repeating the same word is dull and boring and, therefore, you should always use synonyms. I strongly disagree! The reader does not know whether your synonym refers to a new thing or not, and gets confused. So: Consistently use the same word for the same thing, and — equally important — use different words (and math symbols!) for different things! For less important words, though, you may consider using synonyms.

**POOR:** A group of lipids was extracted from the set of membrane lipids. Subsequently, the center of mass was calculated from this pool.

**BETTER:** A group of lipids was extracted from the set of membrane lipids. Subsequently, the center of mass was calculated from this group.

**POOR:** Next, we calculated A; next, we calculated the average.

**BETTER:** Next, we calculated A, then the average.

Of course, repetitions *can* be dull and boring. If the author had read this text aloud, he would have noticed:

**POOR:** Regardless of the importance of ionic solutions for a broad field, the nature of the impact of ions on the water molecules is still not understood on the molecular level. Especially, in recent years a discussion evolved about the water-ion interaction on the molecular level. In this study, using first principles calculations, we elucidate water-ion interaction on the molecular level.

**Avoid vague words.** Indicators of fuzzy writing are ‘*corresponding*’, ‘*relate to*’, ‘*data*’, etc.:

**POOR:** The binary trace of the permeation signal corresponds to the channel states open (‘o’) and closed (‘c’), respectively.

**BETTER:** When permeation is seen, the channel is open (‘o’), otherwise closed (‘c’).

**Linking words** such as ‘*however*’, ‘*nevertheless*’, ‘*in fact*’, etc. tell the reader the *purpose* of a (sub-)clause and help to establish the proper context. They avoid misinterpretation and sometimes even unintended humour.

**POOR:** The main focus of this study has been on  $x < x_0$ ;  $x > x_0$  might yield some interesting results.

**BETTER:** The main focus of this study has been on  $x < x_0$ , but  $x > x_0$  might also yield interesting results.

**However vs. in contrast.** The latter announces a contrast or difference to previously mentioned *facts*; in contrast, ‘*however*’ introduces a turn in the *line of thought* or an *argument* against a previous statement.

**POOR:** From the simulation, we obtained  $\sigma = 3$ ; however, the x-ray analysis yields  $\sigma = 9$ .

**BETTER:** From the simulation, we obtained  $\sigma = 3$ ; in contrast, the x-ray analysis yields  $\sigma = 9$ .

**POOR:** The value  $\sigma = 3$  suggests an exothermic reaction. In contrast, this assumption is inconsistent with the previous results.

**BETTER:** The value  $\sigma = 3$  suggests an exothermic reaction. However, this assumption is inconsistent with the previous results.

**The placement of ‘however’** requires attention.

**AMBIGUOUS:** Like the multi-electron problem, the electromagnetic field is a multi-particle state, however, in contrast to the electronic case, it consists of non-interacting particles.

**A BIT BETTER:** Like the multi-electron problem, the electromagnetic field is a multi-particle state; however, in contrast to the electronic case, it consists of non-interacting particles.

**STILL BETTER:** Like the multi-electron problem, the electromagnetic field is a multi-particle state, but in contrast to the electronic case, it consists of non-interacting particles.

**EVEN BETTER:** Like the multi-electron problem, the electromagnetic field is a multi-particle state. However, in contrast to the electronic case, it consists of non-interacting particles.

**Definite vs. indefinite article.** Textbooks suggest to use the definite article ‘*the*’ to refer to a *specific* or *particular* noun, the indefinite article ‘*a*’ otherwise.

While this rule may be helpful in some cases, you may find it difficult to apply to others. Instead, consider using the definite article only if the reader already has in mind what it is referring to.

In this sentence, the definite article reveals that the author confused her state of knowledge (that there was only one spectrometer at her institute) with that of the reader:

**POOR:** The measurements were carried out with the 100-nm-spectrometer.

**BETTER:** The measurements were carried out with a 100-nm-spectrometer.

[The reader does not know what's in your institute.]

Often additional ambiguity arises when using the definite article in the plural, as it is unclear whether you refer to all items or only a subset.

