SCIENTIFIC WRITING 2.0

A Reader and Writer's Guide

Lebrun Jean-Luc

World Scientific

SCIENTIFIC WRITING 2.0

A Reader and Writer's Guide

The Expanded Edition with Writing Diagnosis Tool on DVD for Mac and PC

This page intentionally left blank

SCIENTIFIC WRITING 2.0

A Reader and Writer's Guide

The Expanded Edition with Writing Diagnosis Tool on DVD for Mac and PC

Jean-Luc Lebrun

Trainer of Researchers and Scientists from
A★STAR Research Institutes
Singapore

Test-tube artwork on cover by Justin Lebrun.

SCIENTIFIC WRITING 2.0
A Reader and Writer's Guide

Printed in Singapore.

Contents

Contents

Contents

Preface First Edition

You know how to write grammatically correct English. Congratulations. You have read the little book of Strunk & White "The elements of style".[1] Perfect. To pursue true writing excellence, you now need to consider the people key to your success: the readers.

What readers fear the most while reading a scientific paper is to 'get stuck' or left behind. They are stuck when the experienced writer zigzags around the familiar obstacles in his or her knowledge field, whilst readers crash into them; and they are left behind when the knowledgeable writer runs where they can only walk.

The knowledge gap that separates you from your readers cannot be ignored. Having enough background knowledge to understand your paper does not, in any way, guarantee the reader that your paper is easy and rewarding. Indeed, much more is required. A scientific paper requires more memory, attention, and time than a typical novel of the same length. Good writing should therefore take into account the reader's ignorance, fatigue, short-term memory, and impatience to reduce their impact.

Unique writing techniques rarely presented in books on technical writing will get you closer to the six qualities that are the hallmark of great scientific writing: fluid, organized, clear, concise, convincing, and interesting (FOCI). Consider sentence structure. Does placing the conjunction 'although' at the head of a sentence provide more reading pull than placing it midway? Consider sentence progression. Does keeping the topic of the first sentence constant throughout a paragraph help the reader progress smoothly through a written argument? Consider the reader's expectations. Can a single word in a sentence trigger large expectations? "Because it was raining that day," creates

---

[1] Strunk W Jr and White EB, The Elements of Style, Penguin Press, New York, 2005.

the expectation that the writer will explain what happened because of the rain. The rest of the sentence brings closure with "the paint did not dry on time." The reader reaches the end of the sentence knowing why the paint did not dry. The first expectation raised is fulfilled, but another expectation arises: the paint did not dry on time for what? Expectations drive reading forward in science as they do in literature. By creating and controlling pull, progression, and expectations, you guide your reader.

Readers have different expectations for each part of a scientific article, from its title to its conclusions. Since ignoring these expectations frustrates readers, the writer should avoid the short introduction that sheds little light on the 'what' and the 'why' of the paper, the abstract that is indistinguishable from the conclusions, the misleading title, the baggy structure, and the immature and unprocessed visuals. This book will help you learn how to put together a coherent paper that satisfies readers.

Fifty plus stories designed to liven up reading, and a hundred plus examples inspired or quoted from scientific articles add spice to this book. Do not let their scientific jargon intimidate you. What matters in these examples is syntax, word placement, and expectations raised.

This book was written at the request of the many scientists who have participated in the scientific writing skills seminars I conduct in various parts of the globe in French or in English. In their assessment of the course, the participants highlighted benefits, some expected and some unexpected. As foreseen, those who had already published papers felt that their writing had improved by keeping the reader in mind. Junior scientists without any publishing experience were relieved that they no longer had to blindly imitate the work of others, not knowing whether what they were imitating was good or bad. Unexpectedly, even senior scientists with great publishing experience found that the seminar had improved their analytical reading skills and had equipped them with a method to conduct better peer reviews.

Before turning the page, words of appreciation are due. More than 2000 international scientists from A-Star Research Institutes helped me understand and care for the scientific reader. Three authors, through their books, influenced the contents of this book: Michael

Alley[2] on scientific writing, George Gopen[3] on reader energy and expectations, and Don Norman[4] on user interfaces. They have my deepest respect. They are the giants on whose shoulders I climbed to discover a new world they had explored well before I did. If, thanks to them, I discovered new techniques that will be of interest to you, may they share the credit.

[2] Alley M, The Craft of Scientific Writing, Springer, New York, 1997.
[3] Gopen GD, Expectations: Teaching Writing from the Reader's Perspective, Pearson Longman, 2004.
[4] Norman D, The Design of Everyday Things, Basic Books, New York, 2002.

This page intentionally left blank

Preface Second Edition

The second edition is always an exciting event for any writer. It is a partial rebirth of sorts. What required remodeling returned to the writer's workbench. Obsolete chapters were dismantled. New chapters were added to reflect the evolution of the author's knowledge that inevitably follows four years of refinement in the science of writing.

In this edition,

- I am more vocal on the need to write each part of a scientific paper with the "attitude" that matches the role played by the part. For example, the introduction and the conclusion of a paper are written in a personal way, whereas the methodology and result parts are written as one would describe a plant in a herbarium.
- I am also clearer when explaining the techniques of fluid writing. This was made possible by the thousands of scientists who identified the causes of text fluidity problems during the training seminars.
- I have expanded the role of Vladimir, our man in the story bubble, to bring home important points.
- I have added a new chapter to expand the horizon of those eager to go beyond the teachings of this book. Over 10 years of constant research on the best practices in scientific writing, I have bookmarked websites that contain excellent articles worth reading, and even printing. This chapter contains my bookmarks, and a brief description of the web pages and recommended books.

- Finally, the important guidelines have been flagged with the following symbol, ☀ and the exercises have been refined. Put into practice what you learn. You have a chance to do just that, whenever you see the coach icon below.

I have formalized quality metrics to assess a scientific paper. They have been partly embedded in a Java prototype thanks to an active student programming team at the University of Eastern Finland (UEF), Joensuu, Finland. The core programming team included four key members, Mr Henri Leisma, Ms Monika Machunik, Mr Arttu Viljakainen and Mr Teemu Turunen, all master's students at UEF. The work was supervised by Dr Tomi Kinnunen while Dr Tuomo Kakkonen provided useful feedback on the NLP methodology. The latest version of the tool, dubbed as the Scientific Writing AssistaNt (SWAN), is available for free at the following web address: http://cs.joensuu.fi/swan/. To contact the development team for further information, you can also write an email to swan@cs.joensuu.fi. You will be pleased to know that no swans or other animals were harmed during the development process of the tool. The Whooper Swan *(Cygnus cygnus)* is the national bird of Finland — and Sibelius is my favorite composer. Here is the logo of the tool.

Naturally, the team welcomes any NLP expert or Java Programmer who knows what this acronym means and would like to improve this application for the benefit of all scientists who write papers. You'll get ample credit! The tool relies on open source code and is therefore available under the terms of the GNU General Public License as published by the Free Software Foundation.[1]

---

[1] It is distributed free in the hope that it will be useful but without any warranty; without even the implied warranty of merchantability or fitness for a particular purpose. See the GNU General Public License for more details(http://www.gnu.org/licenses/).

Part 1
The Reading Toolkit

This title probably conjures up the image of a schoolboy's pencil-case containing a few chosen items designed to help reading: a pair of glasses for better readability of the footnotes and formulas, a bookmark to help with our failing memory, a bag of pekoe tea to boost our attention when it declines, a LED flashlight to continue reading in the dark when away from electricity. The reading toolkit I have in mind, however, is filled with resources of the invisible kind: time, memory, energy, attention, and motivation. A skillful writer uses this toolkit to minimize the time, memory, and energy needed for reading while keeping reader attention and motivation high.

This page intentionally left blank

Chapter 1
Require Less from Memory

Let us start with a Vladimir story.

> ### Reading accident
>
> *Vladimir is reading a paper in the printed proceedings of the conference he attends each year. As it turns out, he is reading your paper! Suddenly, he stops on page three, places his index finger underneath a word, and rapidly scans the text he has just read, searching for something. What he is looking for is not on the page. With his free left hand, he flips back one page, and then another… He stops again. His face lights up. Satisfied, he returns to the page he was reading before the distracting reading U-turn, withdraws the finger still pointing to the problematic sentence, and resumes reading. What happened? A reading accident: the forgotten acronym. Who is responsible?*
>
> *Guess who!*

The Forgotten or Undefined Acronym

Acronyms allow writing to be concise, but conciseness is unhelpful if it decreases clarity. An acronym is always clear within the paragraph in which it is defined. If it is used regularly in the paragraphs that follow, the reader is able to keep its meaning in mind, but if it appears irregularly or if reading is frequently interrupted, the acronym — away from the warm nest of the reader's short-term memory — loses its meaning.

How could Vladimir have lost the meaning of the acronym? He had read its definition at the beginning of the paper, but had forgotten

it by the time he got to page three. Apart from simply forgetting, Vladimir could have missed the definition all together. Picture this.

You drew a brilliant figure that attracts the attention of Vladimir as he first browses through your paper. He is drawn to the caption of that figure and trips over an obscure acronym. Problem is, Vladimir has not yet read the introduction of the paper where the acronym is defined. Therefore, refrain from placing acronyms in captions and subheadings.

Sometimes, the acronym is not defined at all because the writer assumes that the reader is an expert in the field, and already familiar with the acronym. This omission is sure to upset the readers who are not experts because it forces them to search for the definition of the acronym outside of the paper.

Avoiding problems with acronyms is easy:

- ☀ If an acronym is used two or three times only in the entire paper, it is better not to use one at all (unless it is as well-known as IBM, or the paper happens to be extremely short, like a letter or an extended abstract).
- ☀ If an acronym is used more than two or three times, expand its letters the first time it appears on a page or under a new heading so that the reader never has to go far to find its definition.
- ☀ Avoid acronyms in visuals or redefine them when in a title, legend or caption because readers often look at the visuals of a paper before reading it.
- ☀ Avoid acronyms in headings and subheadings because readers often read the structure of a paper before reading the rest.
- ☀ Be conservative. Define all acronyms, except those commonly understood by the readers of the journal where your paper is published.
- ☀ Never leave an acronym undefined unless it is as familiar as U.S.A. And remember that your international reader may not be familiar with local, national, or foreign acronyms (French SIDA is English AIDS).

> ### The Singapore taxi driver
>
> *The other day, while Vladimir was in Singapore, He hailed a taxi to go to a research institute located on the campus of Nanyang Technological University (N.T.U). The taxi stopped. Vlad got in and said "Nanyang Technological University, please". The taxi driver, an old man who had clearly been living many years in Singapore, replied: "I do not know where that is". The answer surprised Vlad. The university is old and well established; surely the taxi driver had taken passengers there before. When Vladimir explained that it was at the end of the expressway towards Jurong, the old man's face suddenly brightened and he said with a large smile, "Ah! N.T.U! Why didn't you say so before!" That day, Vladimir learned that an acronym is sometimes better known than its definition.*

☀ If the acronym is well-known, introduce it before its definition. If it is not well-known, let its definition precede the abbreviated form.

The popular universal learning algorithm SVM (Support Vector Machine) had a profound impact on the world of classification.
The new universal learning algorithm — Support Vector Machine (SVM) — is likely to have a profound impact on the world of classification.

Identify the first appearance of each acronym in your paper. Search for its subsequent appearances. If the acronym reappears in a section of your paper different from the one where it was first defined, redefine it there. That way, the reader never has to go back further than the head of a section to find the meaning of any acronym. If the acronym appears in a heading/subheading, or in the caption of a figure, replace the acronym with its full definition.

The Detached Pronoun

This, *it*, *them*, *their*, and *they* are all pronouns. A pronoun replaces a noun, a phrase, a sentence, or even a full paragraph. Like the acronym, it acts as a shortcut to avoid repetition.

Pronouns and acronyms are both pointers. This characteristic is at the root of the problems they create:

- If you point to where someone was sitting an hour ago to refer to that person, not even the people who were there may remember who that person was. If the pronoun points to a noun already gone from the reader' short-term memory because it

is located 80 words back in the text, the noun-pronoun link is broken. This disconnection is insufficient to discourage readers. They read on, but with less clarity of thought.

- If during your speech you mention and point towards a researcher friend standing in a group of four or five people thirty meters away from you. Only the people who know that person will know whom you are referring to; no one else will. Prior knowledge disambiguates. When the pronoun points to several likely candidates, the non-expert reader — whose incomplete understanding of the text does not allow disambiguation — will pick the most likely candidate and read on, hoping to be right. If that likely candidate is the wrong one, then interpretation errors follow.

> *The fat cat ate a mouse. It had been without food for a week.*

What had been without food for a week, the cat or the mouse? The writer knows, of course, but the reader, left guessing, attempts to guess what '*it*' means. '*It*' may be the cat because the cat was hungry after starving for a week. But wait! A fat cat does not lack food, so the '*it*' may be a weak starving mouse unable to run fast enough to escape. But wait! A mouse with far less reserves than a cat could not survive such a drastic diet, and hence '*it*' must be the cat. The reader is confused.

- Finally some fingers seem to point to nobody, or they point to things still to come.

> *It was deemed appropriate to exclude women from the sample.*

This sentence starts with the infamous '*it*,' the passive voice companion. Who is '*it*'? The person who deemed '*it*' appropriate is hidden from the reader. Is '*it*' the writer? The supervisor? The research department manager?

And some fingers point to things or people still to come. If you point to where someone is going to be standing five minutes from now to refer to that person, the people will have to wait to know who that person is. This creates tension.

> *Our surprise speaker for this evening is a great archer. **He** has been ….*
> *… Ladies and gentlemen, let's welcome Professor William Tell.*

Apart from these situations where the substitute for the pronoun lives in the reader's future, it only takes milliseconds for the brain to choose a likely candidate for the pronoun in the short-term working memory. What influences the choice of candidate?

1) The reader's knowledge. The more superficial the knowledge, the more error-prone the choice will be.
2) The context. When there are multiple candidates, the reader uses context and logic — as in the cat example.
3) The distance between the candidate and the pronoun. Candidates that are many words away from the pronoun are unlikely to be chosen.
4) The grammar. A singular pronoun (this) or a plural pronoun (these, them, they) helps guide the choice, and so does gender (he or she, her, his).

In the following example, try to determine what the plural pronoun '*their*' refers to. The three candidates are in bold. Had the sentence been clear, this task would have been instantaneous. As it is, you will probably struggle; but if you do not, ask yourself how much your knowledge of the field has helped you make the correct choice.

> *The cellular automaton (CA) cell, a natural candidate to model the electrical activity of a cell, is an ideal component to use in the simulation of* **intercellular communications**, *such as those occurring between cardiac cells, and to model* **abnormal asynchronous propagations**, *such as* **ectopic beats**, *initiated and propagated cell-to-cell, regardless of the complexity of THEIR patterns.*

It is difficult to determine the plural noun pointed to by '*their*' because the sentence segment '*regardless of the complexity of their patterns*' could be moved around in the sentence, and still make sense.

> *… to use in the simulation of intercellular communications, regardless of the complexity of their patterns…*
> *…to model abnormal asynchronous propagations, regardless of the complexity of their patterns…*
> *…such as ectopic beats, regardless of the complexity of their patterns…*

Communications, propagations, and beats can all display complex patterns. Let us decide that in this text, '*their*' represents the '*abnormal asynchronous propagations.*'

The ambiguity can be removed in different ways. One could omit the detail. The long sentence would be seven words shorter.

The cellular automaton (CA) cell, a natural candidate to model the electrical activity of a cell, is an ideal component to use in the simulation of **intercellular communications**, *such as those occurring between cardiac cells, and to model* **abnormal asynchronous propagations**, *such as* **ectopic beats**, *initiated and propagated cell-to-cell.*

One could rewrite the part of the sentence that creates a problem to make the pronoun disappear.

The cellular automaton (CA) cell — a natural candidate to model the electrical activity of a cell — is an ideal component to use in the simulation of intercellular communications, such as those occurring between cardiac cells, and to model the cell-to-cell initiation and propagation of abnormal asynchronous events (such as ectopic beats) with or without complex patterns.

One could repeat the noun instead of using a pronoun.

The cellular automaton (CA) cell, a natural candidate to model the electrical activity of a cell, is an ideal component to use in the simulation of intercellular communications, such as those occurring between cardiac cells, and to model abnormal asynchronous events, such as ectopic beats, initiated and propagated cell-to-cell, however complex the propagation patterns may be.

However, these are all quick fixes. The best way to handle this situation is to return to the source of the problem, in this case, the extreme length of this sentence. A complete rewriting is necessary.

Cardiac cells communicate by initiating and propagating electric signals cell to cell. The signal propagation patterns are sometimes complex as in the case of abnormal asynchronous ectopic beats. To model such complexity, we used a Cellular Automaton cell (CA cell).

The length of the rewritten text is only about 70 % that of the original. The paragraph is clearer, shorter, and without the pronouns *'those'* and *'their'*.

Conduct a systematic search for each of the following pronouns in your paper: 'this,' 'it,' 'they,' 'their,' and 'them'.
    If you were the non-expert reader, could you easily identify what the pronoun refers to without ambiguity? If you could not, remove the pronoun and repeat the noun(s)/phrase it replaces. An alternate route consists in rewriting the whole sentence in a way that removes the need for the pronoun.

> Pattern matching
>
> *That day, Vladimir could not understand why the paragraph he was reading was so obscure. The usual culprits were absent: the grammar was correct and the sentence length was average for a scientific article. Then Vladimir roared as if he had aligned three gold bars in a slot machine. His superior pattern matching brain had aligned three synonymous expressions:*
>
> | | | |
> |---|---|---|
> | *(1) Predefined* | *location* | *information* |
> | *(2) Pre-programmed* | *location* | *information* |
> | *(3) Known* | *position* | *information* |
>
> *Each column contained either identical words, or synonyms. Vladimir decided to replace each of the expressions with 'known location'. When the fog created by the synonyms cleared, the structural problems of the paragraph appeared in full light.*

The Diverting Synonym

We all remember (not always fondly) the English teacher from primary school; let's say her name was Mrs Smith. She would read our ten line English essay which took us all weekend to write, and would scold us each time we dared repeat the same noun in successive sentences.

"The fat cat ate a mouse. The cat had not eaten a mouse for weeks."

She would pounce on us and mortify us for having committed the dreaded eighth capital sin: repetition.

"But Little Johnny, you should not repeat 'cat' or 'mouse'. You should use synonyms. For example, you could say 'The feline creature had not eaten a rodent for weeks'". And poor little Johnny, who did not even know what felines or rodents were, marveled at the deep knowledge of his English teacher. The message he carried into his life as a researcher was the very message that Mrs. Smith had drilled into him: do not repeat words but use synonyms to demonstrate your great culture through your vast vocabulary.

In scientific papers, however, synonyms confuse readers, particularly those not familiar with the specialized terms used in your field.

☀ Avoid synonyms. Make your writing clear by consistently using the same keywords, even if it means repeating them from one sentence to the next. As an added benefit, you will lessen the demands on the memory of your readers: fewer synonyms are fewer words to remember and understand.

Search for synonyms. They are usually found in successive sentences. Synonyms of title keywords are often found in abstracts, in headings and subheadings, and in figure captions. Select your set of keywords (particularly those used in the title) and reuse these keywords everywhere: title, abstract, introduction, headings and subheadings, legends, and captions.

The Distant Background

> **The Macintosh factory**
>
> *When Vladimir moved to Cupertino California, in 1986, to work as a summer intern at the headquarters of Apple Computer, he was taken on a tour to visit the Macintosh factory in nearby Fremont. Every day, truckloads of components and parts came in, just enough for one day's production; and every day, containers of Macintoshes were shipped out. The net result: no local storage, no warehousing. Vladimir was witnessing the very efficient technique of just-in-time (JIT) manufacturing.*

Ask anyone where background is located in a scientific paper, and the answer will consistently be '*the introduction*'. Well, yes… and no. If background material is of no immediate use to the reader, it rapidly fades out of memory. Our short-term active memory is far too short-term for such background to remain in for long.

> **The variable types**
>
> *There are two types of variables in a computer program: global and local variables. Why? To allow the program to manage computer memory space more efficiently. Global variables require permanent memory storage whereas local variables free up their temporary memory storage space as soon as the program exits the subroutine where they were used. Could this wonderful concept apply to writing?*

Parking *all* background material in the introductory sections of your paper greatly increases the demands on memory. Background material comes in two flavors: the global background, applicable to the whole paper; and the local or just-in-time background, useful only within a section or paragraph of your paper. The just-in-time background imposes no memory load: it immediately precedes or follows what it clarifies. Here is a just-in-time example.

> *Additional information is readily available from "context" — other words found near the word considered.*

In this example, the word "*context*" is defined as soon as it appears.

☀ When a heading or subheading in your paper contains a new word requiring explanation, explain it in the first sentence under the heading, in a just-in-time fashion.

Lysozyme solution preparation
Lysozyme, an enzyme contained in egg white, …

In this subheading from an article for a chemistry journal, the word '*lysozyme*' is expected to be new to the reader. The writer defines it just-in-time in the first sentence of the section, using what grammarians call "an apposition" — a phrase that clarifies what precedes it. Kept short, appositions are very effective. Long, they are ineffective as the following sentence demonstrates.

> *Lysozyme, a substance capable of dissolving certain bacteria, and present in egg white and saliva but also tears where it breaks down the cell wall of germs, is used without purification.*

Appositions are also ineffective when they slow down reading and lengthen a sentence as in the earlier example.

> *The cellular automaton (CA) cell, **a natural candidate to model the electrical activity of a cell**, is an ideal component to use in the simulation of intercellular communications, **such as those occurring between cardiac cells**, and to model the abnormal asynchronous propagations, **such as ectopic beats**, initiated and propagated cell-to-cell, regardless of the complexity of their patterns.*

Select each informative heading and subheading (i.e., not the standard methodology, results and discussion headings). Read the sentence that immediately follows that heading/subheading. It should provide the background for the heading. If not, modify that sentence.