**POOR:** Simulations were carried out using GROMACS.

**STILL POOR:** The simulations were carried out using GROMACS.

**BETTER:** All simulations were carried out using GROMACS.

**Expectations** are always yours, not necessarily everybody's or the reader's, which should be reflected by the text.

**POOR:** It cannot be expected that this error affects our results.

**BETTER:** We do not expect this error to affect our results.

**Computer programs** are just tools to implement a numerical method or algorithm; they do not represent a method by themselves. Emphasis should therefore be given to the methods rather than to the (often less well-known) name of the software.

**POOR:** Principal components were calculated using the HGSRYE module of the `grombl 2.1 $\beta$`  package, which diagonalises the covariance matrix.

**BETTER:** Principal components were calculated by diagonalising the covariance matrix using the HGSRYE module of the `grombl 2.1 $\beta$`  package [42].

In physics, **classical** is ambiguous.

**POOR:** Classical molecular dynamics simulations were carried out.

**BETTER:** Force field based molecular dynamics simulations were carried out.

**Significant vs. large.** In the quantitative sciences, the primary meaning of 'significant' is 'statistically significant', not just 'large'.

**POOR:** We were unable to estimate the uncertainty of the measured lengths, but observed a significant difference.

**BETTER:** [...], but observed a large / pronounced / marked difference.

'**Data**' is always too vague. Unless the context makes it crystal clear, always state explicitly what type of data is meant.

**POOR:** The data were recorded every 10 ps.

**BETTER:** Atomic positions were recorded every 10 ps.

**While** means both 'during' (temporal) and 'whereas' (contrasting). For the latter meaning, use 'whereas'.

**POOR:** While the conclusion is correct, the results are wrong.

**BETTER:** Whereas the conclusion is correct, the results are wrong.

Also **Since** has both a temporal and a logical meaning. For the latter, use 'because' or (if unambiguous) 'as'.

**POOR:** This approach is more convenient, since it saves one step.

**BETTER:** This approach is more convenient, because it saves one step.

**Avoid the Germanism ‘could’** (German: ‘konnte’). In English, the word ‘could’ primarily means ‘might’ (conditional, ‘könnte’), and, less common, ‘was able to’ (past tense, ‘konnte’). You risk a dramatic and unintended shift of meaning:

**MISLEADING:** I could solve the problem.

[will most likely be understood as *I might solve the problem, but didn’t yet.*]

**BETTER:** I was able to solve the problem.

**Therefore** is particularly tricky. In many cases, it seems to fit at first sight but actually doesn’t. ‘Therefore’ always announces a logical consequence (‘for this reason’), not a motivation or aim.

**POOR:** In a second step, an increased resolution was required. Therefore, additional measurements were performed.

**BETTER:** In a second step, an increased resolution was required. To this end, additional measurements were performed.

Though seemingly innocent, the word ‘therefore’ might even suggest scientific misconduct:

**DANGEROUS:** Theory predicts decreased variations. Therefore, a small  $\sigma = 3 \text{ \AA}$  was measured.

**BETTER:** Theory predicts decreased variations. Accordingly, a small  $\sigma = 3 \text{ \AA}$  was measured.

For a theoretician, **experimental** is just the opposite to calculations. Be more precise, though:

**POOR:** The figure shows calculated and experimental ion currents.

**POORER:** The figure shows calculated and experimentally measured ion currents.

[Pleonasm]

**BETTER:** The figure shows calculated and measured ion currents.

**EVEN BETTER:** The figure shows calculated ion currents and those measured by single molecule electrophysiology on oocytes.

## 5.2. Brevity

“Every word of your text has to ‘perform work’; if it doesn’t, remove it!”\*

I consider this rule super-helpful.

It is quite surprising how simple it is to shorten wordy writing. Often, just removing a few words will produce a more forceful text.

You may think that a complex thought or concept requires similarly complex language to be properly appreciated by the reader. Quite the opposite: How can I value something I don’t understand? More often than not, complex language is used to camouflage or replace unclear thinking.

If your line of thought is clear, so is your text.

### Avoid redundancies.

**POOR:** The idea of a principal component analysis is to characterise the principal components of a data set.

**BETTER:**

[Remove the sentence.]

\*Wolf Schneider, former Director of the Henri Nannen school for journalists.

**Avoid double-negative.**

**POOR:** The single electronic motion is not independent of the other electrons but is correlated to them.

**BETTER:** The motions of the electrons are correlated.

**'Not'** can often be avoided using a prefix.

**POOR:** Rotational symmetry does usually not suffice to [...]

**BETTER:** Rotational symmetry is usually insufficient to [...]

**Trends** should be described as such and not via start and end points.

**POOR:**  $y$  is large for small  $x$  values and small for large ones.

**BETTER:**  $y$  decreases with increasing  $x$ .

Terms such as **algorithm** or **method** are often redundant.

**POOR:** The structure was optimised using the steepest descent algorithm.

**BETTER:** The structure was optimised using steepest descent.

**'In the case of'** should be avoided.

**POOR:** In the case of small velocities,  $x$  vanishes.

**BETTER:** For small velocities,  $x$  vanishes.

**'Was chosen to be'** can be compacted.

**POOR:** The width  $\sigma$  was chosen to be 5 nm.

**BETTER:** A width of  $\sigma = 5$  nm was chosen.

**'In order'** can often be removed.

**POOR:** In order to test this hypothesis, the variance was calculated.

**BETTER:** To test this hypothesis, the variance was calculated.

**'In addition to'** is weak and can be made more forceful.

**POOR:** In addition to the measurements, also the simulations supported the hypothesis.

**BETTER:** Not only the measurements, but also the simulations supported the hypothesis.

Avoid **'Given in Ref. [12]'**.

**POOR:** The OPLS force field parameters given in Ref. [12] were used.

**BETTER:** OPLS force field parameters [12] were used.

### 5.3. Style

Good writing style makes reading easy, efficient, and joyful. It is therefore not a matter of taste, nor is it surplus luxury. It is a duty.

There are many great books on good English usage [6–12], and I am certainly not an expert. Therefore I list only the most frequently reoccurring issues.

**Avoid long sentences.** If carefully composed, complex sentences *may* be comprehensible and even elegant — but don't rely on your ability to do so.

**FAIL:** Here, it was used that  $[H, r] = -ip$  and the fact that the wavelength of the electromagnetic wave is much larger than the extent of the molecular wave function such that  $A \approx A_0$  with  $r_0$  giving now the position of the total molecule, i.e. the center of mass of the molecule.

Better play safe and split it into several shorter sentences. As an additional benefit, you will be forced to re-assess and improve your line of thought beyond the suboptimal level often reflected by your lengthy sentences.

Don't overact, though. A row of short main clauses reads boringly — unless you are Hemingway. Ideally, your text is a vivid mixture of short and slightly longer sentences.

**POOR:** The sequence of energy conversion steps carried out within the  $F_1F_0$ -ATP synthase, first from electro-osmotic energy to mechanical energy by the  $F_0$ -subunit, the presumably mechanical energy transfer to the binding sites within the  $F_1$ -subunit via the  $\gamma$ -subunit, followed by the conversion into chemical energy, as well as the fascinating functional mechanisms involved, a rotating axis, which triggers synthesis within the three adjacent ATP binding sites, combined with a near 100 % efficiency, suggest the term 'molecular machine'

**BETTER:** In summary, the  $F_1F_0$ -ATP synthase is an energy converter involving several conversion steps. First, the  $F_0$ -subunit converts electro-osmotic energy into mechanical energy, which is transferred to the binding sites within the  $F_1$ -subunit via the  $\gamma$ -subunit, and finally converted into chemical energy. Mediated via a rotating axis, ATP is synthesised within each of the three adjacent binding sites in a highly cooperative manner. Combined with a near 100 % efficiency, this complex mechanism suggests the term 'molecular machine'

**Avoid nominalizations** (the German disease). Verbs describe actions and nouns refer to things, as in 'We investigated the sample'. If nominalised to 'We performed an investigation of the sample', however, the action is described by the noun 'investigation', and the actual verb 'performed' does not carry any further information; it is just there because English grammar requires a verb. The result is very weak, clumsy, and tiring text.

Watch out for endings such as '-ised' and '-tion', or empty verbs like 'perform', 'provide', 'correspond', 'exhibit', 'carry out', 'take place', 'exert', 'execute', or 'effect', 'possess', etc.!

**POOR:** This force results in an acceleration motion of the particle, onto which water molecules exert a deceleration effect.