The Broken Couple

> Wasted water, wasted thoughts
>
> *Vladimir stood still, hands under the hot water tap waiting for the water to become warm, wasting cold water down the sink. His problem-solving brain kicked into gear.*
>
> *"Why don't plumbers place the water heater close to the hot water tap, he wondered. Alternatively, I suppose one could use plastic pipes or insulate the heat-sinking metal pipes. Of course, I could always buy a thermostatic tap... with my next bonus check."*
>
> *His musings were interrupted by the arrival of the hot water and by the thought that he already had these thoughts before.*

When reading a sentence in which the verb never seems to arrive, has it occurred to you that your reader may also waste, or worse, be distracted by the words that separate the subject from its verb? Details inserted between the main components of a sentence **burden** (burden comes from the old French *bourdon*, a hum or buzz — but do we need to know that?) **the memory** because they move apart two words that the reader expects to see together, such as the verb (*burden*) and its object (*the memory*) in this sentence.

☀ Keep these happy couples close to each other.

• Verb and its object	• Subject and its verb	• Visual & its complete caption	• Background information and what requires background
• Unfamiliar word and its definition	• Acronym and its definition	• Noun/phrase and its pronoun	...

Separating the subject and the verb, as illustrated in ☛**1**, can be devastating.
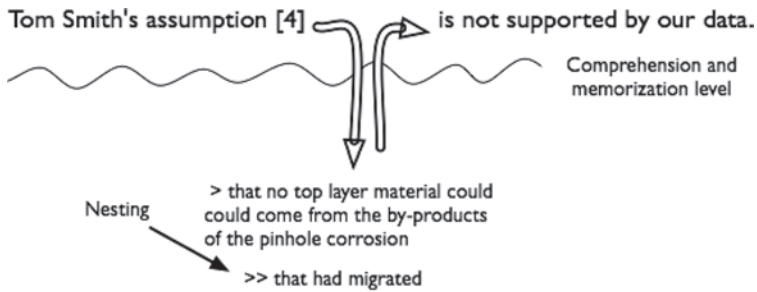
Figure ☛**1**
The nesting of phrases has the same effect as plunging the reader below comprehension level. In the end, what will be remembered is above the comprehension level, and what will be discarded as detail and forgotten is below the comprehension level. Two broken couples are responsible for this: (1) The subject ('Tom Smith's assumption') is separated from the verb ('is not supported') by two nested sentences starting with 'that'; and (2) the phrase 'the by-products' is separated from 'that had migrated' by the phrase 'of the pinhole corrosion,' creating some confusion. It is not the corrosion that migrates, it is the by-products; To avoid the confusing double nesting, the writer could have changed 'that had migrated' into the noun 'migration' as 'that no top layer material could come from the migration of the pinhole corrosion by-products.' The sentence would have been clearer.

> ### Memory registers
>
> *Curiosity pushed Vladimir to open up his antique TRS80 computer to look at the printed circuit board. He could not quite recall what CPU was in use back then. Close to the heat sink, he spotted the long CPU chip, the Intel 8085 microprocessor. He remembered studying its structure back in 1981. That was when he discovered for the first time that rapid access to memory is so critical to the overall speed of a microprocessor, that the central processing unit (CPU) has its own dedicated memory registers right on the chip, under the same roof so to speak. This collocation allows ultrafast storage and retrieval of data from these internal registers compared to the time it takes to store and retrieve data from external memory.*

☀ To increase reading performance through fast memory access, keep syntactically or semantically closely related items on the same page, in the same paragraph, in the same sentence, or on the same line. The reader will appreciate the increase in reading speed.

The Word Overflow

Memorizing takes a great deal of attention. The process is also slow. Have you ever been able to absorb complex road directions without asking the person to repeat them? Our working memory is not very elastic; it cannot accommodate a sudden word overflow as the next example demonstrates. Before you start reading it, a little background will help. This sentence is about molding machines. These machines melt plastic pellets and inject the molten plastic inside molds to make plastic parts.

> *"The main difference between the new micro molding machine design and the conventional « macro » molding machines with reciprocating screw injection system is that by separating melt plastication and melt injection, a small injection plunger a few millimeters in diameter can be used for melt injection to control metering accuracy, and at the same time a screw design that has sufficient channel depth to properly handle standard plastic pellets and yet provide required screw strength can be employed in micro molding machines."*[1]

Chances are you were not able to finish reading this 81-word sentence without having to read again parts of it. Its syntax is acceptable, and the meaning clear enough for a specialist familiar with the machine. But the working memory necessary to process it in one reading is too large for most readers.

When noticing such a long sentence, the writer is often tempted to break it down into two parts, by inserting a full stop in the middle of it.

> *"The main difference between the new micro molding machine design and the conventional « macro » molding machines with reciprocating screw injection system is that by separating melt plastication and melt injection, a small injection plunger a few millimeters in diameter can be used for melt injection* **to control metering accuracy. At the same time** *a screw design that has sufficient channel depth to properly handle standard plastic pellets and yet provide required screw strength can be employed in micro molding machines."*

Of course, it helps, but it rarely is the best way to deal with a long sentence.

---

[1] Zhao J, Mayes RH, Chen GE, Xie H, Chan PS. (2003) "Effects of process parameters on the micro molding process", *Polymer Eng Sci*, **43**(9): 1542–1554, ©Society of Plastics Engineers

☀ <mark>Before rewriting a long sentence to avoid what computer programmers call memory stack overflow and humans call headache, identify the intent of the author.</mark>

This sentence, found in the introduction of the paper, identifies a novel solution. The novelty is buried right in the middle of the long sentence in a place unlikely to attract much attention.

<mark>Before breaking down that sentence into multiple sentences, let's adopt a classic structure: start with the existing problem — then present a new solution.</mark> Unlike the original sentence that immediately plunges the reader straight into the core of the innovation, let us start with the known information — the background familiar to the reader.

> *In conventional « macro » molding machines with reciprocating screw injection, melt plastication and melt injection are combined within the screw-barrel system. In the new micro molding machine, screw and injector are separated. The screw, now redesigned, has enough channel depth and strength to handle standard plastic pellets, but the separate injector, now with a plunger only a few millimeters in diameter, enables better control of metering accuracy.*

The rewritten paragraph has three sentences instead of one, and 67 words instead of 81. The memory is not strained because the punctuation provides enough pauses for the brain to catch up: two commas and one full stop in the original versus five commas and three full stops in the rewritten version. The new version is also clearer and more concise.

Before we leave this example, and since good writing is also convincing writing, I would like to emphasize something the authors of the original sentence did well. They anticipated that the users of the new 'Micro-Molding' machines would want to use the same (cheap) plastic pellets that they use in their old 'Macro-Molding' machines. They may have tons of these pellets stored on wooden pallets in their warehouses. The authors responded to a possible objection by reassuring the users that the redesigned screw is able to handle standard plastic pellets.

☀ <mark>The best way to persuade someone to let go of something familiar and adopt something new, is to anticipate and respond to the</mark>

most common objection people might have when considering the change.

Choose a section of you paper (start with the introduction if you want). Select the sentences that look long. On the Mac version of Microsoft Word, command-click anywhere inside a sentence to select that sentence. Look at the number of words in that sentence (bottom line of the window). If that sentence exceeds 40 words, read it to see if it is clear. When unsure, revise for clarity: identify the purpose of the sentence and look for a classic structure to help you break it down into smaller sentences. Alternatively, you may want to remove the excessive details that make it long, or rewrite the whole paragraph.

In summary, acronyms, pronouns, synonyms, abusive detailing, background ghettos, cryptic captions, disconnected phrases and long sentences, all take their toll on the reader's memory.

Chapter 2

Sustain Attention to Ensure Continuous Reading

Attention-getters

Vladimir and his wife, Ruslana, usually read in bed before going to sleep.

"You haven't finished it yet?" Ruslana asked.

It was more a remark than a question. Her husband had been reading a 10-page scientific article for the past three nights, while in the same amount of time, she had read close to 250 pages of a suspenseful novel. Each evening, she had remained in bed; whereas Vladimir, unable to remain focused, had been in and out of bed, for a drink, a phone call, the late-night TV news, or a snack. She knew the signs. Tired after a long day at the lab, he did not have enough energy to stay attentive for more than 10 minutes at a time. His article required too much time and concentration. She asked Vladimir,

"Have you ever read a really interesting scientific article that you could not stop reading?"

He looked at her, and remained silent long enough for her to know that there could not have been many.

"I can't think of one!" he said finally, "Even my own papers bore me."

"Don't they teach you how to write interesting papers at your Institute, you know, papers that attract the attention of scientists?" Vladimir sighed.

"Attracting is fine. It is sustaining the attention that is hard. I wish I could keep my readers as awake and interested as you are."

She could not resist the tease.

"Oh, I am awake, darling, and interested. Are you?"

Love, drama, and suspense add spice to novels. But how can there be any suspense in a scientific paper when the writer reveals all secrets

right upfront in the title. Imagine Agatha Christie writing a novel with a title like "The butler killed the Duchess with a candlestick in the library". The book would surely make it to the top of the worst-seller list — at least, as far as suspense goes!

It would start with a summary that says "In this book, we will show that it was the butler who killed the Duchess with a candlestick in the library." The introductory chapter would mention that the chamberlain, the gardener, and the maid also live with the Duchess, but none of them have any motivation to kill her. The middle chapter would show the photo of the butler caught in the act of bashing the Duchess over the head with a candlestick. Additional chapters would include the signed testimony of the butler admitting to the murder. And the final chapter would reinforce the facts already established: the weight of the candlestick and the muscular strength of the butler are enough to kill the Duchess. It would end with a proposal for further research, such as the criminal use of candlesticks by Geishas on Sumo wrestlers. You have here a possible rendition of a scientific paper *à la* Agatha Christie, but not by Agatha Christie.

How would Agatha Christie make any scientific paper more interesting? Let me suggest that, for one thing, she would turn your contribution into more of a story. She would bring twists into your story plot, mention unexpected problems, which you, the scientist hero deploying inductive and deductive logic, would victoriously solve. Sometimes, she would have you pause to elaborate or clarify, instead of blissfully moving athletically through your complex story with total disregard towards the breathless non-expert reader. And even though she could not use suspense at the top-level of the story to hook the reader, she would recreate local suspense, whenever the opportunity arose. There would rarely be a dull moment. Indeed, having listened to some riveting Nobel Laureate lectures, I am now convinced that a scientific story need not be boring at all.

Keep the Story Moving Forward

The most challenging word of this heading is the word 'story'. Somehow, scientists do not see the active story as an appropriate

model for a scientific paper. Granted, not all parts of a scientific paper gain clarity and garner interest through the story style (the methodology section rarely does), but all parts can benefit from the techniques presented next.

☀ The great attention-getter, **change**, is what keeps a story moving forward.

Take a change in paragraph, for example, with each change, the story progresses, widens, narrows, or jumps. When ideas stop moving, a paragraph lengthens. Imagine a river. When does water stop moving?

Sometimes, the river widens to form a lake with no discernible current. Ideas stop moving when they stagnate or expand into paraphrases. Sometimes water gets caught in a whirlpool. The author traps the reader in a whirlpool of details before returning to the main idea. Sometimes the river has deep meanders. The author makes an unexpected U-turn to complete an already presented idea. Sometimes slow counter-currents form alongside the riverbank. The author, inverting subject and object positions in a sentence, slows down reading because of constant use of the passive voice — a proven story-killer.

The lake

When there is no purpose (because there is no current), a paragraph grows in length. Ideas are presented in no specific order, or in an order obvious only to the writer. They are sometimes repeated through long paraphrases. This seemingly chaotic and repetitious flow of ideas contributes to the never-ending paragraph that stretches over one entire column of a journal and depresses the reader even before it is read. The reader knows by experience that reading will be slow, and clarity will be low.

> *When ideas are not in motion, a paragraph grows in length. The additional length slows down reading and reduces overall conciseness. **With paraphrasing, the paragraph lengthens without actually moving the ideas forward since the sentences have the same meaning.** …*

The sentence in bold paraphrases what the first two sentences already cover.

The whirlpool

==Nested detailing (when details explain details) is another paragraph lengthening agent.== Nested detailing takes the reader away from the main intent of the paragraph. When the details stop, the reader is ejected back into the main topic stream. In the next paragraph on embryonic cell proliferation in culture dishes, the in-depth description of the culture dish (text in bold) distracts the reader (dish → coating → reason for coating). These details could have been described elsewhere, or simply removed.

> For the next three days, the thirty embryonic cells proliferate in the culture dish. **The dish, made of plastic, has its inner surface coated with mouse cells that through treatment have lost the ability to divide, but not their ability to provide nutrients. The reason for such a special coating is to provide an adhesive surface for the embryonic cells**. After proliferation, the embryonic cells are collected and put into new culture dishes, a process called 'replating'. After 180 such replatings, millions of normal and still undifferentiated embryonic cells are available. They are then frozen and stored.

The omega meander

The reader is distracted when the author makes an unexpected Ω turn to add detail to a point made several sentences earlier. The flow of thoughts is disrupted. In the following example, the sentence in bold should follow the first sentence to remove the meander and linearize the flow.

> After conducting microbiological studies on the cockroaches collected in the university dormitories, we found that their guts carried microbes such as staphylococcus and coliform bacteria dangerous when found outside of the intestinal tract. Since cockroaches regurgitate food, their vomitus contaminates their body. Therefore, the same microbes, plus molds and yeasts are found on the surface of their hairy legs, antennae, and wings. **It is not astonishing to find such microbes in their guts as they are also present in the human and animal feces on which cockroaches feed.**

The counter current

A normal sentence has current. It pulls reading forward from the old information upstream to the new information downstream, from the

sentence' subject to its verb and object. When such a natural order is unduly inverted (with the passive voice for example), the sentence commonly lengthens, and reading slows down.

The cropping process should preserve all critical points. Images of the same size should also be produced by the cropping.

Look at your long paragraphs and ask yourself, am I making a single point here? Can I make that point using fewer arguments, fewer words, or even a figure? Would making two paragraphs out of this one paragraph clarify things and keep ideas in motion? Do I have lakes, whirlpools, meanders, or counter currents?

Twist and Shout

… Shout

'Twist and Shout,' a song The Beatles made popular, is an attention grabbing heading that requires explanation. The twists, in our attention-grabbing context, are the twists in the story plot. The shouts are all things that call the reader to attention.
Shouting is quite easy!

☀ Turn up the visual volume; add subheadings and contributive visuals.

Conquer the mind of the reader with the firepower of visuals instead of the bullets of text riffles. Add subheadings under the standard Methodology, Results, or Discussion headings. Surrounded by vision-enhancing white space, subheadings shout in their bold font suits. Subheadings tell the reader how your story is moving forward. Therefore, make your subheadings as informative and indicative of content as possible. Avoid vague subheadings such as Simulation or Experiment, but also avoid excessively specific ones that contain keywords not even in your abstract.

☀ Bring changes to format and style.
Used in moderation,

1) A numbered list,
2) **Bold**

3) Underlined
4) *Italic text*
5) **a change in font**
6) A box around text

are equivalent to raising the volume of your voice, or changing its pitch or prosody. These changes break the monotony of paragraphs and make things stand out (note that the publisher may limit your choices by imposing a standard format and regulating the use of styles).

☀ Change sentence length, and ask questions.

After a long sentence, and particularly at the end of a paragraph, a short sentence carries much emphasis. It shouts. Why? It does not clog the memory, its syntax is fast to process, and its meaning is easier to understand. The last sentence of the following paragraph is four times shorter than the three sentences that precede it. That short sentence with short words does not shout, it screams!

> *Annotations on paper photos were all manual; they were either implicit (one photo album by event, location, or subject) or explicit (scribbles on the back of photos). In today's digital world, while some annotations are still manual, most, like time, date and sometimes GPS location, are automatically entered in the photo file by the camera. Is it conceivable that one day automatic annotations will be extended to include major life events, familiar scenery, or familiar faces, thus removing the need for manual annotations? Yes, and that day is upon us.*

The theme of the paragraph is photo annotation — yesterday, today and tomorrow. The passive voice used throughout the paragraph is just a writer's tool to achieve the writer's purpose: automation is inevitable. You may argue that it is possible to rewrite this paragraph using the active voice. I agree, and to show you that the active voice is not a panacea, healing all writing problems, here is the "active" paragraph. This example will also help us discover one more attention sustaining technique.

☀ Change sentence syntax and length.

> *Annotating paper photos was a manual task, implicit (photo albums by event, location, or subject) or explicit (scribbles on the back of photos). Annotating digital photos is an automatic task for digital cameras*

(time, date, and sometimes GPS location inserted in the photo's data file). Annotating automatically the previously manually entered data will tomorrow be possible for major life events, familiar scenery, or familiar faces.

All three sentences share a constant topic located upfront ('*annotating*'). All three sentences are written in the active voice. All three sentences have similar lengths. And all three sentences follow exactly the same syntax: subject, verb, and object. Parallelism in length and syntax over more than two sentences loses its magic. The reader is bored. A statement now replaces the suspenseful question. And even though the paragraph is more concise (65 words versus 88), it is now lifeless.

☀ Convey importance with words that act like pointing fingers.

They are quite effective when used sparingly (if everything stands out, nothing does). These words give you a way to reveal the salient facts amongst all the facts.

more importantly, significantly, notably, in particular, particularly, especially, even, nevertheless

☀ Keep the same contribution in front of the reader throughout your paper.

To make sure that the reader never loses sight of the paper's contribution, the writer scientist mentions it in every part of the paper: the title, the abstract, the introduction, the conclusion, the figures, the body of the paper, and even title-echoing words from the headings and subheadings. But if the writer cloaks the contribution with synonyms, if the writer seems to tell a different story in the title, abstract or conclusion, or if the writer digresses, then the reader may wonder what the contribution really is.

… Twist

There are always opportunities for tension in a scientific paper. It may be a limit you are about to circumvent, or an exception you are ready to exploit. It may be a common point of view you are about to change, or a gap you are about to bridge.

☀ Announce contrasted views or facts.

To bring you as close to drama as you'll ever be in a scientific paper, calls for special introductory words.

however, but, contrary to, although, in contrast, on the other hand, while, whereas, whilst, only, unable to, worse, the problem is that …

Things are not always as perfect as they look at first sight.

*"**Although** COBRA (Cost Based operator Rate Adaptation) has shown itself to be beneficial for timetabling problems, Tuson & Ross [266, 271] found it provided **only** equal or **worse** solution quality over a wide range of other test problems, compared with carefully chosen fixed operator probabilities."[1]*

Twists are tales of the unexpected. Too often scientific papers are tales of the expected. If papers are dull, it is because all difficulties encountered during the research have been erased from the public record. Only what works is presented.

☀ Keep enough of the unexpected difficulties to sustain interest and build in the reader's mind the image of a resourceful scientist.

interestingly, curiously, surprisingly, might have (but did not), unexpectedly, unforeseen, seemingly, unusual, different from, …

In the next example, the modal verb '*might have*' intrigues the reader. … *might have, but did not!* It sets the expectation that the writer will explain why the method is not as applicable as originally thought.

*The Global Induction Rule method [3], a natural language processing method, **might have** worked on news video segmentation since news contents can be expressed in a form similar to that used for text documents: word, phrase, and sentence.*

☀ Use numbers.

Too many numbers distract the reader and clog the memory, but a couple of well-chosen numbers grab the attention. Numbers have considerable attractive power.

---

[1] Reprinted with permission from Dr. Mark Sinclair, PH.D thesis, "Evolutionary Algorithms for optical network design: a genetic-algorithm/heuristic hybrid approach", ©2001.

After 60 years old, your brain loses 0.5% of its volume per year.

0.5% makes this sentence precise and dramatic (at least for the elderly), and *'millions'* attracts attention in the following example.

*After proliferation, the embryonic cells are collected and put into new culture dishes, a process called 'replating'. After **180** such replatings, **millions** of normal and still undifferentiated embryonic cells are available.*

In the title and the abstract, you do not have the artillery power of visuals to convince the reader that your paper is worth reading. The next best thing to help you convince the readers that they really need to download your paper and use your findings is… NUMBERS. Is your abstract precise or vague? Would numbers be appropriate in your title? Inside your paper, you have visuals. Some journals propose a ratio of one visual per thousand words. Calculate the ratio between the number of columns taken by visuals (include the visual text captions as part of the space occupied by visuals) and the number of columns taken by paragraph text (exclude the title, abstract and reference parts of your paper). Now look at the most highly cited papers mentioned in the references at the back of your paper, and calculate the same ratio (visual versus text). How does your ratio compare to their average ratio? Do you have enough visuals, i.e., are you sufficiently convincing and clear?

☀ Announce the alternative routes when you are about to change direction in your story plot.

rather than, instead of, alternatively

*"**Instead of** unidirectional motion along a single pathway, can we have unguided motion through the myriad of shapes?[2]"*

Pause to Illustrate and Clarify

After a particularly long or arduous section of any scientific paper, the reader's ability to absorb more knowledge is low. A sign that this is happening is the inability to recall much of what has been read (if it had been clear, memory recall would have been much easier).

[2] Reprinted excerpt with permission from Wolynes PG. (2001) Landscapes, Funnels, Glasses, and Folding: From Metaphor to Software. *Proceedings of the American Philosophical Society* **145**: 555–563.

Another sign is the inability to distinguish the important from the less important (dependencies, causalities, and relationships are not yet fully identified). These low points in knowledge acquisition happen time and time again while reading the paper. At these times, reader attention is easily lost when it ought to be sustained. Maybe it is caused by too much specialized vocabulary or dense theoretical considerations like formulas. The writer should identify such places in the paper, and plan for a pause to clarify and consolidate. To clarify, the writer uses illustrations, examples and visuals. To consolidate, the writer uses brief summaries. Clarifications and consolidations are highly appreciated by the reader, and temporarily boost the reader's attention level.

☀ Summarize (be it in text form or visual form) to clarify what is important.

A summary rephrases succinctly and differently the main points. It gives readers a second chance to understand, and it gives writers the assurance that readers will be able to keep in step with them. The following words perk up the attention of readers because, in them, they see the promise of knowledge consolidation.

To summarize, in summary, in other words, see Fig. X, in conclusion, in short, briefly put, …

Reading is hard, but writing is harder. Distilling years of research in less than 10 pages is a perilous exercise. Like compressed audio files, compressed knowledge loses clarity. Even if the structure of your paper is clear, you need to reintroduce detail into your text to make things clear, easier to grasp, less theoretical, more practical, and more visual.

☀ Use examples to illustrate.

The need for **examples** is not just a by-product of the distillation process. Illustrative details are needed because, more often than not, your readers are not familiar with what is happening in your field of research. They may be scientists in the same domain, but the distance between you and them in terms of knowledge is great, regardless of their academic level. What is tangible and real to you, may just be an idea, or a theory to them.

Your concern for making things clear to the reader is shown through words and punctuation. The words *for example, namely, such as, in particular, specifically,* and the colon, all keep the attention of the reader at a high level because they promise easier understanding, less generalities, and more clarifying details. They bring relief to struggling readers, and wipe out deep concentration furrows in their brow.

These words also reveal your expertise. Non-experts cannot give examples or specific details. Their comfort zone is in the general and the imprecise. Experts, on the other hand, are at ease in the specific, the precise, and the detailed. That is what makes them convincing, believable, and interesting!

Take your paper. Search for the words summarize, *summary, conclude*, and *in short.* If every time the search window returns empty, ask yourself: Are there sections in my paper where I can anticipate that the reader will be unable to see the tree for the forest, to identify what is really central in my argument or findings? If your answer is yes, then end these sections with a brief summary.

Recreate Local Suspense

The structure of a scientific article leaves little room for suspense. The gist of the contribution is revealed immediately in title and abstract, well before the reader reaches the conclusions. Therefore, suspense has to be artificially recreated.

Do you know what might reduce the accelerated shrinking of your brain and possibly slow down the onset of dementia in your old age?

☀ The master of suspense, the most underused and underrated tool in the writer's toolbox is… the **question** mark '?'

It is the quintessential universal trigger to guided thought processes. Questions make people curious. As soon as the writer expresses a question, it becomes the reader's question and remains so until the answer comes.

(a) A question refocuses the mind and prepares it for an answer.
(b) A question establishes the topic of a paragraph clearly.
(c) A question moves ideas in a given direction.

(d) A question lingers in the mind until it is answered – oh, by the way, the answer to the shrinking brain question is *Vitamin B*.

Does one always need a question mark to ask a question?

No. The question mark makes a question explicit, but there are ways of raising implicit questions. In grammar, the only question that comes without a question mark is **the indirect question**. It is asked by the author.

We wondered whether our data cleaning method was valid.

Unexpectedly, *might have*, *surprisingly*, and *curiously* announce **unexpected findings** that raised questions first in the mind of the author and later, by proxy, in the mind of the reader. In the next sentence the intriguing *would appear* raises a question. Manual polishing only appears to be the answer, but clearly, the writer has a better, albeit less obvious answer.

*What method provides enough contact force to polish the highly complex surfaces of large ship propeller blades? Manual polishing with a belt machine **would appear** to be the obvious answer.*

The **not-yet-justified adjectival claim** always raises a question, particularly when it is emphasized as in the following sentence:

*"The energy landscape/funnel metaphor leads to a **very different** picture of the folding process than the pathway metaphor.[3]*

In this case, the question is "What does the picture of the folding process look like when we use the energy landscape/funnel metaphor?" The author prompts the reader to ask that question by making the claim before bringing in the evidence. What a marvelous way to recreate suspense!

*At high temperatures, electron trapping is **not** …*
Electron trapping is unimportant when …

Acting in a manner similar to the adjectival claim, is the **negative statement** that presents what does not work before what works, what is unimportant before what is important. Starting a statement with what doesn't, begs the question of what does!

---

[3] Reprinted excerpt with permission from Wolynes PG. (2001) Landscapes, Funnels, Glasses, and Folding: From Metaphor to Software. *Proceedings of the American Philosophical Society* **145**: 555–563.

The **announcement of change** increases tension and attention.

The new strain of Malaria had become resistant to all existing medicine, but this was about to change.

The **provocative statement** is a trumpet call for justification or clarification. The following statement is as provocative as the title of the book of Thomas Friedman: "The World Is Flat."

HTML 5 is stillborn.

Certain **values in a visual** raise strong questions. Yes indeed, what happened on Thursday?
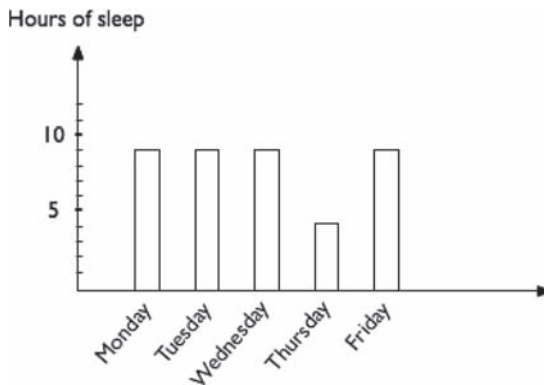
Antagonistic claims create natural suspense. Observe how the author sustains the interest. In five consecutive sentences, mostly written in the active voice, he brings (1) two numbers, (2) a figure, (3) the attention-getters *'but'*, *'not'*, and *'important'*, (4) an antagonistic claim *'whereas'*, *'contradiction'*, and (5) a question suggesting one of the reasons for the observed difference in results.

When trying to improve the clusters obtained by the Clusdex method, Strunfbach et al. reported a 27% increase in error rate when using our annealing method [6], whereas we had observed, not an increase, but a 12% decrease in error as seen Fig. 3. Their findings represented an important contradiction. After examining their data, we discovered that they had used the raw data while we had removed the outliers. Should outliers be kept or removed?

Roadblock. Different methods, different results, and no way to compare them create an impasse.

"As was pointed out (3), it is a challenging task to compare the results of these profiling studies because they used different microarray platforms that were only partially overlapping in gene composition. Notably, the Affymetrix arrays lacked many of the genes on the lymphochip microarrays …"[4]

Search your entire paper for the question mark. If you do not have any, ask yourself why? What are the essential questions you are raising and answering in your paper? Shouldn't you find a way to express them in question form? Do you recreate suspense and raise questions by other means such as roadblocks, provocation, visual questions, negative or adjectival claims?

At the beginning of this chapter, we identified that the structure of a scientific article leaves little room for suspense. The gist of the contribution is revealed immediately in title and abstract, well before the reader reaches the conclusions. Therefore, you, the author, have to reintroduce tension and suspense to sustain interest. Yes, it may feel unnatural; it may also add length to your paper. But the benefits far outweigh the costs. Because attention-getting schemes tend to increase contrast, clarity also increases. More attentive, the reader gets more out of your paper.

In this chapter, we also examined many ways of sustaining the initial attention of motivated readers as they plow through the tough parts of your paper where their attention drains at a fast pace. We used words, punctuation, syntax, text style, page layout, structure, examples, summaries, visuals, and questions. Since the attention of your reader is not deserved but earned through hard work, it is time for you to practice. It is time to wake up the comatose reader with adrenaline-loaded questions. It is also time to pour life-restoring liquid details onto desiccated knowledge.

Read your paper to identify the parts the reader might find hard to understand (give it to read to somebody else if you are unsure). In these parts, modify your text accordingly to increase attention and facilitate understanding (examples, visual aids, questions, new subheadings, etc.).

---

[4] Wright G, Tan B, Rosenwald A, Hurt E, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B Cell Lymphoma. *Proceedings of the National Academy of Sciences*, **100**: 9991–9996.

Disclaimer: *Attention getters are only effective if used sparingly. The author of this book will not be responsible if excessive use of attention-getters in a paper distracts readers and makes them lose focus. Examples of such excessive use include writing 'importantly' seven times in a long section, turning a paper into a cartoon through over-abundant visuals, using bold and italic in so many places that the paper looks like a primary school paper, or starting three consecutive sentences with 'however'.*

This page intentionally left blank

Chapter 3
Reduce Reading Time

Reading time is reduced when the writer is concise and to the point. Conciseness is seen by some authors as a mark of respect for the reader, as Pascal's quote illustrates.

> Blaise pascal
>
> *The 17th century French mathematician and philosopher Blaise Pascal apologizes to his reader in lettres provinciales. He writes: "This letter is longer, only because I have not had the leisure of making it shorter." Conciseness seen as politeness! Boileau, a French writer from the same century, has harsh words: "Who knows not how to set limits for himself, never knew how to write."*

Although time is measured in seconds, reading time is much more subjective. Readers experience the passage of time differently for different reasons: familiarity with the topic, linguistic skills, reading motivation, reading habits, or the reader's tiredness. Using techniques, the writer is able to reduce both objective and subjective time while increasing overall reading pleasure.

Readers are less sensitive to time when they look at visuals. Somehow, the brain is more engaged looking at a visual than reading paragraphs of text. Good visuals are information burgers devoured like fast food. The eyes rapidly scan them, detecting patterns as they go. Their message reaches the brain faster because processing them takes place in the highly developed visual cortex designed to handle **high information bandwidth**. For text or speech, the brain sequentially processes one word or syllable at a time — a low information bandwidth process. Should you need convincing, conduct the following experiment. Which of the following two representations

gives you the largest amount of information in the shortest amount of time: the following text

> *In our experiment, for an up flow velocity of 0.10 m/h, the observed normalized tracer concentration of the effluent increased rapidly from 0 to 0.4 after 15 hours. The increase slowed after 38 hours when the concentration reached 0.95. It peaked at 1.0 at 90 hours. Following which, the concentration curve decreased steeply down to reach zero asymptotically at 180 hours. The calculated data and the observed data were closely related. However, when compared with the calculated data, the observed data seemed to lag when the concentration dropped.*

or figure ☛1?

When reading is slow, readers become impatient. 🔖 examining what makes reading slow is essential if the writer wants to give the reader a pleasurable reading experience.

Reading is slow when the writer creates reading accidents, such as undefined acronyms, ambiguous pronouns, long sentences, missing logical steps, or unfilled knowledge gaps.

Reading is slow when the sentence is too complex or abstract. Pascal wrote that *"Memory is essential for the operations of reason".*
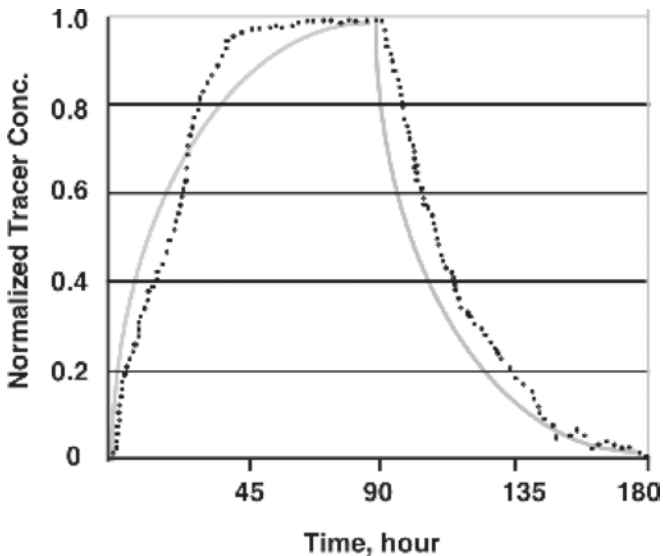
Figure ☛1
Tracer concentrations (dotted line) for an effluent velocity of 0.10 m/h. Calculated data (solid line) lag experimental data when concentration drops.

Complexity makes great demands on anyone's memory. It considerably slows down reading.

Reading is slow when the structure is insufficient. Readers spend more time when they cannot easily find what is of interest to them for lack of a clear structure with enough headings and subheadings.

☀ Make sure you have a detailed, informative structure (more on that later).

Structural elements are quick to read, short and simple in syntax, easily understood by the reader, and easy to locate and identify. For example, the following bold font subheading receives great attention because the indentation creates a white space acting as an eye-trap.

Reading is slow when the reader is not as familiar with the topic as the expert is. For the people new to your field, paradoxically, a longer introduction reduces the overall time required to read your paper simply because it sets the foundations needed to understand the rest of the paper. Veteran writers, after writing their paper to achieve maximum clarity and conciseness to satisfy experts, spend more time on their final drafts to cater for the needs of the non-experts. Here is what UC Davis Professor Dawn Sumner has to say about her writing process:

> *"I leave [my latest draft] for a while, and then I think about what my audience already knows, and I give them all the information that they need [so that] the flow follows for someone not as familiar with the topic as I am, and then [I] sort of reiterate that."*

Let me argue that anyone's writing will never be excellent until, like Professor Sumner, the writer keeps the reader in mind. Thinking about what the reader may not know allows you to identify the knowledge gaps that will slow down reading, because the reader has to work hard to understand how your sentences are logically linked.

☀ Sentences that fill the knowledge gaps, and explanations of gap-creating keywords always speed up reading.

Reading is slow when extra processing time is required because of the complexity of sentence syntax. For example, readers, experts or non-experts, struggle with long compound nouns. These require extra brain cycles to decode, which slows down reading.

☀ Unpack long compound nouns and clarify them by adding a preposition (of, on, to, with) to speed up reading.

Compare '*electric-stimulus* responsive *hydrogels*' with '*hydrogels* responsive **to** *electric stimulus*'. It is immediately clear which of the two is faster to read.

Reading is slow when writing lacks conciseness. As Pascal points out, a lengthy paper takes less time to write than a short one, but it takes more time to read. Identifying the sources of excess length at a global level is the first step towards conciseness, but it is as difficult as determining the causes of a bulging stomach. The need for the diet is clear, but the fat can come from so many sources. Where is the fat in your writing?

- Length is caused by the thousand words that should have been a diagram, a graph, or a table.
- Length grows out of a structure still in the formative stage, where information is needlessly repeated in different sections.
- Length is born out of the slowness of the mind, as it warms up and spreads a fog of platitudes, particularly in the first paragraph following a heading or subheading.
- Length is the fruit of unrealistic writer ambitions, aiming at cramming in a single paper the contribution of several papers.
- Length is the fruit of hurriedness, since it takes time to revise a paper for conciseness.
- Length happens when the reader is given details of the unnecessary kind, details that do not enhance the contribution.

Can your writing be too concise? Yes, it can, and your text would then lose clarity. Here are four good reasons to justify lengthening a paper.

- Lengthen to write a longer introduction that really sets the context and highlights the value of your contribution. Your contribution is like a diamond. To hold and display it, you need a ring (the ring of related works).
- Lengthen to highlight aspects of your contribution in every section of the paper (different angle and level of detail). Each facet of a diamond contributes to its sparkle. Likewise, each part of a paper presents the same contribution at a different angle.
- Lengthen to go beyond stating results in the abstract, and reveal the potential impact of your contribution on science. Would you give an uncut diamond and ask readers to polish it themselves?
- Lengthen to provide the level of detail that enables research colleagues to independently assess your results or at least follow your logic.

Read your paper. Are you repeating details? If you are, revise the structure to avoid repetition. Do you feel that readers of the journal in which you publish your paper already know what the first paragraph of your introduction says? If you do, cut it out. Is the last paragraph of your introduction a table of content for the rest of the paper? If it is, cut it out. Are you bored reading your own prose? If you are, it is time to replace it with a visual. Are all details essential to your contribution? Read the whole paper again and cut ruthlessly the details that explain details.

This page intentionally left blank

Chapter 4
Keep the Reader Motivated

Attention is precious. One should not waste it. It directs brain activity. Attention is a thought traffic controller. If attention wanes, our train of thoughts could derail or be redirected away from reading. Yet, for all its importance, attention is governed by a powerful ruler: motivation. Consider reading as a system with inputs and outputs, as shown in ☛**1**.

Motivation is one of the four critical inputs to the system. The reader's initial need or interest feeds it. The fun of gaining knowledge (feedback loop) keeps motivation high or even increases it for example when expectations are exceeded or when goals are met quickly. Motivation decreases when expectations are not met (syntax is too obscure, initial knowledge is insufficient, or paper is disappointing), when alternatives to reading become more attractive, or when the reader is tired.

> ### Wrong title and unmet expectations
>
> *Vladimir is a young US researcher new to the field of English speech recognition. When Popov, his supervisor, sent him two months ago to the ICASSP conference in Paris to catch up with the latest happenings in the field, he could not have foreseen that this trip to France was about to put Vladimir in some sort of trouble.*
>
> *Today, young Vladimir is searching online journal databases for general articles on automated speech recognition in over-the-phone plane reservation systems using dialogues. Fifth in a long list of titles, he spots the title: "Over-the-phone dialogue systems for travel information access." He smiles. All the keywords he typed are there. He orders the paper through Joan, the research center's librarian.*
>
> *A day later, the paper lands on his desk with a yellow post-it note attached to the first page that says, "French girl friend you met in romantic Paris?"*

> *Vladimir does not understand. He is happily married to Ruslana, a Russian. Puzzled, he removes the post-it that covers the abstract and starts reading. The abstract is at odds with the title. He had hoped for a general article, but he finds that it is about French speech in dialogue systems. His eyes move to the name of the first author: Michelle Mabel. A French woman! Darn! No wonder the librarian is teasing him. Why else would he be reading an article so foreign to his research field? Should he start reading? Or should he worry about the rumor that is probably going around the lab about a torrid extra-marital affair with a French woman? … (To be continued)*

Dash or Fuel the Hopes of Your Readers, Your Choice

Motivating the reader starts with the title of your paper. It provides the initial reading impetus. The reader will scan hundreds of titles and select only a few. Imagine for a moment that the reader found your title interesting. You have what all authors dream of: the reader's attention. So it is now up to you. Are you going to dash your reader's hopes, or on the contrary, fuel them?

Dash — by a title that is not representative of the rest of the article
Fuel — by a title that is representative of the rest of the article

> *Vladimir decides to read it anyway. After all, the paper is only five pages long. He should be able to get through it fast. He will just skip the parts that do not interest him.*
>
> *Half an hour later, he is only in the middle of page 2 reaching the end of the introduction, gasping for a graphic or a diagram to make things clearer. He glances anxiously at his watch. He has a meeting coming up with his team in 20 minutes. (To be continued)*

Dash — by making clear that reading the article will require more time than anticipated
Fuel — by making clear that reading the article will require less time than anticipated

THE READING SYSTEM
a black box perspective

Knowledge →

Motivation/goals →

Time →

Scientific paper →

BRAIN
Reading box

Knowledge
Enhanced
or Modified

Emotions →

Notes →

Great visual Processing speed
(figures & tables)

Boosting Human Factor

Limited eyesight
Limited memory
Limited knowledge
Slow reading speed for text
Limited attention span (fatigue)
Impatience (time-to-result driven)

Limiting Human Factors

Light
Noise
Comfort
Interruptions

Limiting Environmental factors

Figure ☛ 1

Reading, considered as a system, has four inputs and three main outputs. Prior to reading, each input has an initial value. That value will change over time because the outputs influence the inputs. For example, the more knowledge you get from the paper, the more knowledge you put back to facilitate further understanding. External factors influence the reading process. They either lubricate the process or create friction and inefficiencies. They indirectly affect the pace of absorption of knowledge, and therefore motivation, a critical input to the system. If reading was a transistor, motivation would be its base current that either shuts down or promotes the reading activity.

The introduction mentions French phonetics, and the differences in accents between the Chtimi and the Marseillais dialects. His knowledge of France is limited to football players and perfumes. He has heard of Zinedine Zidane, and Chanel #5 (he bought it in Paris for Ruslana's birthday), but that's about it. He glances at the reference section. No help there. He looks up Wikipedia online. No help there either. If he asks Joan, the librarian, she will ask him about his French girl friend, so he scraps that idea. By now, his motivation is at its lowest point, so he skips a few paragraphs and jumps to the part about dialogue modeling. (To be continued)

Dash — by not giving the reader the baseline knowledge required to read the paper

Fuel — by providing the reader with the baseline knowledge needed to read the paper

However, he did not know that his motivation could drop lower than its previous low level. The key paragraph that seemed to be precisely in his area of interest is totally obscure. He spends a good 5 minutes on it, and then gives up. (To be continued)

Dash — by using prose so obscure or complex in syntax that the reader gets discouraged and becomes unsure whether he or she understands correctly

Fuel — by using prose so clear that the reader is encouraged and sure that he or she understands correctly

He finally decides that the article is too specific. The semantic modeling will not apply to English at all. He won't be able to use it. And his meeting starts in a few minutes. A colleague who is also going to the meeting appears in his cubicle, looks quickly at one of the subheadings of the paper and says, "Hey, Vlad! I didn't know you were interested in French. Got a French girl friend? Does Ruslana know?"… Vladimir drops the paper in the trashcan. "It's a mistake," he says. (To be continued)

Dash — by making the reader doubt the validity, or applicability of the contribution

Fuel — by demonstrating the quality, validity, or applicability of the contribution

He rushes to his meeting. As he enters the room, all his colleagues shout what sounds like 'Bonjour'. Before he answers, he looks around and relaxes; his boss has not yet arrived. He says, "It is not what you think. The title was misleading." They all laugh. At that moment, his boss enters the meeting room and hands Vladimir a paper: "Here, read this," he says. "I have not read it, but it seems perfect for your research. It's by… um, a French author." The whole group collapses in laughter.