**BETTER:** This force accelerates the particle, which is slowed down by water molecules.

**EVEN BETTER:** Accelerated by this force, the particle is slowed down by the water molecules.

**POOR:** [...] provide a reliable determination of the molecular transform.

**BETTER:** [...] reliably determine the molecular transform.

**POOR:** The results are in agreement with experiment.

**BETTER:** The results agree with experiment.

**The passive voice is not forbidden.** Don't avoid the passive voice by all means, as your teacher may have told you. If the grammatical agent is really unimportant, the passive voice is preferred.

**POOR:** We incubated the sample at 37 C for 45 minutes.

**BETTER:** The sample was incubated at 37 C for 45 minutes.

Also, the passive voice can serve to put the focus on where it should be:

**POOR:** The main resources in the ocean are plankton and oxygen. Whales eat plankton, whereas fish need oxygen.

**BETTER:** The main resources in the ocean are plankton and oxygen. Plankton is eaten by whales, whereas oxygen is needed by fish.

Don't overdo it, though! Consider alternatives:

**GOOD:** The third chapter describes the main results.

Use **simple language**. Prefer a simpler word whenever you can. Often this will be one with a Germanic instead of a Latin root.

**POOR:** Due to the fact that the experimental data acquisition was erroneous on numerous occasions, we discontinued the investigation and utilised ameliorated equipment.

**BETTER:** Because the measurements were often wrong, we stopped the study and used a better ruler.

However, **avoid lab slang**.

**POOR:** We took 300 structures from the simulation.

**BETTER:** We extracted 300 structures from the simulation.

**POOR:** Energy values were written out every 100 steps.

**BETTER:** Energy values were recorded every 100 steps.

## 6. MINOR ISSUES

Why should you care about trifles? After all, you are a scientist who has more important things to do.

In fact you should, for three reasons.

First, each spelling error, grammar slip, or missing comma triggers an ‘interrupt signal’ in the reader’s brain, which is annoying and distracts from the content. Many colleagues find it quite difficult and exhausting to focus on the scientific content of a draft in poor English. Second, a wrong word at a critical position can easily misdirect the reader. Third, the reader — e.g. a referee of your manuscript — may consider you a careless writer. Not knowing anything about how carefully you perform your science, what do you think the referee would assume?

**Equations and mathematical expressions are part of a sentence**, which dictates proper punctuation.

**POOR:** [...] which follows from the equation.

$$ma = F \quad (1)$$

**BETTER:** [...] which follows from the equation

$$ma = F. \quad (2)$$

**Vectors.** In hand-writing, vectors are usually indicated by an over-arrow or underline; however, most Journals and books print vectors and matrices in boldface.

**WRONG:**  $\vec{x}$ ,  $\underline{x}$

**CORRECT:**  $\mathbf{x}$

In most cases, indices of vectors are not vectors.

**MOST LIKELY WRONG:**  $\sum_{i=1}^n \mathbf{x}_i$

**MOST LIKELY CORRECT:**  $\sum_{i=1}^n \mathbf{x}_i$

**Subscripts** in mathematical expressions. In math mode,  $\LaTeX$  treats each letter as a mathematical symbol. For subscripts that are mathematical symbols, such as in  $a_{ij}$ , this is fine; however, for words or abbreviations in subscript you have to switch to text mode to avoid italics and strange spacing.

**WRONG:**  $N_{affinity}$   $\LaTeX$ :  $\$N_{\text{affinity}}\$$

**CORRECT:**  $N_{\text{affinity}}$   $\$N_{\text{\textterm{affinity}}}\$$

**Citations are like footnotes** and cannot be the subject of a sentence.

**POOR:** In [20,42] it was shown that both approaches are valid.

**BETTER:** As shown by Zwanzig [20,42], both approaches are valid.

**EVEN BETTER:** Both approaches are valid [20,42].

### No abbreviations at the beginning of a sentence.

Most Journals abbreviate Fig., Tab., Eq., Sec., etc., except as the first word of a sentence. Also, use a non-breaking space in 'Fig. 5'; e.g. in  $\LaTeX$ , write 'Fig. ~5'. Similarly, use a non-breaking ('protected') space between a number and its unit. E.g. in  $\LaTeX$ , write '34~nm'. If you really must use Word, write '34°nm'.