Dash — by boring the reader with a style lacking dynamism, a sentence structure lacking variety, new information lacking emphasis, and text lacking illustrations

Fuel — by captivating the reader with a dynamic style, a varied sentence structure, new information properly emphasized, and text rich in stories and illustrations

I deliberately chose a story to illustrate the various disappointments researchers experience while reading certain papers. How much more dynamic is the language of stories compared to the stiff classic scientific writing style. Deviating from the norm is often frowned upon (like ending sentences with a preposition). Yet, one section of your paper is ideally suited to accommodate such deviations: the introduction. You have a story to tell: the story of why you embarked on your research, why you chose a particular method, etc (see chapter on the introduction). Since it is a story, use the narrative story style to write it.

You can now see that your writing sustains the motivation of the reader through a combination of writing style, honest title, judicious detail and background, clear contribution, and good English.

Meet the Goals of Your Readers to Motivate Them

In our story, Vladimir (a newcomer to the field) is interested in general background. There are many kinds of readers, all coming to your paper with different motives and different levels of expertise. Satisfying and motivating them all is an impossible exercise if you do not really understand what readers hope to find in your paper. The following scenarios will help you understand their goals.

The field intelligence gatherer

Hi! I am a scientist working in the same field. I may not be doing exactly the same research, but I am a regular reader of the journal you read and attend the conferences you attend. I was the guy sitting on the fifth row facing you when you presented your paper in Korea last year. I have read most of your abstracts to keep up to date with what you're doing.

First among the six reader profiles examined is the intelligence gatherer. Such scientists are interested in anything in abridged form: your abstract or conclusions, sometimes the introduction. They probably will not read your whole paper.

The competitor

Hi! You know me, and I know you. Although we have never met face to face, we reference each other in our papers. By the way, thanks for the citation. I am trying to find a niche where you are not playing, or maybe I'll fix some of your problems in my next paper. Hey, who knows, maybe you are onto something I could benefit from. I'd love to chat or work with you on a common paper one of these days. Interested?

Even if some of your background is missing, competitors are able to fill in the blanks without your help. They will read your paper rapidly, starting with the reference section to see if their name is in it and if their own reading is up-to-date. They may also use your list of references to complete their own list. They will probably skip your introduction. Occasionally, competitors may be asked to review your article before publication.

The seeker of a problem to solve

Hi! You don't know me. I am a senior researcher. I just completed a major project, and I am looking for something new to do. I am not quite familiar with your field, but it looks interesting, and it seems as though I could apply some of my skills and methods to your problems and get better results than you. I am reading your paper to find out.

Problem seekers may read the discussion, conclusions, and future work sections of your paper. Since their knowledge is not extensive, they will also read the introduction to bridge their knowledge gap.

The *solution seeker*

Help! I'm stuck. My results are average. I am pressured to find a better solution. I need to look at other ways of solving my problem. I started looking outside my own technology field to see if I could get fresh ideas and methods. I'm not too familiar with what you're doing, but as I was browsing my list of titles, I discovered that you are working in the same application domain as I am.

Solution seekers will read the method section, the theoretical section, and anything else that can help them. They could be surgeons looking for artery modeling software, or AIDS researchers who heard that small-world networks have interesting applications in their field. Their knowledge gap may be very large. They are looking for general articles or even specific articles, which they will read in part, expecting to find a clear but substantial introduction with many references to further their education.

The *young researcher*

Hello! I'm fresh out of university, and quite new to this field. Your paper looks like a review paper. That's exactly what I need right now. Nothing too complicated. Just enough for me to understand the field, its problems, and the solutions advocated by researchers. That will do just fine!

Young researchers will read the introduction and (perhaps) follow your trail of references. They do not expect to be able to make sense of everything the first time, but what little they can understand, they will be happy with. Their knowledge gap is great.

The *serendipitous reader*

Hi! Cute title you've got there. I had to read your paper. Such a title could only come from an interesting writer. I thought I would learn a few things, a paradigm shift maybe. I'm not sure that I will understand any of it, but it's worth a try. Last time I did that, I learned quite a lot. The paper had won the Best Paper Award in an IEEE competition. I studied the paper. Although I did not understand much, I got quite a few hints on how to improve my scientific writing skills!

My point is this: researchers will come to your paper with different motivations and needs. A common mistake is to imagine the reader as another you, the competitor in this story or someone who

knows your topic as well as yourself. As the author, you would be wise not to rush through the introduction and the list of references. You would also be wise to provide enough detail so that other researchers can check and validate your work: 'little validation, little value'.

Ask people to read your paper. Ask for their opinion. Is it written for experts like yourself or will researchers new to the field be able to benefit from your paper? Are they motivated to read the rest of your paper after reading your introduction? Ask readers to circle the parts of your paper they found difficult to understand. Don't ask them to explain why, though. They probably would not give you a good answer. But then examine each of these circled parts and revise them drastically.

Chapter 5
Bridge the Knowledge Gap

> Apple computer
>
> *In the past, the computer Gods in the mainframe world, could only be approached by the grand priests in white coats serving them. When the personal computer arrived, the Gods did not quite fall off their pedestals. They just moved from the computer room temple to the living room shrine. The computers had tamed their owners. Occasionally, through what looked like sorcery to humans, some PC owners managed to tame their computers. Application programs responded with docility to the secret incantations they typed. That knowledge, known only to them, had corrupted their virgin brains. They no longer remembered what it was like not to know. Then came the Macintosh, to give hope again to the rest of the human race. History tells us the Mac dispelled the Orwellian vision of 1984, but it did not quite manage to allay the fears of the uninitiated although the iPod and the iPad did.*

Crossing the knowledge chasm between you and your reader is not easy. The reader knows less than you, but how much less?

- It depends on you. If the new knowledge you are contributing is significant, the knowledge gap between you and your readers is large.
- It depends on your readers. If their own knowledge of your field is small, they may not be familiar with the vocabulary or methods used. As a result, their initial knowledge gap is large, even if the additional knowledge you bring is modest.

You need to evaluate the gap to make sure your paper reaches the readers described in the previous chapter: the field intelligence gatherer, the competitor, the problem or solution seeker, and the young researcher. Of course, you could assume that the reader is

knowledgeable enough to follow your paper, but is this assumption valid? What do you, the writer, ***know for sure*** about your readers' initial knowledge?

You know that your readers found interesting one or several keywords in your title. You know that your readers have *enough* knowledge to tackle and explore *parts* of your paper. You know that your readers read the journal in which your paper is published or attend the conference where your paper is presented. You know that their work is related to the domain covered by the journal or conference. For example, the readers who attend the International Symposium on Industrial Crystallization are in chemical engineering. They know the tools and techniques used in the domain. They know the meaning of centrifugation, phase separation, concentration, calorimeter, and polarized light microscope. They know the principles of science, how to conduct experiments, and how to read a concentration and temperature chart. They know English — the Queen's English, the President's English, or some flavor of broken English.

Now that we have ascertained what you know for sure about your readers' initial knowledge, what then do you ***not know for sure***?

The answer is short and simple: **everything else**. Indeed, everything else cannot be assumed to be known. Even though it is tempting to believe that readers have the same level of knowledge as the one you had at the start of your project, nothing could be further from the truth. Readers are not younger versions of you.

Now that we have looked at their initial knowledge, let us consider your contribution. How much do they know about it? Nothing! Your contribution is unknown to them, **just as it was unknown to you before you started the research that led to your paper**.

Let us suppose that the title of your article is the following:

"Phase transitions in lysozyme solutions characterized by differential scanning calorimetry."[1]

Some readers may be more familiar with characterization techniques than they are with lysozymes. Therefore, they do not know which

---

[1] Lu J, Chow PS, Carpenter K. (2003). Phase transitions in lysozyme solutions characterized by differential scanning calorimetry. *Progress in Crystal Growth and Characterization of Materials* **46**: 105–129.

data, method, or experiment best applies to lysozyme, nor would they know what others before you have done in this domain or what specific problems remain unsolved. That is precisely what they will discover while reading your article.

Bridge to Ground Zero

I hope you now see that, on the whole, the gap between your elevated knowledge and your readers' basic knowledge (ground zero) is wide.

Since it is impossible to guess how wide the gap is, you will have to set the minimum knowledge required, the reasonable ground zero on which you will build new knowledge. It would be unreasonable to write your paper for college students or for scientists who are not regular readers of the journal your paper targets.

To put it in an easy to remember formula:

Reader Knowledge Gap = The new knowledge you acquired during your project + The new basic knowledge the reader requires to reach your starting point.

Ground zero will be conditioned by the number of pages given to you by the journal. The more pages you have, the more you can lower ground zero or increase the size of your contribution. For short papers for example, you can settle for a higher ground zero or a smaller contribution, and you can use the short reference section of your paper as a knowledge bridge. References are convenient shortcuts that tell the non-expert reader *"I'm not going to explain hidden Markov models. Indeed, I'm going to use the acronym HMM when I refer to them. This should be common knowledge to anyone working in the domain of speech recognition. Go and read reference [6] that represents the classical work on the subject if you need more basic information."*

Assuming the editor gives you enough pages, you may decide to lower ground zero by providing extra background knowledge instead of asking readers to get up to speed by themselves. Writers often provide this knowledge in **a background knowledge section** that immediately follows the introduction. This section is a great place to summarize what readers would have learned had they had the time to read the articles you mention in your reference section

(you can safely assume they will not read your [1], [2], and [3] before reading the rest of your paper). This background knowledge section is not part of your contribution, but it is necessary to understand it.

Ground zero is set by the latest books or review articles written by domain experts. If such articles are not available, then a look at the latest conference proceedings in your domain area should give you a couple of general articles that will accomplish the same function. Ground zero keeps moving up. Science is built on science, and scientists are expected to keep abreast of what is happening in their domain.

Ground zero is set by the type of journal. Some journals are multidisciplinary. The journal of bioinformatics, for example, is read by two distinct types or readers with widely different backgrounds: computer scientists and the biologists. Some journals have such a broad audience (Nature for example) that the whole paper has to be written for non-experts. Just-in-time background is extensive in these journals.

Bridge to Title Words

Readers have chosen to read your paper because certain keywords in its title attracted their attention. This very fact is of utmost importance. Because of it, and as incredible as it sounds, the knowledge gap of the reader is predictable. Each specific keyword in your title attracts two types of readers: the readers who are very familiar with the keyword and do not require any background, and the readers who would like to know more about that keyword and who indeed require background. Identify these keywords, and you have an idea of the background you need to cover in your paper. It is that simple.

Let us use a sample title to identify its potential readers. But before we start, some background might be helpful. To manufacture a non-stick frying pan, you need to deposit a hard coating on top of the metal pan. That coating is hydrophobic, which simply means water-repellent. The process to deposit the coating is called the sol-gel process. The *sol* is a chemical solution serving as a precursor which, after it enters into reaction with the *gel* made of polymers, will form the hydrophobic hard coating.

"Hydrophobic property of sol-gel hard coatings*"

What are the keywords?

1) Sol-gel: a subset of technologies for making coatings
2) Hard coatings: a subset of surface coatings
3) Hydrophobic property: a subset of the properties of surface coatings such as hardness, etc...

The knowledge field presented in this paper is established by these three keywords. Only experts among the readers have enough knowledge in all three. It should come as no surprise to find among the readers people with knowledge lacking in one or more of these keywords. And it should also be expected that, given the name of the conference (Technological Advances of Thin Films and Surface Coatings), all participants are somewhat familiar with the main knowledge sets: coatings, coating technologies, or coating properties.

Now ask yourself what these readers would like to know, and which part of your paper is going to provide them with an answer.

Hydrophobic property: What in the sol-gel process influences the hydrophobicity of hard coating? How is hydrophobicity related to other properties of the coating such as surface structure, transparency, or mechanical properties? How is hydrophobicity measured and quantified?

Hard coatings: How are hard coatings made by sol-gel technology? How does hydrophobicity affect or confer the hardness of the coating? How hard is the coating made of sol-gel material?

Sol-Gel: What are the sol-gel precursor materials used? How are the sol-gel coatings solutions prepared? What are the coating processes used?

Here is what should NOT be covered in the background because it falls outside the scope of your paper: hydrophobicity of soft coatings, other properties of materials independent of hydrophobicity, or hard coatings produced by other processes than sol-gel.

---

* The paper was presented by its author, Dr. Linda Wu, at the 2nd International Conference on Technological Advances of Thin Films and Surface Coatings (Thin Films 2004) in Singapore.

In this exercise, the title allowed you to identify five overlapping reader profiles, with different background needs:

1) People interested in sol-gel
2) People interested in hard coatings but only vaguely familiar with sol-gel
3) People interested in the hydrophobic properties of any coating material hard or soft
4) People interested in the mechanical properties of sol-gel hard coatings eager to know if the mechanical properties are compromised by the hydrophobic properties of the coating
5) People interested in the hydrophobic properties of sol-gel hard coatings (your contribution).

Now that you understand the principle, conduct the same exercise on the title of your own paper.

☀ Before you write your introduction, use the title keywords to identify who your readers are, and provide the background they probably need to understand and benefit from your paper.

Just-in-Time Bridge by Way of Local Background

In the chapter on memory, we determined that there were two types of background: the global background found in the introductory parts of your paper, and the local background found wherever just-in-time information is needed. The local background includes the definition of highly specific keywords that only experts use. Problem is, you no longer realize that non-experts don't know these keywords. To use a metaphor, imagine yourself standing next to your reader at the foot of a hot-air balloon. You are about to embark on your research. You climb inside the gondola. As research progresses, the landscape you see is different from the one seen by the reader on the ground. Your discoveries raise your knowledge level as you drop the sandbags of your ignorance. That ignorance now becomes your reader's ignorance. Your balloon rises slowly (or rapidly, for one does not usually dictate the pace of discoveries). By the time you are ready to write, you have risen far above the reader, as in ☛1.
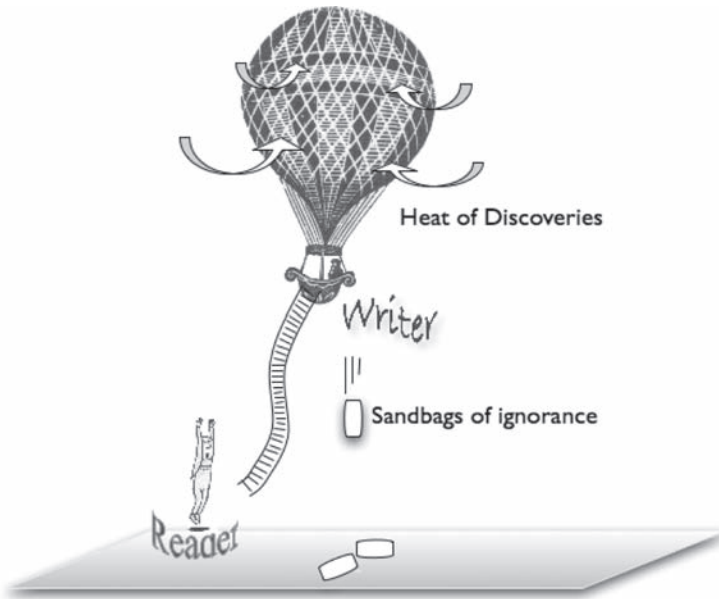
Figure ☞1
Knowledge elevates the writer above the reader.

Your job as a writer is to bridge the gap created through months of research. You have to throw a figurative ladder to readers so that they can come on board your hot air balloon gondola. How far down should you throw the ladder? Down to ground zero. Often the ladder is too short. Ground zero is not properly identified — the background is for tall experts only. Or, it may be that ground zero is properly identified, but rungs on the ladder are missing. These rungs correspond to the local background. Readers remain suspended in mid-air, stuck in the middle of a section of your paper, frustrated, trying to get on board but unable to climb further — you skipped an essential logical step or piece of background information that prevents them from completely benefiting from your paper, or you used a new word they do not understand.

Asking a reader to be familiar with reference [X] before being able to understand the rest of your paper is equivalent to asking the reader to get off the ladder, go to the library, read *the whole article* [X], and climb the ladder again. Add the missing rung, and keep on climbing.

☀ It would be more advisable, space allowing, to briefly summarize in your paper whatever reference [X] contains that is of interest to the reader.

The authors could have written this:

The dynamic behavior was expressed in the Unified Modeling Language (UML, Booch et al. (1999)). The notation used in Figure 3 is that of UML sequence diagrams.

Instead, they wrote this (note the footnote "*")

The dynamic behavior was expressed in the Unified Modeling Language (UML, Booch et al. (1999)). The notation used in Figure 3 is that of UML sequence diagrams.*

(in footnote) For those not familiar with the notation: Objects line up the top of the diagram. An object's messaging and lifeline boundary is shown by a vertical dashed line starting below the object. Object activity is shown by the activation bar, a vertical rectangle drawn along the lifeline. Horizontal arrows issued from a sender object and pointed to a receiver object represent the messages sent.

There is no need to build an elaborate rung on your ladder. Your goal is not to have the reader marvel at the exquisite design of a ladder rung, its elaborate anti-skidding grooves, its ornate structure… The role of the ladder is to take your reader inside the gondola of the air balloon where your contribution is revealed and understood. So here are the parting words at the end of this chapter. They come from biologist and writer, Thomas Henry Huxley.

"The rung of a ladder was never meant to rest upon, but only to hold a man's foot long enough to enable him to put the other somewhat higher."

Read your paper. Is your introduction too short? Is it motivating? Have you identified a ground zero that is reasonable to expect from your reader? Have you identified the background the keywords of your title make necessary? Have you identified the intermediate discoveries that removed the sandbags of your ignorance and elevated your knowledge above that of the reader? Should these constitute part of your background?

Chapter 6
Set the Reader's Expectations

> ### The train
>
> *Imagine the mind of the reader as a train. The author provides the set of tracks and the signal boxes. What could go wrong?*
>
> 1. *No tracks—Readers, left on their own, have to pick the loose rails lying around to create their own set of tracks. Very slow going.*
> 2. *No signal box—expectations are not set, so readers chug along slowly, not knowing where to go next.*
> 3. *Faulty signal—Readers, misled, are going down the wrong tracks.*
> 4. *The train is in a tunnel—Readers tolerate being left in the dark for a short while, as long as the end of the tunnel is in sight, and clarity returns.*

Creating and fulfilling the reader's expectations guarantee that reading will be fluid, interesting, and fast. That is why setting expectations is such an essential technique. In this chapter we will discover what sets expectations, and once we know, we will review how to set them. But first, we need to refresh some grammatical knowledge because grammar plays a role in the alignment of thoughts between reader and writer. A well-written sentence sets clear expectations that align the reader's thoughts and feelings with the writer's thoughts and feelings. Alignment is necessary to avoid distortions, misinterpretations, or ambiguity.

Expectations from Grammar

Main clause–subordinate clause

A **clause** contains a subject and a verb.

Learning needs to be semi-supervised.

Subject   Verb

A **subordinate clause** does not stand alone; to be understood, it needs a **main clause**.

Subordinate clause *Because variation within each class is large,* Main clause *learning needs to be semi-supervised.*

A sentence may have two **independent clauses**, which stand alone. They are separated by punctuation and often have a contrasting hinge (*'however'* and *'but'*).

Independent clause Generally, supervision is required Punctuation ; Other independent clause however, for classes with low variations it is not required.

Independent clause Generally, supervision is required Punctuation , Other independent clause but it is not required for classes with low variations.

These independent clauses are kept within the same sentence to establish the contrast between required and not required. Note that they share the same subject (*'supervision'*).

The placement of words in a sentence influences expectations

Do you want to find out how grammar sets your expectations? I suggest that we take a small survey together. Use a pencil if you wish to be able to erase the marks after the survey. You will see four sentences, each one on the topic of *Evolutionary Algorithms* — computer techniques based on biological evolution to find the optimal solution to some problems. After slowly reading each sentence, mark down how you feel about them. Do not re-read multiple times or ponder the matter at length before answering, because readers generate impressions "on the fly," as they read. One last piece of advice: if you want to make sure that the answer to a question does not influence the answer to the next question, take a short pause, look away from the book, take a sip of water, and return to the next sentence.

Feel good about Evolutionary Algorithms? Tick the box under the happy face ☺
Feel neutral about Evolutionary Algorithms? Tick the box under the neutral face ☺

Feel negative about Evolutionary Algorithms? Tick the box under the sad face ☹

(1) Evolutionary Algorithms are sufficiently complex to act as robust and adaptive search techniques; however, they are simplistic from a biologist's point of view.

☺ ☻ ☹
□ □ □
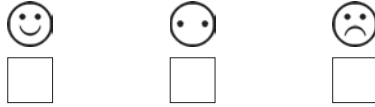
(2) "Evolutionary Algorithms are simplistic from a biologist's point of view, but they are sufficiently complex to act as robust and adaptive search techniques."[1]

☺ ☻ ☹
□ □ □
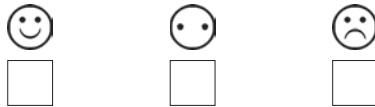
(3) Although Evolutionary Algorithms are sufficiently complex to act as robust and adaptive search techniques, they are simplistic from a biologist's point of view.

☺ ☻ ☹
□ □ □
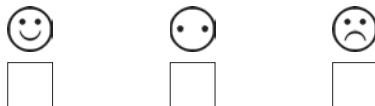
(4) *Evolutionary Algorithms are simplistic from a biologist's point of view, although they are sufficiently complex to act as robust and adaptive search techniques.*

☺ ☻ ☹
□ □ □
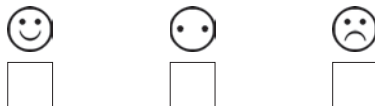
Now look back to your answers. Did you tick the same box for each sentence? If not, why not?