**POOR:** Fig. 2a compares the achieved accelerations of the previously recorded particles to the ones in Tab. 5.

**CORRECT:** Figure 2a compares the achieved accelerations of the previously recorded particles to the ones in Tab. 5.

**Hyphens and Dashes.** There are three different dash types. The hyphen or divis ('-',  $\LaTeX$ : -) is used for concatenating nouns and to split words at the end of the line. Ranges between two numbers are indicated by the slightly longer n-dash ('-',  $\LaTeX$ : --), and the m-dash ('—',  $\LaTeX$ : ---) separates parts of sentences.

**POOR:** The cut-off is chosen - typically - in the range 10—12 Å.

**BETTER:** The cut-off is chosen — typically — in the range 10–12 Å.

**Short space after an abbreviation.** A long space is used after the period at the end of a sentence;  $\LaTeX$  takes care of that automatically. A short space follows the period of an abbreviation, however, and you need to tell  $\LaTeX$  using '\ ' or '~'.

**POOR:** To see this, cf. Fig. A. [cf. Fig. A.]

**BETTER:** To see this, cf. Fig. A. [cf.\ Fig.~A.]

## 7. READY? – NOT YET!

So your draft is ready, finally! Off it goes to your supervisor, or to the Journal right away.

Not so fast!

Pause for one or a few days to erase your brain. Become a reader who has not seen the text before and does not know anything about your project.

'Test-run' the whole text on this reader's mind; emulate her mind and meticulously keep track of how her memory changes with each piece of information given by your text. For any term and expression of your text, don't assume the reader's mind will get a copy of what you have in mind and want to communicate. Instead, ask yourself what the most likely association of the term will be in *light of the current state of the reader's mind*. As non-native speaker, check a dictionary, or google in which context your term or expression is most often used. Check that all terms, concepts, and methods are explained at the appropriate level of detail. See if all logical steps are laid out in proper sequence, and check for redundancies.

For this test-run to work, it is essential that you read the whole text in one run, uninterrupted, without distraction, and well-rested. Don't attempt to cure any deficiencies of the text right away — just mark the position and read on, to the end of the text. Only thereafter start improving the text. Iterate until the emulated reader's mind is exactly where you want it to be.

Scientific writing is a programming job — don't sell untested software!\*

### LAST RULE:

You may deviate from any rule — *if you know precisely why*.

\*Are you experiencing a sense of déjà vu right now? Hasn't this trick been already mentioned? Correct. I really want to make sure you apply this rule. It is so important, but rarely mentioned in textbooks.

(Solution to the semantics puzzle on p. 15: (1) Energy landscapes do not *turn* into hyperplanes, they are *defined* on hyperplanes; (2) it is the energy landscape which is complex, not the hyperplane; (3) statements may be true or false, but rarely dynamical — what the author had in mind is a statement *about* dynamics; (4) as far as I know, no statement has ever been made *within* a hyperplane.)

## APPENDIX: FREQUENT MISTAKES

Do use a spell-checker before giving your manuscript to anyone else! However, these common mistakes may remain undetected:

**Incorrect tense:** Experiments and observations by you or others were performed *before* you wrote the paper (hopefully!) and thus are described in past tense. The conclusions you infer from those hold true at any time and, therefore, are written in present tense.

**POOR:** All simulations are performed under constant pressure conditions, which suggested that transcription activation was entropy-dominated.

**CORRECT:** All simulations were performed under constant pressure conditions, which suggests that transcription activation is entropy-dominated.

**Which vs. that.** A subordinate clause that can be omitted without rendering the main clause meaningless starts with '*which*'. If the subordinate clause is essential for the definition of the subject or object of the main clause, use '*that*'.

**POOR:** We used the simulation, which showed the smallest fluctuation.

**BETTER:** We used the simulation that showed the smallest fluctuation.

**POOR:** From the simulations, we calculated average fluctuations that helped us to avoid the experiment.

**BETTER:** From the simulations, we calculated average fluctuations, which helped us to avoid the experiment.

In some cases, both is possible, but your choice may change the meaning:

**GOOD:** The 10 ns simulation, which showed the largest fluctuations, was used.

[There is only one 10 ns simulation and you provide additional information.]