All four sentences present two identical facts: (1) Evolutionary Algorithms are simplistic from one point of view, and (2) Evolutionary Algorithms are sufficiently complex to act as robust and adaptive

---

[1] *(2) is reprinted with permission from Sinclair M. (2001) Evolutionary Algorithms for optical network design: a genetic-algorithm / heuristic hybrid approach. (name of university). Thesis.*

search techniques from another point of view. If you agree with one point of view (for example, if you find them simplistic), then all sentences 1 to 4 should carry the same message and be equally perceived, whatever the order of the words in the sentence. Yet, this is not the case, is it! You find some sentences favorable and others unfavorable. Something influences the way you perceive the facts presented. What is it? The answer is found in the grammar.

From a grammatical perspective, sentences 1 and 2 are similar in that they both have two independent clauses; what changes is the order of the clauses.

Looking at the opinions of a panel of 33 readers in Table ☜1, one sees the strong polarization of answers between the first two sentences: Sentence 1 is predominantly neutral to negative whereas sentence 2 is predominantly positive.

Table ☜1
Influence of the placement of the information in a sentence

	Sent.start	Sent.end	Main	Sub.	Neg.	Neutral	Pos.
Sentence 1	Pos.	**Neg.**	Pos/Neg		**15**	14	4
Sentence 2	Neg.	**Pos.**	Neg/Pos		3	4	**26**
Sentence 3	Pos.	**Neg.**	Neg.	Pos.	**22**	9	2
Sentence 4	Neg.	**Pos.**	**Neg.**	Pos.	12	**19**	2

To understand what happens, let's sort the sentences according to two criteria: the majority vote of the panel, and the way the Evolutionary Algorithms is featured at the end of each sentence — simplistic being negative and sufficiently complex being positive

Sentence (1): Panel ☹ and end sentence ☹ ✓ MATCH
Sentence (2): Panel ☺ and end sentence ☺ ✓ MATCH
Sentence (3): Panel ☹ and end sentence ☹ ✓ MATCH

People who took the survey seem to align their opinion on the opinion expressed at the end of each sentence.

☀ To encourage the reader to agree with you on what is important, place the information you consider important at the end a sentence.

We have not yet looked at sentence 4.

Sentence (4): Panel ◌ and end sentence ☹ ✗ NO MATCH

Why doesn't sentence (4) fully behave like the other three? What changes in the sentence to change the views of the panel?

Evolutionary Algorithms are simplistic from a biologist's point of view ← **Main clause, Subordinate clause** → *although they are sufficiently complex to act as robust and adaptive search techniques.*

The sentence the panel sees as neutral, represents a balance of opinions. It seems to be the result of a fight between two sources of influence with a bias towards the negative opinion. The negative point of view is expressed in the main clause and it seems that the panel of readers is influenced more by what comes in the main clause than by what comes in the subordinate clause. The reason people are indecisive is explained by the fight for influence:

Evolutionary Algorithms are simplistic from a biologist's point of view is in a weak position at beginning of sentence, tempered by the strong influence of the main clause; and *although they are sufficiently complex to act as robust* and *adaptive search techniques* has a strong position at the end of sentence tempered by the weak influence of the subordinated clause. They balance each other.

If what we suggest is valid, then sentence (3), which like sentence (4) has a main and subordinate clause, should verify the strong role of the main clause, and it does. In sentence (3), the information that gets the most votes from the panel is in the main clause.

☀ To encourage the reader to agree with you on what is important, place the information you consider important in the main clause.

But why stop here! We can extract more interesting guidelines from this fruitful example. You may have noticed in Table ☞**1** that the number of indecisive people is larger for sentence (1) than for sentence (2). Why is that?

It may be a combination of two factors: punctuation and the diluted strength of the adverb '*however*'. Whenever the brain encounters a punctuation like the period or the semi-colon, it stops and processes the meaning of the sentence. In the case of sentence (1), the brain gets a favorable impression of evolutionary algorithms at the semicolon pause. The second part of the sentence, starting with 'however' in the strong position at the end of sentence, attempts to reverse that already very positive impression. It succeeds partly (negative score greater than positive score), but not totally. The reason for that partial failure is due, in my opinion, to "however', an adverb overused to bring contrast, and often used wrongly to just change

topic (fake contrast). As a result the people who were positive only move to the neutral attitude, whereas the people who were neutral remain neutral or turn negative.

For sentence (2), the situation is quite different. It is very positive. Why? Again, it may be a combination of two factors: punctuation and the undiluted strength of the conjunction *'but'*. There is no semantic closure after a plain comma, the brain continues reading before coming to that final closure at the end of the sentence where the strong positive impression is formed. To reinforce that positive impression, the hammer-like *'but'* *rebuttal* squashes the negative impressions.

☀ To encourage the reader to agree with you on what is important, place the information you consider important after the conjunction 'but', and/or right before a period or a semi-colon.

There is still the matter of variability in the answers of the panel — the people who did not go with the majority vote. What influenced their opinions? There are several factors, which I list here.

1) People's temperament and outlook on life. Some are influenced more by what is positive than by what is negative.
2) People's job. The biologist may place emphasis on the negative statement whereas the computer scientist may have more positive feelings for the algorithm.
3) Some people consistently voted against the majority vote. I discovered that those who voted this way were the non native English speakers whose native language was based on Sanskrit. In their grammar, the important information is mostly placed at the beginning of a sentence, not at the end. So be aware that if you are in this situation after taking this test, your foreign grammar influences the way you write English, and that may confuse readers.
4) Your answer to the first question may bias your answer to the second question, and so on (remanence effect). That is why I regularly change the sentence order when conducting the test. The good news is (as far as you are concerned) that the guidelines stated here are confirmed by the 1000 + scientists who took the test.

☀ Consider starting sentences with a subordinate clause so that they end with a convincing main clause.

60

Forget about Mrs. Smith, your English teacher who told you to "**N**ever start a sentence with **B**e**C**ause", and she stressed every syllable to show you she meant it. Of course, had you asked "Why, Mrs. Smith?" She would probably have responded: "Because I say so".

The following words create a subordinate clause at the head of a sentence so that the main clause can end the sentence. They also make excellent attention-getters, and really shine at setting expectations.

Because…If…Since…Given that…When…Although…Instead of…While…

Take a word like *because*. Placed at the beginning of a sentence, *because* announces an upcoming consequence in the main clause. It sets a delay between the time the expectation is raised and the time it is fulfilled. That delay creates tension and momentum. The tension acts like a metallic spring: it pulls reading forward. In the real world, the length of a spring matters less than its strength; likewise, a sentence's length matters less than the tension created by the arrangement of its words.

☀ To build suspense within a sentence, start the sentence with a subordinate clause. This creates a tension that the end of the sentence releases.

Dynamic sentences have words that set expectations

Which words create expectations in the next sentence?

Up to this point, we have only considered basic filtration techniques.

If you answered: the locution *up to this point*, the adverb *only,* and the adjective *basic*, you were absolutely right. *Only* shows that more is coming. *Up to this point* works in tandem with *basic*. Together, they confirm that basic filtering techniques are no longer going to be covered. More advanced filtering techniques will now be considered. How do we know that? Take away the adjective *basic*.

Up to this point, we have only considered filtration techniques.

With *basic* missing, the reader expects that we are leaving the topic of filtration techniques to look at other techniques. So it is the adjective *basic* that sets expectations. Now test yourself with the next sentence. Which are the words that generate expectations?

Dengue fever epidemic in India does not occur at the beginning of the monsoon season.

The words that set expectations are the negation *does not* and the noun *beginning*.

1) Expectation of explanation — Why doesn't it occur at the beginning?
2) Expectation of elaboration — When exactly does it start?

This sentence indirectly tells us that dengue fever epidemic occurs, not at the beginning but sometimes *during* the Monsoon season. Its purpose is to emphasize when it starts. The writer prepares the reader to an explanation of the conditions for the epidemic to take place.

☀ Adjectives, adverbs, and nouns associated with a negation or with a pejorative meaning, allow fast transfer of expectations to their opposites.

Compare contrasted (A) with flat (B).

> **(A)** *Trapping is unimportant at high temperatures where there is plenty of energy to escape. But* **trapping** *leads to very slow dynamics at low temperature.*
>
> **(B)** **Trapping** *is important at low temperature because it leads to very slow dynamics, as there is not much energy for the molecules to escape.*

1) The total number of words for both texts is identical: 24 words. Paragraph (A) has two sentences: the first has one main clause and one subordinate clause, and the second has one independent clause. Each sentence is short (14 and 10 words). Sentence (B) is long (24 words) and the main clause has two cascaded subordinated clauses ("because", "as") making it more complex.

☀ Sentences that end with nested sub-clauses create weak expectations.

2) With (A), the reader is alerted on the role of low temperatures. Sentence (B) is less contrasted because low and high temperatures are no longer compared; it creates the expectation that the next sentence will keep the same topic, i.e. trapping. Sentence (B) does not create the same compelling

wish to know more about "very slow dynamics" because it is tucked away in the middle of the long sentence.

☀ Unguided, the default expectation is that the topic does not change from one sentence to the next.

3) But the punctuation has changed. The full stop in the original (A) has been replaced by a comma in (B). The full stop is better. It gives more breathing space than the comma. During this very useful breathing space, the brain processes the sentence and lets its meaning sink in. Based on this fresh understanding, the brain then sets expectations for the next sentence.

☀ Place the punctuation where it creates expectations.

Cognitive neuroimaging

Michael works in a cognitive neuroscience laboratory. He explores the brain with functional MRI, and seeks to understand what happens in our working memory. I ask him what happens when we read. Michael, an extremely well organized man, retrieves from his computer two papers from Peter Hagoort: "Integration of Word Meaning and World Knowledge in Language Comprehension," and "How the brain solves the binding problem for language: a neuro-computational model of syntactic processing."

Somewhat intimidated by the titles, I ask if he would not mind explaining in layperson's terms what happens when we read. Still facing his Macintosh Macbook Pro, he quickly thinks and asks, "Do you use Spotlight?" I reply, "Of course." Any Macintosh owner with the latest operating system is familiar with the search function of Spotlight, the little white magnifying glass inside a blue spot located in the top right corner of the Mac menu bar. "Look here," he says. I get closer to his screen. "As I type each letter in Hagoort's name, the search engine immediately updates the search results. H, then HA, then HAG. Notice how the list is now very small; one more letter and we will have zoomed down to Hagoort's papers."

As soon as he types the letter O, the list shrinks down to a few items, and among them, Hagoort's papers. He turns towards me as I sit back into the chair facing his desk. "You see," he says, "it looks as though the Mac tries to guess what you are looking for. Similarly, while you read, your brain is active, forever seeking where the author is going with his sentence. It analyses both syntax and meaning at the same time, going from one to the other transparently."

Expectations from Science

Verbs, adjectives and adverbs make claims

The next example combines expectations from science and expectations from grammar. Do you remember the example with Tom Smith (Fig. ☛1 of Chapter 1)?

Tom Smith's assumption [4] that no top layer material could come from the by-products of the pinhole corrosion which had migrated is not supported by our data.

The sentence is written in a passive voice: the subject (the "*assumption*") is not the doer of the verb ("*supported*"). The real doer — the "*data*", is placed after the verb. '*By'* is often a marker of the passive voice ("*by our data*").

Because two subclauses and 17 words separate the subject from its verb, the sentence is difficult to read. Despite all its problems, that sentence still manages to create an expectation. A scientific claim is made: the data **does not support** Tom Smith's assumption. As scientists, we expect the writer to support his claim with data. We certainly do not expect to hear about Tom Smith's assumption because the lengthy insert does that already.

Here is the sentence again, but this time with verb closer to subject, and with added dynamism thanks to the active voice. This sentence is written in the present tense, a tense usually used for claims. The auxiliary "*do*" shows certainty.

Our data reveal that, contrary to Tom Smith's assumption [4], the pinhole corrosion by-products do migrate to form part of the top layer material.

Have the expectations changed? What do you expect the writer to introduce in the next sentence? The data or Tom Smith? The majority of you will answer, "The data". There is a slight difference with the first sentence, though: the expectations are more defined now. The reader expects proof of the claim regarding migration, or proof that the material found on the top layer comes from the pinhole, not just any data. Tom Smith is now between commas, a side remark, and nobody cares about his assumption. The findings of the author are stated, not those of Tom Smith.

But imagine for an instant that you are the writer and you want to focus the reader's attention, not on the data, but on Tom Smith's assumption. How would you rewrite the previous sentence?

There are two ways of rewriting the sentence; both have in common the placement of Tom Smith at the end of the sentence.

> *Our data reveal that the pinhole corrosion by-products migrate to become part of the top layer material, contrary to Smith's assumption [4].*

> *Our data reveal that the pinhole corrosion by-products migrate to become part of the top layer material. These findings contradict Tom Smith's assumption [4].*

In this last sentence, Tom Smith's assumption is no longer a side remark, it is the main point and it comes in a small package: a short punchy sentence. The reader is now curious. What did Tom Smith assume? Why is there a contradiction? We know what the findings of the writer are. What we do not know clearly is what Tom Smith assumed.

☀ Expectations of change are mostly set by the new information found at the very end of a sentence.

So far, we have looked at claims made by verbs (*'is not supported,' 'do migrate'*). Adjectives and adverbs also make claims. Readers of scientific papers have different expectations than readers of novels. When the novelist writes: "the ferocious dog", the reader's imagination creates a ferocious dog from past face to face encounters with such dogs or from movies where such dogs are seen. It may not be the type of ferocious dog the writer had in mind, but who cares — the more ferocious the reader makes it, the better! Unlike the scientist, the novelist does not have to convince the reader of the dog's ferocity by measuring the surface of the barred dental area, the number of milliliters of saliva secreted per minute, or the dog's pupil dilation — all this on live ferocious dogs of course. *In vivo* data collection is a dangerous occupation, especially if you want your population to be representative, and go beyond the Chihuahua to include German Shepherds, Doberman Pinschers, and Rottweilers.

Adjectives or adverbs are subjective. What is robust to you may be fragile to me. What is very fast to you may be moderately fast to me. In Science, adjectives and adverbs are claims. When the claim

stated is widely accepted as true, however, the scientist does not expect the writer to justify it. But remember, what is obvious to an expert, may be non-obvious to a non-expert who might still want you to justify your adjectival claim. It is safer to justify than to assume no justification is needed. And it is safer to define an adjective than to let the reader define it.

> *When feed gas changes from light to heavy gas, the plant load decreases by a small fraction (4.7 %).*

In this example, the adjective *small* is immediately quantified. Elsewhere in the paper is a table that gives the definition of what constitutes heavy, medium, and light feed gas, using feed gas components (mole fraction) expressed in % or ppm.

☀ Quantify adjectives when possible or define them.

Read your abstract and your introduction. Highlight all adjectives in fluorescent yellow, and adverbs in fluorescent red. If your paper glows in the dark, then you have work to do. Examine each adjective and adverb. Are the claims justified? Would removing an adjective make you more authoritative? Could each adjective be replaced by a fact?

The main sections of a paper create their own expectations

The next few examples show that expectations are also guided by the part of the paper the sentences appear in, in this case respectively, the introduction, the methodology, and the abstract.

This sentence, the first in the introduction, makes an adverbial and adjectival statement that creates a strong expectation regarding the goal of the paper.

> *'Traditionally, airplane engine maintenance has been <u>labor-intensive</u>.'*

Both "Traditionally" and "has been" confirm that the labor intensive situation is as true today as it was in the past. The reader does not consider *'labor-intensive'* as a scientific claim but as an established fact. This sentence sets the context. Its topic is airplane engine maintenance. The reader expects novelty. In this case, the expectation is that the author has found a novel way to make airplane engine maintenance less labor-intensive by using robots, by reducing the number

of parts, or by improving the quality and durability of airplane engine components. But wait! How would the reader know that the contribution is about less labor-intensive maintenance schemes, and not more labor-intensive? Where does the reader get these expectations from? They are born out of general culture and understanding of economics in developed nations, i.e. the economies of scale allowed through automation.

☀ Context of statement and general culture shape expectations.

This other example from the methodology section of a scientific paper establishes expectations around a time-ordered sequence. The clue that a time sequence is in progress is clear. Each sentence starts with, or contains, an expression related to time: 'day one,' 'next three days,' 'after,' 'over the next…'.

> *On day one, thirty embryonic cells are placed in the culture dish. For the next three days they are left to proliferate. After proliferation, the cells are collected and put into new culture dishes, a process called replating. Over the next three weeks, 180 such replatings produce millions of normal and still undifferentiated embryonic cells.*

☀ Expectations of sequential progression require sequence markers such as time markers, verb tense, or consecutive numbers. In this case, expectations are set by the first words of the sentence (not its last words).

In this final example, an abridged abstract, the author laid out the sentences according to a specific order expected by the reader: that of a scientific paper: aim, methodology, result, and significance, which also corresponds to that of the scientific process: observation, hypothesis, experiment, result and interpretation.

> **[Observation]** *The dengue genome forms a circle prior to replication, as is the case for the rotavirus. Since one end of the circling loop is at the 3′ end of the genome where replication takes place,* **[Hypothesis]** *we wondered if the loop had an active role to play in the replication.* **[Experiment]** *After comparing the RNA synthesis capability of various whole and truncated dengue genomes using radio-labelled replication arrays,* **[Results]** *we found that a region other than the 3′ end of the genome had an even larger role to play in the replication: the 5′ end of the genome. Although far away from the 3′ end, it seems to loop back into it.*

[Discussion] Thus, it may be that the promoter site for RNA synthesis resides in this unusual location. Looping would then be a means of bringing the promoter to where it can catalyze rapid duplication.

☀ Expectations of logical progression require sequence markers such as logical markers ('if,' 'then,' 'thus,' 'therefore'), verb tense or auxiliary to express suggestions ('would,' 'could,' 'may'), or consecutive numbers expressing logical steps which are not necessarily based on a time sequence.

Chapter 7
Set Progression Tracks for Fluid Reading

Giant George

I mentioned in the preface of the book that I am very much indebted to Dr. George Gopen. His book "Expectations: Teaching Writing from The Reader's Perspective" introduced me to sentence progression schemes built around the concept of topic and stress. To demonstrate the power of Dr. Gopen's impact on my writing, I need to reveal this somewhat embarrassing personal story.

Back in 2002, I rewrote a convoluted paragraph from a Ph.D thesis. It was beautifully rewritten: clear, concise. The students in my class were in awe. What I had not told them was that the rewriting had taken me close to one hour and twenty minutes of solid work involving many drafts. Naturally, they all wanted to know what technique I had used to rewrite such a masterpiece, but all I could do was wave hands and claim that when you "feel" something is still not right, you go on to the next draft. That was true, of course, but not encouraging to the students because my response meant that if they did not have that feel, there was no hope for them. Writing was an art, not a science.

Looking back to what happened then, I realize how much ground I have covered. Not only can I rewrite the same paragraph in less than twenty minutes, but I can also explain how to apply the progression technique to any paragraph. It is no longer art, it is science.

When readers cruise down your paper in fifth gear, it is because you have created a highway for their thoughts to travel on at great speed, a highway that stops their mind from wandering where it should not go. Sometimes, while reading certain papers, I feel as if I am driving in the fog at a crawling speed across a muddy field, trying to follow

somebody else's tracks. In Science, unlike literature, you guide your readers along a clearly lit well-signposted highway. How to build such a highway is described in one word: progression.

Progression is the process of transforming what is new into what is known. It builds a coherent context that allows readers to travel light and read on with minimum cognitive baggage. When readers start a sentence, a paragraph, or a section in your paper, they relate what they read to what they know. This progressive anchoring of new knowledge onto old knowledge is an essential learning mechanism.

Progression schemes come in two flavors: topic-based progression and non-topic based progression such as progression by explanation, logical step, or transition word. Often times, these two schemes are at work together within a same sentence. In this chapter they are considered separately to enable you to identify them more easily.

The reader may wonder if there is a connection between expectation and progression. Sometimes they are joined and work in synergy. For example, when a sentence starts with *first*, it sets the expectation that the following sentences will cover what comes after; the sequential progression fulfills the expectation. Sometimes expectation and progression are distinct and work separately, hopefully in the same direction. It does happen that they pull the reader in different directions, which is not advisable, of course. Progression should always support the expectation, not the opposite.

There are other differences between the two. Expectations have a longer reach. For example, a question's reach extends far beyond the next sentence. Progression is purely local: between two phrases within a sentence, two sentences within a paragraph, or two paragraphs within a section. Expectations open the reader's mind. Progression keeps the reader's mind on your rails.

If your expectation is to now plunge into the classes of progressions in the upcoming paragraph, I need to reset your expectation. You first need to review a few grammatical points to clarify the meaning and role of topic and stress. In France, schoolchildren discover all about topic and stress (in French *Thème* et *Propos*) in their grammar book at the age of 14. By the time they go to university, they have forgotten all about them! The same probably applies to you, so here is a quick refresher course.

Topic and Stress

Topic and stress have a well-defined meaning in grammar. In a simple sentence, the topic corresponds to the known information located at the head of a sentence (often the subject of the main verb), and the stress is the new information located after the topic (the active verb and the rest of the sentence).

| **Topic** > *But trapping*
| **Stress** > *leads to very slow dynamics at low temperature.*

As soon as you read *But…*followed by a noun, you know that the previous sentence mentioned that noun; therefore, the topic is old information. It comes upfront in this sentence. It is the subject of the verb *leads*.

☀ To identify a topic, evaluate whether what it describes is known and is at the beginning of a sentence. The stress is new to the reader.

The following sentence raises slightly different expectations based on the information the reader already knows (the sentence topic). Before you look at the sentence, some background might be useful. When one uses an image (instead of text) to query a database of images and retrieve the image closest to the query image, one uses a "content-based image retrieval system". Medical images usually belong to one of a finite number of categories, whereas general image retrieval systems have an infinitely large number of categories. What are all the expectations the next sentence is likely to raise?

For content-based medical image retrieval systems, the number of classes is limited.

- • Expectation 1: Other image retrieval systems

*"For content-based **object** image retrieval systems, however, the number of classes is **unlimited**," or "The number of classes is **infinite,** however, for content-based **object** image retrieval systems."*

- • Expectation 2: Quantification of claim

*"With **64 classes**, the whole human body …"*

- • Expectation 3: Justification of claim

71

*"**This is because** medical images are dealing with a finite number of body parts: those visible using current imaging techniques."*

- Expectation 4: Consequence of the limitation

*"**This guarantees** that an optimum search path exists."*

Out of the four candidates, the most likely ones will depend on the sentence structure, the placement of the topic and the stress, and the information presented in the previous sentences (context). I will let you decide which expectation each sentence triggers. My candidates may not be yours - but what is important for you to discover during this exercise is that expectations change even when all sentences carry the same raw information using the same words.

1. Classic case of Topic followed by Stress

(A) | ***Topic >** For content-based medical image retrieval systems, the number of classes* | ***Stress >** is limited.*

Out of the four expectations listed, which one(s) do you have for the sentence that follows (A), given the sentence stress?

(B) | ***Topic >** For content-based medical image retrieval systems,* | ***Stress >** the number of classes is limited.*

Out of the four expectations listed, which one(s) do you have for the sentence that follows (B), given the sentence stress?

(C) | ***Topic >** The number of classes* | ***Stress >** is limited for content-based medical image retrieval systems.*

Out of the four expectations listed, which one(s) do you have for the sentence that follows (C), given the sentence stress?

2. Split Topic or Split Stress

What if we had sandwiched the stress between a split topic? Does that change **expectations**?

(D) | ***Stress >** For content-based medical image retrieval systems,* | ***Topic >** the number of classes* | ***Stress >** is limited.*

Why would the writer start the sentence with a stress, creating a bump in the traditional topic-followed-by-stress road? It is difficult to ignore a bump in the road. It slows down

reading because it resets the context. By modifying the normal order, the writer is emphasizing the first segment, alerting the reader that it will become the topic of the next sentence.

(E) | ***Topic >*** *The number of classes* | ***Stress >*** *is limited* | ***Topic >*** *for content-based medical image retrieval systems.*

The writer centers the limitation away from *the number of classes* towards *the content-based medical image retrieval systems*. The old topic, now in a stress position is re-stressed, rejuvenated, and ready to reappear in topic position in the next sentence.

3. Inverted Topic

This situation is to be avoided for it violates reader expectations. Let us change example.

__The cropping process__ should preserve all critical points. Images of the same size should also be produced __by the cropping__.

These sentences do not seem well balanced, would you not agree? This is because, in the second sentence, the already known information (*the cropping*) is at the end in a stress position, a place traditionally reserved for new information, whereas the new information is in the place reserved for context. This inversion delays the understanding of the sentence until its topic finally arrives to clarify everything.

Here are three ways to correct the problem:

Change the voice in the sentence from passive to active or vice versa, thus straightening the inverted topic and stress by bringing the known information to the head of the sentence.

The cropping process should preserve all critical points. It should also produce images of the same size.

Invert the order of the sentences to re-establish progression.

Images of the same size should be produced by the cropping. The cropping should also preserve all critical points.

Combine the two sentences into one.

The cropping process should preserve all critical points and maintain the size of images.

☀ Do not invert topic and stress in a sentence.

In examples (A) to (E), the verb that starts the stress is weak (*to be*). When the verb that starts the stress is strong, it may overshadow the rest of the stress and override weaker expectations.

> *Applying Kalman filters* **reduced** *the noise in the data sent by the low-cost ultrasonic motion sensors.* **The reduction** *was sufficient to bring down the detection error rate below 15%.*

☀ Use strong verbs to control expectations. 💬

Three Topic-Based Progression Schemes to Make Reading Fluid

Progression around a constant topic

The scheme is straightforward. The subject of the sentence is repeated in successive sentences, directly or by a pronoun, or a more generic name. The reader is already familiar with the topic and reading is fluid. In this example, 'trapping' (of the molecule) is the constant topic.

> *"***Trapping** *is unimportant at* **high temperatures** *where there is plenty of energy to escape. But* **trapping** *leads to very slow dynamics at* **low temperature**. *In the case of liquids, this* **trapping** *causes the glass transition — a dramatic slowing of motion on cooling."*[1]

In the first sentence, the author skillfully sets expectation for what comes next: an explanation of why trapping is important at low temperatures. In the second sentence, this expectation is reinforced by '*but*' to enhance the contrast. The adverb '*very*' creates the expectation of an illustration of very slow dynamics. The third sentence provides the illustration: *"the glass transition, a dramatic slowing of motion on cooling"*. Note the use of an apposition to explain *glass transition* in a just-in-time fashion. Note also the pattern: stress — topic — stress of the final sentence, allowing the writer to emphasize that the example given only applies to liquids.

---

[1] Reprinted excerpt with permission from Wolynes PG. (2001) Landscapes, Funnels, Glasses, and Folding: From Metaphor to Software, *Proceedings of the American Philosophical Society* **145**: 555–563

Progression through partial aspects or subclasses of the main topic

In this progression, the main topic is usually announced in the first sentence, and subsequent sentences examine aspects of the topic. In the following example, the first sentence is about visuals. The next two sentences review two aspects of visuals: their placement, and their convincing power.

> **Visuals** are star witnesses standing in the witness-box to convince a jury of readers of the worth of your contribution. **Their placement** in your paper is as critical as the timing lawyers choose to bring in their key witness. But most of all, **their convincing power** is far beyond that of text exhibits.

Chain progression

In a chain progression, topic and stress are daisy-chained. The stress at the end of a sentence becomes the topic at the beginning of the next sentence. This frequently used progression scheme is easy for readers to follow. It is illustrated in the next paragraph.

> *"The Brownian motion of the protein strand will carry it willy-nilly between various shapes, somehow finally getting it to settle down into a much less diverse family of shapes, which we call the* **"native structure" of the protein***. The average* **native structures of many proteins** *have been inferred experimentally using X-ray crystallography or NMR."*[2]

The elements in a daisy chain do not need to be repeated word for word from one sentence to the next. Often, the verb in a previous sentence becomes a noun at the head of the next sentence.

> *Applying Kalman filters* **reduced** *the noise in the data sent by the low-cost ultrasonic motion sensors.* **The reduction** *was sufficient to bring down the detection error rate below 15 %.*

Sometimes, part of the previous sentence (bolded in the following paragraph) is brought forward in the next sentence by way of a pronoun such as '*This*'.

---

[2] *ibid.*

Topic-Based Progression schemes
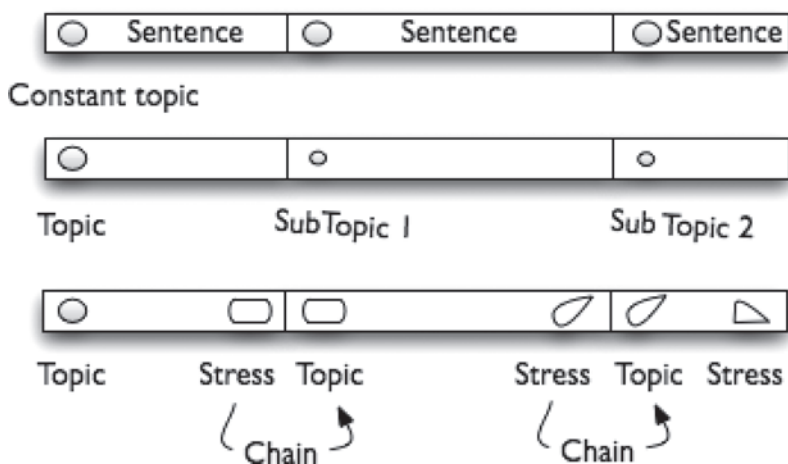
Figure ☞1

Three topic-based progression schemes. A sentence is represented by a rectangle; its topic by an object at the head of the sentence. In the chain progression, the topic retains the same shape as the previous stress to indicate that the stress of one sentence becomes the topic of the next sentence.

> *"The above observations can be generalized to a rather important conclusion. If **large mole differences between species exist in a data set** (and **this** is often the normal case for catalytic reactions), then the reactions involving both major and minor species should be rewritten to include only the latter. **This** should solve the problem of abnormal gradients in the extent of reactions for most cases."*[3]

Observe how the authors capture the readers' attention with words of importance ('generalized to a ***rather important*** conclusion'), or words attractive to scientists ('*solve the problem* of abnormal gradients', '***generalized*** to a rather important *conclusion*'). Don't fall in love with this paragraph though. It has its problems. It could be more concise: the *rather* in '*rather important*' is unnecessary. It also contains a wonderful example of an ambiguous pronoun in the last sentence. What does '*This*' represent? If you answer '*the reactions*,'

---

[3] Widjaja E, Li C, Garland M. (2004) Algebraic system identification for a homogeneous catalyzed reaction: application to the rhodium-catalyzed hydroformylation of alkenes using *in situ* FTIR spectroscopy. *Journal of Catalysis* **223**: 278–289, with permission from Elsevier.

read the sentence again because it is not the correct answer. '*This'* refers to the rewriting.

Each of the three topic-progression schemes (Fig. 1) influences expectations and vice versa.

The constant topic progression answers the need to know more about the same topic (expectation of elaboration — breadth). The topic/subtopic progression answers the need to go deeper into the topic (expectation of elaboration — depth). The linear progression answers the need to see how things are related (expectations of relatedness and outcome). Non-thematic progressions have even more direct ties with expectations, as we shall see.

Non Topic-Based Progression Schemes

Progression through explanation and illustration

The second sentence in the following paragraph introduces a progression of a new type: the **explanation**.

It usually follows a question or a statement that acts as a question.

Although the new algorithms have worst-case polynomial complexity, in practice they are shown to perform faster than the linear time algorithm. This is because their construction methods have good locality of reference and they have taken the computer's memory architecture into consideration.[4]

In this example, the first sentence says: the new algorithms should have performed worse than the existing ones, but they outperform them. The reader is intrigued and waiting for an explanation. It comes in the second sentence. Notice that there is also a chain link: the main clause of the first sentence is represented by the pronoun "this" at the beginning of the second sentence. More pronouns ("their" and "they") strengthen the link to the topic of the first sentence (the new algorithms).

The progression by illustration is the easiest to establish: just write (Figure #)! But you can also announce the progression with

---

[4] Gopalkrishnan V, Sim K. (2005) Efficient Suffix Tree Construction for Sequence Repeats. Joint International Conference of InCoB, AASBi, KSBI (Bioinfo), Busan (Korea). (with K. Sim)

the transition words *for example, similarly, similar to, an analogy for, etc.,* which prepare an explanation based on a related example, an analogy, or a metaphor.

> ### What! Metaphors?
>
> *These last two chapters seem to discourage calling on the reader's imagination. One must define ferocity. One must keep the reader on track and channel reader thoughts. Does this mean that, you, the writer scientist, should keep reader imagination at bay because Science is objective? Many examples in this book come from an article written by Professor Wolynes entitled "Landscapes, funnels, glasses, and folding: from metaphor to software". Rather than trying to convince you, I will quote here the first few sentences of his article. "Of all intellectuals, scientists are the most distrustful of metaphors and images. This, of course, is our tacit acknowledgment of the power of these mental constructs, which shape the questions we ask and the methods we use to answer these questions."*

Time-based progression

Another instance of a non topic-based progression, the time-based progression, is commonly found in the methodology section of any scientific paper.

> *"The protein when it is **first made exists** in an extraordinarily large variety of shapes, resembling those accessible to a flexible strand of spaghetti. (2) The Brownian motion of the protein strand **will carry** it willy-nilly between various shapes, somehow **finally** getting it to settle down into a much less diverse family of shapes, which we call the "native structure" of the protein."[5]*

In this time-based progression, the passage of time is expressed by changes in the verb tenses from the present to the future, and by the adverbs *first* and *finally*. Note also a topic/stress to subtopic progression: the first sentence links protein and strand while the second sentence looks at an aspect of the protein strand — its motion.

---

[5] Reprinted excerpt with permission from Wolynes PG. (2001) Landscapes, Funnels, Glasses, and Folding: From Metaphor to Software, Proceedings of the American Philosophical Society **145**: 555–563

Although words such as *first, to start with, then, after, up to now, so far, traditionally, finally, and to finish* mark the start, the middle, or the end of a time step, time is often implicit. The reader scientist understands that the writer is following the logic of time when narrating the various steps of an experiment. Most often, the passage of time is established by changing the tense of a verb, from the past to the present, or from the present to the future.

Logical Sequential Progression

An enumeration is a type of sequential progression: from the first to the last. Progression can be numerical, but it can also follow an order defined by the author (the elements of a list, for example). In the next sentence, the author announces two factors that contribute to the propagation of dengue fever before covering each one in turn.

Two factors contribute to the rapid spread of dengue fever: air transportation, and densely populated areas.

And even this fortune cookie message attributed to R. Lincoln meets expectations.

There are two rules for success. 1) Never tell everything you know.

The progression can follow implicit logic, or explicit logic such as the cause and effect relationship announced by the words *thus, therefore, because…*

Progression through transition words

Progression is sometimes announced by special words called transition words, such as *in addition, moreover, furthermore, and, also, besides*. These linking words are a topic of controversy among people who teach writing. Some say, they are just a convenient way to ignore progression; they artificially establish a link where none exists. In truth, it is often the case. I recommend that when you see these transition words, you try to replace them with an implicit progression, such as a sequential step or a topic progression. If you cannot replace them, it may be that an explicit progression using these transition words is necessary, as in this conclusion paragraph:

Our method determined the best terminal group for one specific metal-molecule coupling. In addition, we confirmed that Smith's formula for

calculating molecular binding energy is less computationally intensive and more accurate than the frequently used formula proposed by Brown [8].

The two sentences are independent. *'In addition,'* here, emphasizes the extent of the contribution of the paper.

Pause in Progression

Progression is not a goal in itself. Sometimes the reader, despite the best efforts of the writer, gets lost because of insufficient knowledge or because too many distractions occurred while reading. To let the non-expert reader catch up, the writer needs to pause, in particular where the concepts take much concentration to understand, even for an expert! During that pause, the reader can consolidate the acquired knowledge with a writer's supplied summary or illustration. To announce that pause, use words such as *to summarize, briefly put,* and *for example* to name a few.

Troubleshooting Progression Problems

Unstructured paragraph

The following long paragraph has 97 words and five sentences. A tell-tale sign of progression problems is the reliance on transition words (bolded) to glue together sentences instead of tying these through a strong logical, structural bond (hence the need to *re-structure*). But before we start working on it, some background is necessary.

When the service light on the dashboard blinks, you take your car to the garage for a check-up. The mechanic opens the car bonnet (hood), finds a secret socket you never knew was there and connects it to an electronic diagnostic machine by way of a long cable. The machine spews out a paper ribbon telling you what requires maintenance, and when to replace a number of critical parts. How does the machine know that? It knows because statistics of parts failure have been collected over many years on thousands of similar cars and used to create "prognostic models".

"1. Formation of prognostic knowledge is concerned with extracting knowledge from historical data in a maintenance and diagnostic system. 2. Different prognostic approaches can be used based on the characteristics of the equipment and the nature of the failure problems. 3.

For example, in cases where complete knowledge of THE EQUIPMENT is not available, it will be difficult to produce a comprehensive model for model-based prognosis. 4. *However*, it is possible to use experience-based or AI-based approach [6] to extract the PROGNOSTIC KNOWLEDGE. 5. *Furthermore* it is possible to gradually build the comprehensive prognostic system by combining the results of MULTIPLE APPROACHES."[6]

The words in all caps represent the old information from previous sentences, and the words in **bold** are words acting as transition words. Apart from sentence 3 which seems to be in a chain progression with sentence 2, the old information is systematically placed at the end of the sentence in an inverted stress/topic position. To compound the problems, sentences 1 and 2 bring totally new information and have unclear connections, and the transition words (*'for example,' 'however,'* and *'furthermore'*) hide the lack of progression.

For each sentence, let's determine both topic and expectation.

(1) *Formation of prognostic knowledge is concerned with extracting knowledge from historical data in a maintenance and diagnostic system.*

This first sentence announces the main topic of the paragraph by defining it; it helps the reader. Its brand new topic is "**Prognostic knowledge**". The reader cannot form a clear expectation because there are four possible following topics: *The formation of prognostic knowledge, prognostic knowledge, extraction methods,* or *maintenance and diagnostic systems.* As a result, the reader defaults to the standard expectation: a repeated topic.

(2) *Different prognostic approaches can be used based on the characteristics of the equipment and the nature of the failure problems.*

The brand new topic is "prognostic approach". Approach is one of these words that should be banned from scientific English. What does the approach refer to? The formation of prognostic knowledge, the extraction methods? To the reader, this is unclear, yet what is clear is what is expected in the next sentence: one of the approaches (whatever that means) for one characteristic of equipment and type of failure.

---

[6] Zhang DH, Zhang JB, Luo M, Zhao YZ, Wong MM. (2004) Proactive Health Management for Automated Equipment: From Diagnostics to Prognostics. *Proceedings of Eighth International Conference on Control, Automation, Robotics and Vision* (ICARCV2004), pp. 479–484. Copyright 2004 I.E.E.E

(3) *For example, in cases where complete knowledge of the equipment is not available, it will be difficult to produce a comprehensive model for model-based prognosis.*

What should have been the topic (one announced prognostic approach) is in the wrong stress position: "*Model-based Prognosis*". The reader anticipated that topic (the prognosis approach) to be upfront in the sentence. The reader, encouraged by the transition words '*For example,*' thinks the writer will present one approach; instead, he is given the negative result of one approach, i.e*., a difficult-to-produce comprehensive model*. The disorientation does not stop there. Instead of being given equipment characteristic or failure type, the reader is asked to move up to a higher conceptual level: the *complete knowledge* of characteristics of equipment and types of failure. As a result, the reader is further disoriented.

The expectation set for the next sentence is unclear but the reader may think that the writer will introduce other "approaches" that address the difficulty.

(4) **However**, *it is possible to use experience-based or AI-based approach [6] to extract the prognostic knowledge.*

Again, the topic *Prognostic knowledge* (old known information) is placed at the end of the sentence in stress position. The transition word '*However,*' announces a way to address the difficulty. The expectation is now reset to prepare the reader to know more about these methods.

(5) **Furthermore** *it is possible to gradually build the comprehensive prognostic system by combining the results of multiple approaches.*

Inverted topic / stress situation again. Is the comprehensive model the same as a comprehensive system? The adjective is the same, but the nouns are different. Knowledge, system, approach, everything is mixed.

A simple look at the succession of sentence topics reveals a lack of structure. This paragraph needs re-structuring.

(1) Prognostic knowledge (The whole sentence contains new information)

(2) Different Prognostic approaches (The whole sentence contains new information)

(3) Chain but on a part of the stress of sentence 2 (the characteristic of equipment)

(4) Prognostic knowledge (topic of sentence 1 but inverted in stress position)

(5) Multiple approaches (topic of sentence 2 but inverted in stress position).

Before we restructure, let's take stock of what we know:

For complete historical knowledge, model-based prognosis works and is comprehensive. For incomplete historical knowledge, models don't work well, and AI or experience-based methods are there to bridge the prognostic knowledge gap until such times when enough historical data are available to make the prognosis system comprehensive again. To reach comprehensiveness over time, however, will require the integration of all knowledge extracted from all methods.

Notice that we have a time-based structure with an end (sufficient historical data — model is enough), a beginning (no historical data (new methods are required), and a middle (insufficient historical data for some equipment pieces or failure types (combination of methods is required). We can now rewrite based on a solid structure.

*[**Revised — version 1**] When complete historical data from a maintenance and diagnostic system are available for all failure types and equipment parts, comprehensive prognostic knowledge is easily formed and used in model-based prognosis. When little or no historical data are available for a failure type or equipment part, prognostic knowledge extraction methods — experience-based or AI-based [6] — are helpful. Combining the prognostic knowledge from all methods allows the gradual build of a comprehensive prognostic system.*

The revised paragraph is without transition word (in bold in the original). On the positive side, it has only 74 words versus 100 words in the original — a 26 % word count reduction. It has 3 sentences instead of 5. On the seemingly negative side, the number of words per sentence is higher (25 versus 20), and the attention-getting words (for example, however) are gone. Yet, readers find the text clearer

because of the parallel structure. Therefore, **structure brings clarity, and clarity is not always related to the length of a sentence.**

> *[Revised — version 2] Knowledge on which to base a prognosis is extracted from the data retrieved from maintenance and diagnostic systems. When historical data from these systems is available for all equipment parts and failure types, models are sufficient to build a comprehensive prognostic system. When data are missing, the prognostic system is gradually built using a mixture of Experience-based, AI-based techniques, and models.*

63 words versus 100 words in the original — a 37% word count reduction. With an average of 20 words per sentence, it is similar to the original, but it is clearer and more concise.

Interrupted topic

Sometimes, the progression breaks as in example (A) below. It stops for a sentence or two, and then is resumed. In such situations, readers rapidly lose their sense of direction. Somewhere, somehow, one or two links in the progression chain are broken, but where? How does one identify a broken link?

1. Underline the topic of each sentence in a paragraph.
2. Circle the topics that are not part of a topic-based progression (i.e. not connected to topic or stress of the previous sentence).
3. See if these topics are in a non-topic based progression (pause, explanation, time or logical sequence). If they are not, then congratulations, you have just located a broken link or an inverted topic / stress situation.