**ALSO GOOD:** The 10 ns simulation that showed the largest fluctuations, was used.

[There are several 10 ns simulations and you *define* which of those you used.]

**Each vs. every vs. all:** '*Each*' refers to an individual thing in a group, '*every*' to several.

**POOR:** Every simulation was analysed independently. Each simulation showed this feature; in this respect, every simulation was similar.

**CORRECT:** Each simulation was analysed independently. Every simulation showed this feature; in this respect, all simulations were similar.

'**This**' preferably refers to a specific object, not to whole statements.

**POOR:** A large error was obtained. This casts doubt on the conclusion.

**BETTER:** A large error was obtained. This result casts doubt on the conclusion.

Many non-native speakers find the use of **respectively** a bit strange.

**POOR:** Solid and dashed lines show minimisation and equilibration.

**BETTER:** Solid and dashed lines show minimisation and equilibration, respectively.

**On the other hand ...** must be preceded by 'On the one hand ...'. Should the latter seem somehow inappropriate, so is — most likely — the former. Use a different linking phrase instead.

**POOR:** From the simulations, large radii of gyration were obtained. On the other hand, the measurements yielded much smaller values.

**BETTER:** From the simulations, large radii of gyration were obtained. In contrast, the measurements yielded much smaller values.

#### Place vs. position vs. location vs. coordinate.

**WRONG:** The place/location of the atom.

**CORRECT:** The position of the atom.

**POOR:** The center of mass has a more positive  $x$ -position.

**BETTER:** The  $x$ -coordinate of the center of mass is more positive.

#### Space vs. room vs. region.

**POOR:** There is not much space for improvement.

**BETTER:** There is not much room for improvement.

**WRONG:** The eigen vectors define the conformational room.

**CORRECT:** The eigen vectors define the conformational space.

**WRONG:** The ion moves into a room with a strong electric field.

**BETTER:** The ion moves into a region with a strong electric field.

#### Plane vs. surface vs. area.

**WRONG:** The surface of this plane is  $10\text{ m}^2$ .

**CORRECT:** The area of this plane is  $10\text{ m}^2$ .

**WRONG:** The two vectors lie within the same surface.

**CORRECT:** The two vectors lie within the same plane.

#### Differ vs. vary. Vary refers to changes over time.

**POOR:** Simulations were carried out for varying resolutions.

**BETTER:** Simulations were carried out for different resolutions.

**POOR:** In the simulations, the spring position differed with time.

**BETTER:** In the simulations, the spring position varied with time.

#### Comparable does not mean similar; it just means that things can be compared.

**POOR:** The calculated force was comparable to the measured one.

**BETTER:** The calculated force was similar to the measured one.

#### Affect vs. effect.

**WRONG:** This affect effects the results.

**BETTER:** This effect affects the results.

[just to illustrate the words, still twists your tongue ...]

#### High vs. large.

**POOR:** The value turned out to be rather high.

**BETTER:** The value turned out to be rather large.

['High values' would refer to moral values!]

**OK:** The energy barrier is rather high.

**In front of** establishes a relation in *space*, **before** in *time*.

**WRONG:** Minimisation was carried out in front of equilibration.

**CORRECT:** Minimisation was carried out before equilibration.

**WRONG:** Note the flexible gate before the binding site.

**CORRECT:** Note the flexible gate in front of the binding site.

**Intend vs. intent vs. indent.**

**WRONG:** We intended to typeset with more intends. At least, that was our indent.

**CORRECT:** We intended to typeset with more indents. At least, that was our intent.

**Calculate/compute vs. evaluate.**

**POOR:** This trajectory was evaluated several times.

**CORRECT:** This trajectory was computed several times.

**POOR:** This expression was calculated for different parameter sets.

**CORRECT:** This expression was evaluated for different parameter sets.

**Compared to vs. compared with:** Use 'to' if the emphasis is on similarity, and 'with' if you want to point to a difference.

**POOR:** Brain tumors are relatively rare compared to more common cancers.

**CORRECT:** Brain tumors are relatively rare compared with more common cancers.

**Necessary vs. required:** The former means something is essential and cannot be omitted for the desired result; the latter refers to rules or regulations that could be waived.