(A) *After conducting microbiological studies on the cockroaches collected in our university dormitories, we found that their guts carried staphylococcus, members of the coliform bacilli, and other dangerous micro-organisms. (2) Since they regurgitate their food, their vomitus contaminates their body. (3) Therefore, the same microbes, plus molds and yeasts are found on the surface of their hairy legs, antennae, and wings. (4) To find such micro-organisms in their guts is not surprising as they are also present in the human and animal feces on which cockroaches feed*

Table ☛2 reveals the problem. Did you notice that the stress of sentence 1 is the topic of sentence 4. Sentences 1 and 4 should be

Table ☛2

Topic/stress table reveals progression issues.

TOPIC	STRESS	SENTENCE#
microbiological study on cockroaches	**micro-organisms in guts**	1
cockroaches	regurgitation & vomitus contamination of cockroach body	2
micro-organisms from vomitus	parts of cockroach body	3
Micro-organisms in guts	human feces as cockroach food	4

next to each other in chain progression strengthened by a logical progression (effect-to-cause). Sentences 2 and 3 are difficult to separate because they are in a logical progression ("since" … "therefore").

Table ☛3

Topic/Stress Table-based reveals clear progression — chain and constant topic.

TOPIC	STRESS	SENTENCE#
microbiological study on cockroaches	**micro-organisms in guts**	1
micro-organisms in guts	human feces as cockroach food	2
micro-organisms in guts	parts of cockroach body	3
micro-organisms in guts	regurgitation & vomitus contamination of cockroach body	4

> **[Revised]** *After conducting microbiological studies on the cockroaches collected in our university dormitories, we found that their guts carried staphylococcus, members of the coliform bacilli, and other dangerous micro-organisms. To find such micro-organisms in their guts is not surprising because they are also present in the human and animal feces on which cockroaches feed. But why are these micro-organisms, plus molds, plus yeasts, also found on the surface of their hairy legs, antennae, and wings? The micro-organisms are transported out of the guts when the regurgitating cockroach vomits its food and contaminates its body.*

The suspense has been increased by turning sentence 3 into a question, that sentence 4 answers. Sentences 1 and 2 are in chain progression. Sentences 2 and 3 share a constant topic. Sentence 4

is in a non-topic based progression with 3 through the explanation (see Table ☛3).

The next paragraph is about a tropical and sub-tropical disease called dengue fever. Compare the original version with the final version to discover how the text improved.

Original text

The transmission of the dengue virus to a human occurs through the bite of an infected female Aedes mosquito. In addition, the disease spreads rapidly in densely populated areas because of the lack of effective mosquito control methods, the increase in air travel, and poor sanitation in areas with a shortage of water. The mosquito becomes infected when it feeds on a blood meal from a human carrier of the virus. The virus multiplies inside the infected mosquito over three to five days and resides within its salivary gland.

Follow these steps to analyze the original text.

- Identify the author's intention, and the main points of the paragraph.
- Identify closely related sentences by the words they share. Identify potential sentence topics — usually the words repeated from one sentence to the next.
- Restructure the text to achieve progression and set desired expectations around one topic.

Solution.

- Author's intention and key points

The author first presents the cycle of propagation of the virus (point 1), because it is linked to prevention and control of the spread of the disease (point 2). Point 1 describes the human-mosquito-human transmission cycle with sentences (1), (3), and (4), and point 2 reveals the causes for the spread of the disease with sentence (2).

- Topic and progression scheme

There are at least two possible topics: the virus (repeated three times) and the mosquito (repeated four times). Either one can be the constant topic in all sentences. Regarding sentence order, since the paragraph (not shown here) that follows our sample paragraph describes what the community can do to stop the

propagation of dengue fever, the last sentence must be similar to sentence (2).

- Text restructure taking into account the reader's prior knowledge and anticipating what the reader may not know

Having monitored how thousands of scientists rewrite this paragraph, I would like to warn you about potential time-wasting pitfalls. Some of you may think that reordering the sentences to fix the lack of progression is enough. You clearly see that the second sentence is out-of-place and move sentence 2 to sentence 4. However, this creates a disconnection between part 1 of the paragraph (the transmission cycle) and part 2 (spread of the disease). Why? After moving sentence 2 to the end of the paragraph, sentence 4 now becomes sentence 3 which ends with "and resides within its salivary gland." Moving from the *salivary gland* of a mosquito to *the spread of the disease* is a conceptual jump too great for the reader. Therefore, do not describe the cycle using the sequence Mosquito infects Human — Human infects Mosquito, but instead choose the sequence Human infects Mosquito — Mosquito infects Human. That way, the salivary gland sentence will be the second sentence in the paragraph, in the middle of the sequence. In that position, it does not create any problem.

The sentence that ends the paragraph has three problems.

In addition, the disease spreads rapidly in densely populated areas because of the lack of effective mosquito control methods, the increase in air travel, and poor sanitation in areas with a shortage of water.

1. The bogus transition word "in addition" must go. But even when it is removed, the sentence starts with the wrong topic: the disease. Either you write an additional sentence with words extracted from the last sentence to act as glue between the two parts (cycle and spread), or you find in that last sentence old information that can be brought upfront and used as the sentence topic.
2. Examine the information given in that sentence. Is all of it useful? Remember that the next paragraph following the one you rewrite is about <u>community</u> action to stop the spread of the disease.
3. You are probably new to the topic of dengue fever. Yet, you know that mosquitoes need water to breed their larvae.

Knowing this, does everything in that sentence make sense? Is there a knowledge gap that needs to be bridged?

Final text (version one — Virus is the constant topic)

The dengue virus *from a human carrier enters the female Aedes mosquito via the infected human blood she feeds on.* **The virus** *then multiplies inside the mosquito's salivary gland over three to five days.* **It** *is transmitted back into another human through the saliva injected by the infected mosquito when she bites.* **The virus** <u>spreads</u> *rapidly in areas where large numbers of humans and mosquitoes cohabit.* <u>*This spread*</u> *is accelerated with human travel (air travel particularly), ineffective mosquito control methods, and poor sanitation in areas with water shortages.*

Progression is built around a constant topic. The progression is also a time-based progression (the transmission cycle) and a logical progression (amplification: from limited to extended, from specific to general). The fourth sentence acts as a bridge between the two parts. It is built by removing an element from the last sentence of the original version — '*densely populated areas*' — and developing it as a full sentence.

This paragraph is understandable by experts who have lived in tropical countries, but not by "the rest of us". Some say: I know that mosquitoes breed in water, so if there is a shortage of water, how do mosquitoes reproduce? The lack of knowledge leads to a lack of understanding even though the text is clear.

Final text (version two — Mosquito is the constant topic)

The female Aedes **mosquito** *feeds on the infected blood of a human carrier of the dengue virus. Inside the* **mosquito** *salivary gland, the virus multiplies over a period of three to five days. Subsequently, when the infected* **mosquito** *bites, its saliva carries the virus back into another human. Ineffective* **mosquito** *control methods cause the dengue virus to spread rapidly, particularly in densely populated areas at the end of the rainy season due to the sporadic rains. At that time,* **mosquitoes** *enjoy abundant breeding sites such as puddles of stagnant water and large opened vats storing rainwater for the dry season.*

In the second version, the mosquito is the constant topic in all sentences. The connection between the end of the transmission cycle and the spread of the disease is established simply by bringing upfront in the rewritten sentence an element — *'ineffective mosquito control'* — that was in the middle of the last sentence in the original version. Some information from the original sentence was removed, and extra details were added to make this paragraph clearer and to prepare the transition to the next paragraph on community action. The disappearance of *air travel* is intentional since air travel is not directly relevant to the villagers who are part of the community; and so is the removal of *'water shortages',* which is counterintuitive since mosquitos need water to breed. The additional information clarifies the link between the shortage of water and the spread of the disease.

Final text (version three — sequence progression and Cycle topic)

*The reproduction of the dengue virus relies on a three-step **cycle**. **First**, the virus enters the female Aedes mosquito when the mosquito feeds on the infected blood of a human host. Once inside the mosquito salivary gland, the virus **then** multiplies over a period of three to five days. The **cycle** is **complete when** the virus returns to another human through the saliva injected by the infected mosquito when it bites. This **cycle** is repeated rapidly (sometimes at epidemic speed) in densely populated areas when people travel, and when mosquito control is ineffective, particularly in areas with water shortages where mosquitoes breed in open air water storage containers.*

In this case, the progression is sequential: three steps. The reader is counting. When the sequential progression ends, the constant topic "Cycle" returns at the end of the paragraph.

Final text (version four — greater knowledge gap — children with some prior knowledge on mosquitos)

People sick with Dengue fever carry the dengue virus inside their blood. When that blood is sucked by a female mosquito called Aedes Aegypti — that's her family name — the blood carries the virus inside of her. It does not harm her, it simply invades her body. Fifteen days after it got in, it

arrives inside her saliva and stays there to multiply, and multiply, and multiply for 3 to 5 days. So when the female mosquito bites someone else, her saliva carries a whole army of dengue viruses into the blood, enough to make that person sick. Where many mosquitos and people live together, dengue fever spreads very fast. It spreads even faster when people travel by plane, or when people are careless and let the Aedes Aegypti mosquitos lay thousands of eggs on the side of open water containers like old tires or flower pot plates.

Children require less sophisticated vocabulary. Carrier, infected, densely populated, control method, are adult words. They need spoken language, not written language. They need intermediary explanations, examples, numbers, personification, and drama

One final word of caution: Do not attempt to "fix" progression problems in a paragraph without taking into account the topic of the next paragraph. Progression applies between paragraphs as much as it applies between the sentences of a paragraph. Progression problems are not always fixed by moving sentences around. In many cases, an unclear text needs complete restructuring. To restructure, understand the author's intention, and identify the key points of the argument presented and their underlying logical connections.

Take ten consecutive sentences from the discussion of your paper. Identify the topic and the stress of each sentence. Can you identify a progression scheme? Are some sentences totally disconnected from their predecessors? Are sentences artificially connected by a transition word hiding a progression problem? Rewrite these sentences to restore normal progression.

Chapter 8
Detect Sentence Fluidity Problems

Here is a very important exercise you don't want to miss because it allows you to evaluate how well you set expectations. Follow it step by step. It will only take you ten minutes.

Step 1:
Choose a paper you wrote, or if you have never written a paper, choose any paper in your field. Take a pencil. You will need it to underline words in each sentence, and to draw one of the smiley faces at the beginning of each sentence.
☺ ☹ ☹

Step 2:
Select the first eleven sentences from the introduction of the paper.

Step 3:
Set your eyes on the first sentence of the introduction.

Step 4:
Read the sentence and block eye access to the next sentence with a piece of paper. After you understand the sentence, underline the words that clue you on the topic of the next sentence. If you have an expectation, go directly to Step 6 (skip the next step).

Step 5:
You have no expectation for what could come next. Move to the right the paper hiding the next sentence until you get to the first verb of the sentence, and go no further. If you now have an expectation,

move to step 6, but if you still have no expectation, draw a ☺ at the beginning of the sentence you examined (not the one you partially uncovered). Go directly to Step 7 (skip the next step).

Step 6:
Test your expectation to see if you guessed right. Look at the complete sentence you were hiding. If it fulfills your expectations, draw a ☺ at the beginning of the sentence you examined (not the one you just read to check whether your expectation was right), but if it breaks your expectations, draw a ☹ instead.

Step 7:
If you have finished you should have a smiley at the beginning of the first 10 sentences; go to Step 8. If not, go back to step 4 and continue with the next sentence.

Step 8:
Compute your score. For each ☺ give yourself a pat on the back and add three points to your total score. For each ☺ subtract one point, and for each ☹ subtract two points. If the score is twenty or more, this writing is brilliant. If the score is less than twelve, consider rewriting the ☺ and ☹ sentences to better control the expectations of the reader.

Here is an example:

> *The sky was blue. (No idea, let me look up to the next verb)*
> *The sky was blue. The cat was (still no idea. I'll put ☺ in front of the first sentence)*
> ☺ *The sky was blue. The cat was trying to catch a bee on a flower. (I'll bet the cat will be stung by the bee!)*
> *The sky was blue. The cat was trying to catch a bee on a flower. The bee stung its paw. (Yeah! here is a ☺ I was right)*
> ☺ *The sky was blue. ☺ The cat was trying to catch a bee on a flower. The bee stung its paw. (The cat limped away)*
> ☺ *The sky was blue. ☺ The cat was trying to catch a bee on a flower. The bee stung its paw. The gardener dug a hole. (☹ Wrong!)*
> ☺ *The sky was blue. ☺ The cat was trying to catch a bee on a flower. ☹ The bee stung its paw. The gardener dug a hole.*
> *Total score: [−1, +3, −2] = 0!*

Now that you are finished with the exercise, wouldn't you like to find out the cause of these faces — in particular the neutral and sad faces? You could do it by yourself, going over each sentence to determine what created your expectations and how the writer violated them. You would then become aware of the causes of fluidity problems. In fact, why don't you do it now!

Many scientists did this exercise during the writing class. They identified the common problems. So, when you are done with the exercise, turn to the next heading to check whether the problems you found are also described here.

Reasons for No Expectations

Rear-mirror sentence

"The above framework is compliant with traditional digital signature system structures."[1]

In this sentence, the author adds one final detail that ends a rather complete presentation of the framework. The sentence looks backward, not forward. It sets no expectation.

Descriptive sentence

"The coordinates of each point in a uniform dataset are generated randomly in [0, 10000], whereas, for a Zipf dataset, the coordinates follow a Zipf distribution skewed towards 0 (with a skew coefficient 0.8)".[2]

Sentences that state facts or data create little expectation.

Paragraph break

By definition, the paragraph is a unit of text with a single theme or purpose. When the theme or purpose changes, the paragraph changes. But how are these two themes connected? Do you let readers find out by themselves? If you do, is the connection obvious even

---

[1] Qibin Sun and Shih-Fu Chang, A Robust and Secure Media Signature Scheme for JPEG Images, Journal of VLSI Signal Processing, Special Issue on MMSP2002, pp. 306–317, Vol. 41, No. 3, Nov., 2005.
[2] Papadias Dimitris, Tao Yufei, Lian Xiang, Xiao Xiaokui The VLDB Journal: the international journal on very large data bases, July 2007, v.16, no. 3, pp. 293–316.

to a non-expert? It is best to make that connection explicit in one of two ways: use the last sentence of the paragraph to introduce the theme of the next paragraph, or use the first sentence of the new paragraph to introduce the new theme and show how it relates to the previous paragraph.

In this next example, the connection is not explicit. The paragraph states that there are well-known algorithms and others not as well-known (we presume). But we do not know what to expect next.

> *"Many clustering algorithms are proposed to group co-expressed genes into clusters. Some well-known examples include hierarchical clustering [1, 2], K-mean clustering [3], and self-organizing maps [4, 5]. Bayesian networks [6] and graph theoretic approaches [7–9], model-based methods [10–12], and fuzzy clustering [13] provide additional methods toward clustering.*
>
> *Various measurements have been employed to score the similarities between pairs of gene expressions."*

Enumeration with no predictable item length

If each item of a numbered enumeration consists of one sentence, the reader very rapidly discovers the pattern and expects the situation to continue. But if each item is covered by a variable number of sentences, the reader does not know what to expect after the end of each sentence: a new item, or the next sentence on the same item.

> *In science, two factors contribute to long sentences: precision and intellectual honesty. Adding precision to a noun through a modifier lengthens the sentence. For example "an R400-7 fiber optic reflection probe with 6 illumination fibers and one read fiber" is significantly more precise than "a fiber optic probe". Such precision may be necessary to convince the reader of the quality of the fluorescence measurement. Intellectual honesty lengthens a sentence through the addition of detailed qualifiers and provisos.*

In this example, the reader expected the writer to move to the "intellectual honesty" factor, right after the "for example" sentence. But instead, the writer adds a sentence on the probable need for precision. The reader, whose expectations are now broken, resets expectations, but no longer knows what to expect after

the comment. This situation could have been avoided by removing the sentence. It is weak ('may be' expresses uncertainty), and non-essential.

Vortex of chained explanatory details creating an expectation void

It is a case of "one thing leads to another", but with no end in sight. The reader follows but does not know what to expect next. A good example has already been given. It is repeated here.

> *For the next three days, the thirty embryonic cells proliferate in the culture **dish**. The dish, made of plastic, has its inner surface **coated** with mouse cells that through treatment have lost the ability to divide, but not their ability to provide nutrients. The reason for such a special **coating** is to provide an adhesive surface for the embryonic cells. After proliferation, the embryonic cells are collected and put into new culture dishes, a process called "replating".*

The first sentence establishes cell proliferation as the paragraph topic. The next two sentences take us down a series of chained details: from embryonic cells to culture dish, from culture dish to dish coating, and from dish coating to adhesive surface. The reader does not know what to expect next. The return to the proliferation topic is not surprising, but it is not expected.

Vague, general statement

> *With the rapid development of computer communication and Internet, the distribution of digital images is pervasive.*

The next sentence is about watermarking. Who would have guessed! The statement is general and typical of the first vacuous sentences of an introduction. From such a vague statement, no expectations can be derived.

Long sentence diluting expectations across many topics

> *"In this paper, we focus on extending the implementation of the Run Time Infrastructure (RTI) to relax the time synchronization among federates, particularly focusing on RTIs that support the conservative simulation protocol for their time management service, for example the DMSO's*

RTI. Each federate regulates the time of the federation with a look-ahead value."[3]

What is the second sentence to be about: RTI, federates, relaxing time synchronization, the conservative simulation protocol, the time management service of the DMSO's RTI? The reader has no clue what to expect because there are too many options in this long sentence filled with details. The second sentence is about the method used by each federate to regulate time. Who would have guessed? I thought that the writer would explain why there was a need to relax time synchronization. I would have given that sentence a ☹ but the experts in the field have given it a ☺ (**Expectations vary from one reader to the next, depending on their background knowledge.**)

Reasons for Betrayed Expectations

Unannounced topic change and ambiguous pronoun

The images in our set are normalized to have a consistent gray level when output. The coordinates of the eyes are automatically detected and set to a fixed position. They are then resampled to a given size. After normalization…

The topic of the first sentence suggests two possible candidates for the next sentence:

1. The benefit of having a consistent gray level
2. Additional normalization step on the images, such as image contrast

The "eye" topic of sentence 2 does not fit expectation 1 or 2. It could have, had the writer specified that the images were face images. Only after readers finish reading sentence 2, do they realize that it ties in with candidate 2 — the additional normalization step. Based on this, readers expect two new candidates for the next sentence:

1. Yet another normalization step

---

[3] Boon Ping Chan, Junhu Wei, Xiaoguang Wang, (2003) Synchron, Synchronization and management of shared state in hla-based distributed simulation, Proceedings of the Winter Simulation Conference,S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

2. A reason for fixing the eye coordinates, or the location of the fixed position

The pronoun '*they*' starts the sentence. Readers think the pronoun refers to '*the coordinates of the eyes*', and they expect a reason for fixing these coordinates. But they reach an impasse when the verb "resampled" arrives because coordinates are unlikely to be resampled. Readers briefly stop and re-read to discover that '*they*' refers to the images mentioned in sentence 1. The pronoun was ambiguous. The third sentence was simply another normalization step (expectation 1).

Here is a possible rewrite.

All face images are normalized (1) to display consistent gray levels, (2) to automatically place the eyes at the same location in each image, and finally (3) to set each image to a 128 × 128 size through resampling. After normalization, …

Text precision is increased. The sentence is long, but precise and easy to follow. The enumeration and the parallel syntax set the expectations that we are moving from one normalization step to the next.

Adjectival or adverbial claim not followed by evidence

Adjectives or adverbs are words that set expectations in three ways:

- They make a claim that requires validation, as in "*The world is flat*".
- They state an existing (often well-known) situation judged unsatisfactory that requires correction, as is "*Airplane engine maintenance is very labor-intensive*". In this case, *very labor-intensive* sets the expectation for its opposite, less labor-intensive. Note the role of very as a judgmental word to enhance the expectation for a more satisfactory solution.
- They state a situation in a negative or a pejorative light that requires the need for a change towards the positive or the ameliorative as in "*trapping is unimportant at high temperatures.*" The sentence sets the expectations that trapping may be important at low temperatures.

> *Drop test simulations that rely on a fine mesh of 3D finite elements are very **time-consuming**. They are required to save the time and cost involved in actual physical tests conducted at board level. We propose a simplified model that considers the board as a beam structure…*

The adjective *time-consuming* makes a claim associated to the Finite Element Method. Readers in that field are already aware that fine mesh 3D modeling is very time-consuming. They expect the writer to present faster methods in the next sentence. It comes… in the third sentence. You may object that the writer uses a good progression scheme, the constant topic progression: '*drop test simulations*', and the pronoun '*They,*' although the pronoun is ambiguous because it could also replace *finite elements*. To answer the objection, it is worth emphasizing that **expectations import more than progression**. To meet the expectations of the reader, these sentences must be rewritten. Here is a possible rewrite.

> *Simulations based on the analysis of a fine 3D mesh of finite elements save the cost, but not the time involved in conducting physical board level drop tests. Saving time is possible if the model is simplified. We propose a model that simplifies the board by making it a beam structure…*

In this rewrite, chain progression is set, and expectations are met.