**POOR:** To drive a car, fuel is required and a driver's license is necessary.

**CORRECT:** To drive a car, fuel is necessary and a driver's license is required.

**Ideal vs. optimal:** The best option irrespective of any restriction is the ideal one, under *given* circumstances you achieve the optimal one.

**POOR:** Given the low bandwidth, our software achieves ideal scaling.

**CORRECT:** Given the low bandwidth, our software achieves optimal scaling.

**Include vs. add.**

**EITHER:** One hundred water molecules were added to system A.

[Result: system A + 100 water molecules]

**OR:** One hundred water molecules were included within system B.

[Result: system B unchanged]

**On vs. at.** The proposition '*on*' has a strong spatial connotation, thus:

**POOR:** Pronounced dynamics were seen on different time scales.

**BETTER:** Pronounced dynamics were seen at different time scales.

**WRONG:** To focus at something.

**BETTER:** To focus on something.

**Minimum:** The singular of the Latin '*minima*' is '*minimum*'; that of '*spectra*' is '*spectrum*'.

**WRONG:** Of the three minima, the second was an absolute minima.

**CORRECT:** Of the three minima, the second was an absolute minimum.

[You did not even think of '*minimums*', did you?]

**Principle vs. principal.**

**WRONG:** In principal, principle components are independent.

**CORRECT:** In principle, principal components are independent.

**ALSO CORRECT:** The principal components are principally independent.

**A vs. an.** The rule "use 'an' before a vowel" follows pronunciation, not spelling.

**WRONG:** A MD simulation.

**CORRECT:** An MD simulation.

**CORRECT:** An hour ago.

**CORRECT:** A uniform acceleration.

Read it aloud, then decide.

Finally, here are a few tips for my German friends:

The German '*Impuls*' translates into **momentum**.

**WRONG:** The force is due to a change of impulse.

**CORRECT:** The force is due to a change of momentum.

Also, the German '*Drehmoment*' translates into **torque**.

**WRONG:** We exerted too much angular momentum.

**CORRECT:** We exerted too much torque.

**Eventually vs. possibly.**

**POOR:** Eventually, the dice will show a six on the first throw.

**CORRECT:** Possibly, the dice will show a six on the first throw.

**CORRECT:** Eventually, the dropped tomato will hit the floor.

The English **Billion** is the Spanish '*millardo*', the French '*milliard*', and the German '*Milliarde*'.

**WRONG:** A second is a billion picoseconds.

**CORRECT:** A second is a billion nanoseconds.

**CORRECT:** A second is a trillion picoseconds.

**Become** is not the translation of the German '*bekommen*', but of '*werden*'.

**DANGEROUS:** I want to become a beefsteak.

**CORRECT:** I want to get a beefsteak.

**ALSO CORRECT:** The figures became larger.

## Suggested Reading

- [1] M. Alley. *The Craft of Scientific Writing*. Springer, 1998.
- [2] D. Lindsay. *Scientific Writing = Thinking in Words*. CSIRO Publishing, 2011.
- [3] J. Schimel. *Writing Science*. Oxford Univ. Press, 2011.
- [4] Hilary Glasman-Deal. *Science Research Writing for Non-Native Speakers of English*. Imperial College Press, 2009.
- [5] M. J. Katz. *From Research to Manuscript*. Springer, 2009.
- [6] W. Zinsser. *On Writing Well*. Harper Perennial, 2016.
- [7] W. Strunk Jr. and E. B. White. *The Elements of Style*. Tisot, 1999.
- [8] J. M. Williams and J. Bizup. *Style: Lessons in Clarity and Grace*. Longman, 2016.
- [9] University of Chicago. *The Chicago Manual of Style*. University of Chicago Press, 2017.
- [10] M. Swan. *Practical English Usage*. Oxford University Press, 2005.
- [11] J. Butterfield (Ed). *Fowler's Dictionary of Modern English Usage*. Oxford Univ. Press, 2015.
- [12] L. Truss. *Eats, Shoots and Leaves*. HarperCollins, 2009.

## Acknowledgements:

I am most grateful to the many students who anonymously provided the examples used in this text, and to Floris Buelens, Aljaž Godec, Bert L. de Groot, Leonard Heinz, Petra Kellers, Carsten Kutzner, Benedikt Rennekamp, Carmen Rotte, Martin Reinhardt, and John Seddon for many helpful comments and suggestions.