Unclear answers to clear questions

> *Beyond which threshold value should the degree of asymmetry of the brain be considered abnormal? The answer depends on what kind of brain information is required. When one studies pathological abnormality, false positives and false negatives should be kept low. The threshold value differentiating normal from abnormal asymmetry could be estimated from patient data, but how sensitive or how specific that value would be, is unclear.*

The starting question sets an expectation for a precise numerical value. The *it-depends* answer is unclear and disappoints. Indeed, the second sentence acts as a second indirect question: What is the brain information required? Again, the answer disappoints, because the writer gives an unexpected answer that reveals a concern for the statistical quality of the data. The reader could not possibly have guessed that, and is therefore surprised by the false positives/false

negatives sentence. At this point, the reader usually gives up and, having been surprised twice in a row, takes a neutral attitude. The writer then tells the reader that a threshold value is of little use if it is not sensitive or specific. Here is a possible rewrite:

> *Assuming that the degree of brain asymmetry is a reliable indicator of brain pathology, can a diagnostic of abnormality be based on an asymmetry threshold value? To answer that question, it will be necessary to determine whether such a value derived from patient data has enough sensitivity and specificity.*

Unjustified choice

> *"The methods for cursor control [Brain Computer Interface] come under two categories, regression [references] and classification [more references]. Each of them has its merits. In our study we adopt the classification method,..."*[4] *(the text that follows describes the method, but does not give the reason for the choice).*

The writer presents the choice he faced: regression or classification. Both have their own advantages. The reader has two possible expectations:

- The writer will give the advantages and disadvantages of each. Obviously, the writer did not intend to do that. It would be better to remove the sentence "*Each of them has its merits*" because it creates the wrong expectations.
- The writer will choose one category of methods and justify it.

Which disadvantage, or which advantage led the writer to prefer the classification method? The writer does not say. Note that even if the second sentence "*Each of them has its merits*" is removed, the reader will still want to know why the writer chose the classification methods over the regression methods. Whenever you announce and make a choice, the reader wants to know why, even if the choice is arbitrary.

---

[4] Zhu X, Guan C, Wu J, Cheng Y, Wang Y. (2005) Bayesian Method for Continuous Cursor Control in EEG-Based Brain-Computer Interface. *Conf Proc IEEE Eng Med Biol Soc* **7**: 7052–7055.

The methods for cursor control [Brain Computer Interface] come under two categories, regression [references] and classification [more references]. We decided to adopt the classification method because…

Broken repeated patterns and paraphrase

The emergence of Hidden Markov Models (HMM) in the 1970's allowed tremendous progress in speech recognition. HMMs are still the de-facto method for speech recognition today. However, some argue that HMMs are not a panacea. At the same time, today's speech recognition systems are far smarter than those in the earlier days.

The first sentence establishes the paragraph topic: HMMs. It places it in a good light. The reader expects the praise to continue, and it does in sentence 2. The reader is given two expectations for sentence 3: (a) examples of the current use of HMMs in speech recognition, (b) or the reasons for the continued use of HMMs (maybe recent improvements to the basic HMM method).

In sentence 3, the writer surprises the reader by announcing the limitations. Expectations reset, the reader now expects the writer to name a problem where HMMs do not excel. Unfortunately, in sentence 4, the writer seems to change topic. "*At the same time*" is a phrase similar to "*in addition*": it masks the lack of a proper transition.

When breaking a pattern, it is better to announce the break in the same sentence as shown below.

Although HMMs are still the de-facto method for speech recognition today, they are not a panacea. As speech recognition applications increase in sophistication (automatic language recognition, speaker authentication), HMMs need to blend harmoniously with other statistical methods, in real-time.

Lack of knowledge (or unexplained word) and synonyms

Blood vessels carry blood cells as well as platelets, which are not as numerous as blood cells (ratio of 1 to 20). Platelets, besides being a source of growth factors, are also directly involved in homeostasis by aggregating to form a platelet plug that stops the bleeding. A thrombocyte count is usually included in a blood test as it is a means to determine diseases such as leukemia.

The third sentence blocks the non-expert reader who does not know that thrombocyte is synonymous with platelet. In this case, the writer

did not wish to repeat platelet because it had already been mentioned twice in the previous sentence. Filling a paper with synonyms increases the knowledge gap and increases the memory load. As a rule, settle on one keyword (the simpler one) when two synonyms are available, and use it consistently in your paper.

> *A platelet count is usually included in a blood test as it is a means to determine diseases such as leukemia.*

But if you wish to introduce a new keyword, define it in a just-in-time fashion as shown below.

> *A platelet count (also known as a thrombocyte count) is usually included in a blood test as it is a means to determine diseases such as leukemia.*

This page intentionally left blank

Chapter 9
Control Reading Energy Consumption

> ### Réponse hémodynamique
>
> *The article by Peter Hagoort that Michael had given me to read was truly fascinating. It described what happens in our brain when, during reading, it encounters strange things such as "the car stopped at the casserole traffic light". Something similar happened to me when I stumbled over the word 'hemodynamic' in the article. Google took me to the website fr.wikipedia.org/wiki/Réponse_hémodynamique, and things became very interesting. I discovered that when reading becomes difficult, the body sends a little more blood (i.e. glucose and oxygen) to the brain. It does not take blood from one part of the brain to send it to another part so as to keep energy consumption constant; it simply increases the flow rate. Following the trail like a bloodhound, I discovered a French article written by André Syrota, director of the life science division at the Atomic Energy Commission, indicating that our brain's additional work could consume the equivalent of "147 joules per minute of thought."*[1]

How tired will your readers be at the end of their reading journey? How well did you manage their time and energy? As George Gopen points out, reading consumes energy.[2] Reading scientific articles consumes A LOT MORE ENERGY. Therefore, how do you reduce the reading energy bill, and how do you give your reader the assurance that plenty of energy refueling stations will be available along the long and winding road of your text?

---

[1] http://histsciences.univ-paris1.fr/i-corpus-evenement/FabriquedelaPensee/affiche-III-8.php

[2] George D. Gopen. 2004. "Expectations: teaching writing from the reader's perspective". Pearson Longman, p. 10.

The Energy Bill

Let, E_T, be the total energy required by the brain to process one sentence. E_T is the sum of two energies:

1) The syntactic energy E_{SYN} spent on analyzing sentence structure,
2) The semantic energy E_{SEM} spent on connecting the sentence to the others that came before it, and on making sense of the sentence based on the meaning of its words.

Gopen considers these two energies to be in a "zero-sum relationship".[3] This means that if E_{SYN} becomes large, it will be at the expense of E_{SEM}: the more energy is spent on the analysis of the syntax of a sentence, the less energy will be left to understand the meaning of the sentence.

$$E_T = E_{SYN} + E_{SEM}$$

E_T is finite and allocated by the brain to the reading task. Similarly to our lungs, which give us the oxygen we need one breath at a time, the brain has enough energy to read one sentence at a time. E_T cannot increase beyond a certain limit fixed by physiological mechanisms: to increase the blood flow rate takes a few seconds and the size of the blood vessels in the brain (although extensible) is limited.

Your job, as writer, is to make sure that $E_{SYN} + E_{SEM} < E_T$ at all times by minimizing both the syntactic and the semantic energy required to read.

What would consume excessive syntactic energy, E_{SYN}?

- Anything ambiguous or unclear — a pronoun referring to an unclear noun, a convoluted modified noun, an ambiguous preposition
- Spelling or light grammar mistakes — missing definite article *the*, wrong preposition such as *in* instead of *on*, misused verb such as *adopt* instead of *adapt*

---

[3] George D. Gopen. 2004. "Expectations: teaching writing from the reader's perspective". Pearson Longman, p. 11.

- Incomplete sentences, i.e. missing verb
- Anything taxing on the memory — long sentences (usually written in the passive voice), formulas, sentences with multiple caveats, provisos and qualifiers, sentences with deeply nested subordinates
- Grammatical structures from a foreign language applied to English without change
- Missing or erroneous punctuation

What would consume little syntactic energy, E_{SYN}?

- Short sentences with simple syntax: subject, verb, object

    New ideas disrupt the logical flow of sentences.

- Sentences with a predictable pattern established with words such as *although*, *because*, *however*, or *the more…the less*.

    The more energy is spent to analyze the syntax of a sentence, the less energy is left to understand what the sentence means.

- Sentences with subject close to verb, and verb close to object

    Motivation allocates the total energy, ET, to the reading task.

- Sentences with good punctuation

    The reader has three choices: give up reading, read the same sentence again, or read what comes next.

What would consume great semantic energy, E_{SEM}?

- Unknown words, acronyms, and abbreviations
- Absence of context to derive meaning
- Lack of prior knowledge to understand, or to aid understanding
- Lack of examples or visuals to make the concept clear
- Overly detailed or incomplete visuals
- Reader forgetful of content previously read
- Reader in disagreement with statement, method, or result
- Very abstract sentences (formulas)
- Sentences out of sync with reader expectations

What would consume little semantic energy, E_{SEM}?

- Sentence with a well-established context

  Total reading energy for a given sentence, E_T, is the sum of two elements: the syntactic energy, E_{SYN}, spent on analyzing its syntax, and the semantic energy, E_{SEM}, spent making sense of the just analyzed sentenced.

$$E_T = E_{SYN} + E_{SEM}$$

- Reader familiar with the topic or the idea

  The songbird flew back to the nest to sit on three little eggs, two of its own, the third one from a cuckoo.

- Sentence that explains the previous sentence

  Therefore if E_{SYN} becomes large, it will be at the expense of E_{SEM}. The more energy is spent to analyze the syntax of a sentence, the less energy will be left to understand the meaning of the sentence.

- Sentence that prepares the ground (through progression or setting of context)

  Sub-clauses that pull reading forward often follow a predictable pattern; they start with a preposition such as "although," "because," "however," or "if".

- Short sentences (with known vocabulary)

  It does not. The reader is surprised.

What would get the reader into trouble?

Energy shortages occur when $E_{SYN} + E_{SEM} > E_T$

- E_{SYN} is unexpectedly large. As a result, what remains of E_{SEM} is insufficient to extract the complete meaning of the sentence.
- E_{SYN} is normal; but a new word, acronym, abbreviation, apparent contradiction, or concept requires additional brain effort (saturated memory, or failure to find associative link with known data). The reader runs out of E_{SEM}. The semantic energy gas tank is empty before the sentence is fully understood.

When this happens, the reader can make one of three choices: give up reading, read the same sentence again, or read what comes next, hoping to understand later.

Giving up reading is tragic. It is a consequence of repetitive and successive breakdowns in understanding. The text becomes increasingly obscure, and the reader finally gives up reading.

Reading the same sentence again happens if motivation is high. The reader is determined to understand because much understanding is expected from the difficult sentence. Rereading after mastering a difficult syntax consumes no syntactic energy because the sentence syntax is now familiar, and the reader spends all of his or her energy on understanding only.

$$E_{SYN} = 0, \text{ and therefore } E_T = E_{SEM}.$$

The simile — reading as consuming brain energy — is in line with what Science observes. The brain hard at work consumes more energy.

Punctuation: An Energy Refueling Station

The period, a full-stop to refuel

When the period arrives, the reader pauses and refills his or her energy tank before reading the next sentence. It gives the reader a chance to conclude, absorb, consolidate the knowledge just acquired and anticipate what comes next (from expectations or progression).

The semicolon, a fuel stop for topping a half-full tank

Surprisingly, searching for a semicolon in a scientific paper is often rewarded by the infamous beep that says: "None found, can I search for anything else?" Periods, colons, and commas seem to be the only punctuation used by scientists. Scientists, by nature logical, should be fond of semicolons, not only to strengthen their arguments, but also to make their text less ambiguous and to carry the context forward at little cost.

Semicolons are close cousins to the period. They stand at a place of semantic closure, like the period. They end and start a clause (part

of a sentence with a subject and verb). Unlike the period, their role is to unite, join, or relate; while the role of the period is to separate. The clauses on each side of a semicolon are often compared, contrasted, or opposed. Often, the first clause in the sentence makes a point, and the clause (or clauses) after the semicolon refines, details, or completes the point. Semicolons are often found close to conjunctions and conjunctive adverbs, such as '*but*,' '*consequently*,' '*however*,' '*therefore*,' '*thus*,' or '*nonetheless*'.

> *The calculated data and the observed data were closely related; however, the observed data lagged when concentration dropped.*

The first clause in the above sentence provides context to the whole sentence. It shares its subject with the clause that follows the semicolon. These two clauses are also closely related semantically, much more so than two sentences separated by a period. Therefore, since the context does not vary within the sentence, reading is faster and easier.

A semicolon has more than one use. When a sentence needs to be long to keep together a list of sentences, the semicolon does the job magnificently.

> *Information with visual impact requires creativity, graphic skill, and time. Because most of these are in short supply, software producers provide creativity, skill, and time-saving tools: statistical packages that crank out tables, graphs and cheesy charts in a few mouse clicks; digital cameras that in one click capture poorly lit photos of test bench equipment replete with noodle wires (I suppose the more awful they look, the more authentic they are); and screen capture programs that effortlessly lasso your workstation screens to corral them for your paper.*

The three ":" "!" "?" fuel stops and the comma

Other punctuation marks provide an opportunity to refuel: the colon, the question mark, and the unscientific exclamation mark (I wonder if Archimedes would have damaged his reputation as a scientist had he ended his "Eureka" statement with an exclamation mark). The question mark, the most underused punctuation mark, causes the reader to pause and think while also setting clear expectation for an answer. The colon introduces, explains, elaborates, recaps, and lists. Unlike the semicolon, it can be followed by a phrase that lacks

a verb. Like the semicolon, it is preceded by a whole clause (not one truncated as in the next example).

And the results are:

In a correct sentence, the main clause is not truncated.

And the results are the following:

Colons are much liked by readers: they announce clarification or detail. Colons are also the allies of writers. They help introduce justi-fication after a statement.

Commas help to disambiguate meaning, pause for effect, or mark the start and end of clauses. But, for all their qualities, there is one that commas cannot claim: semantic closure. Readers cannot stop at a comma and decide that the rest of the sentence can be understood without reading further.

In this chapter, you have been given many tools to reduce the reading energy bill. Imagine your writing as a piece of cloth, and the brain of the reader as an iron. If your writing has the smoothness of silk, the iron can be set to the lowest temperature setting. If it has the roughness of overly dry cotton, not only will the iron have to be set to the highest temperature setting, but you will also put the reader under pressure and demand steam to iron out the ugly creases in your prose. Either **you** spend the time and energy, or **the reader** does.

Ask a reader to read your paper and to highlight in red the sen-tences not clearly understood (high semantic energy), and in yellow, the sentences that slowed down reading because of a dif-ficult syntax (high syntactic energy). Then correct accordingly.

Some punctuation helps reduce reading energy. Search for ':' and ';'. Do you have enough of them? If not, look for opportuni-ties to use them, particularly in long sentences.

This page intentionally left blank

Part 2
Paper Structure and Purpose

Each stage in the construction of a house contributes to its overall quality. Similarly, each part in an article contributes to the quality of the whole, from the abstract (the architect blueprint) and the structure (the foundations), to the introduction (the flight of steps and the porch), the visuals (the light-providing windows), and finally the conclusions (the house key).

The art of construction is acquired through a long apprenticeship. You may be attracted by the time-saving expedient prefab or by the imitation of other constructions of uncertain architectural quality (the paper your supervisor handed out to you as a template for your first paper). Beware of shortcuts. A hastily assembled paper, once analyzed, often reveals major cracks and faults: its shapeless structure is like baggy jeans that would fit just about any frame; its graphics and other visuals have a mass-produced look and feel.

When your house is built, when your paper is finally published, how will the readers feel after visiting your house? In your lobby, you have left a guest book, in which your visitors can enter their remarks. Here are two entries. Which one would please you the most? And what kind of scientific paper would you associate such remarks with?

> To Whom It May Concern,
>
> *Concrete slabs and parched grass don't make much of a garden. I really don't know how you dare call that plywood platform, a porch. It is laughable.*
>
> *The inside of the house is dark and gloomy. You need more windows. And by the way, I did not take anything from the fridge. There was nothing in it, and it wasn't even plugged in. I could not find the things you claimed were in the house. I looked everywhere, but everything was in such a mess, I gave up. There were large cracks in some walls (not line cracks). Your house is not safe. And while we are on the subject of walls, why did you paint them all a dull gray? Anyway, I'm out of here. Thanks for the spare key, but I left it behind. I don't intend to come back.*

> Dear House Owner,
>
> *Your house is such a delight. I wished I had one like that. Your flower garden and your large porch are so inviting. Your indoor is exquisitely designed. Everything is at the right place. I immediately found what I was looking for. But I did not expect the basket of fresh fruit and the cool complementary drinks in the fridge. How lovely! I love your large bay windows. They make the house look so bright and colorful. There is not one dark corner!*
>
> *Thank you for your spare key. I also took your business card to advertise this lovely house of yours, which makes people feel so welcome.*
>
> *With deepest thanks,*
> *Professor Higgins*

What makes a good house depends on its design, of course, but also on the attitude of the house owner towards his guests. ==A good writer is one who cares for his readers; someone who anticipates their needs. Good writing is not just a matter of writing skills; it is also a matter of attitude towards the readers.==

To construct a satisfactory paper, one must understand the role each part plays for the reader and the writer. And to assess the quality of each part, one must establish evaluation criteria. The next chapters

fulfill these objectives. Numerous examples are given to analyze and to help you distinguish good writing from bad writing. By the end of the book, you will be ready to write your paper.

First Impression

Today, as the city's bowels demonstrate their usual constipation, the pouring rain adds a somewhat slimy aspect to the slow procession of traffic. Professor Leontief does not like arriving late at the lab. He hangs his dripping umbrella over the edge of his desk, at its designated spot above the trashcan, and he gently awakens his sleepy computer with some soothing words "Come on, you hunk of metal and silicon oxide, wake up."

He checks his electronic mail. The third e-mail is from a scientific journal where he helps out as a reviewer. "Dear Professor Leontief, last month you kindly accepted to review the following paper and submit your comments by….". He looks at his calendar and realizes that the deadline is only 2 days away. A cold chill runs up his spine. He hasn't even started. So much to do with so little time! Yet, he cannot postpone his response. Being a resourceful man, he makes a couple of phone calls and reorganizes his work schedule so as to free an immediately available 2-hour slot.

He pours himself a large mug of coffee and extracts the article from the pile of documents pending attention. He goes straight to the reference section on the last page to see if his own articles are mentioned. He grins with pleasure. As he counts the pages, he looks at text density. It shouldn't take too long. He smiles again. He then returns to the first page to read the abstract. Once read, he flips the pages forward slowly, taking the time to analyze a few visuals, and then jumps to the conclusions that he reads with great care.

He stretches his shoulders and takes a glance at his watch. Twenty minutes have passed since he started reading. By now, he has built a more or less definitive first impression on the paper. Even though the article is of moderate length, it is too long for such a small contribution. A letter would have been a more appropriate format than a full-fledged paper. He will have to inform the writer, using diplomatic skills not to be discouraging for he knows the hopes and expectations all writers share. What a shame, he thinks. Had he accepted the paper, his citation count would have increased. Now the hard work of thorough analysis lies ahead. He picks up his coffee mug and takes a large gulp.

The reviewer is a busy person, definitely time-to-result driven. A survey I regularly conduct shows that reviewers take on average 20 minutes to get a "publish or perish" impression of the paper. Naturally, some get their first impression in a few minutes and some read everything before they form an opinion.

The survey also shows that reviewers only read certain parts of the paper to derive that impression: the title, the abstract, the introduction, the conclusion, and the structure, which they discover by turning the pages and reading the headings and subheadings. They also look at some visuals and their caption. Forming a first impression on partial cursory reading is an efficient way to save time on the whole review process. Once that impression is derived, the task that follows is narrowed down to the justification of that impression:

1) If it's a good impression (**publish**) then look for all the supportive evidence that strengthens your first impression, and minimize any problem encountered by considering it minor;
2) If it's a negative impression (**perish**) then look for all the good reasons to postpone publication: identify drawbacks, inadequate data, logical errors, methodology inconsistencies, etc.

The saying is right: "First impressions count!" There is much supportive evidence for it in cognitive psychology. Enter the keywords "Halo effect" or "Confirmation bias" in your search engine, and be ready to be amazed!

In part II, the book selectively covers the parts of a paper that are read during the rapid time where the first impression is formed. This choice has been guided by the comments I received from scientists who have published papers. They stated that the methodology and results sections of their paper were the easiest and the fastest to write. It was the other parts that were difficult and took time: the abstract, the introduction, and the conclusions. As to the title, the structure, and the visuals, they recognized at the end of the course, that they had underestimated the key role these parts played in creating the first impression.

After reading these chapters and doing the exercises, your writing should have significantly improved. At that time, the difference between making ripples and making waves in the scientific community is less a matter of writing than it is a matter of scientific excellence, which I leave in your capable hands!

Chapter 10
Title: The Face of Your Paper

When I think about the title of a paper, quite naturally, the metaphor of a face comes to mind. So many features of a title resemble those of a face. *From your face, people get a first impression of you.* Likewise, a title contains the first words the readers will see. It gives them a first impression of how well your paper satisfies their needs and whether your paper is worth reading or not. *Your face sets expectations as to the type of person you are.* Your title also reveals what kind of paper you have written, its genre, its breadth, and its depth. *Your face is unique and memorable. It is found on your passport and various official documents.* Your title is unique. It is found in references and databases. *What makes your face unique is the way its features are harmoniously assembled.* What makes your title unique is the way its keywords are assembled to differentiate your work from the work of others.

When I was 12 years old, I stumbled upon a strange book on the shelves of my local library. It was about morphopsychology, the study of people's characters as revealed by the shape of their faces. It was fun trying to associate a face with a character. Discovering a paper from its title should also be fun. To include you in the fun, I have turned the following section into a dialogue. Imagine yourself as the scientist being asked the questions. How would you answer?

Six Titles to Learn About Titles

Author: Greetings Mr. Scientist. I'd like to introduce a series of six titles and ask you one or two questions about each one. These titles may be in areas you are not familiar with, but I'm sure you'll do fine. Are you ready?

Scientist: *By all means*, *go ahead*!

Author: All right then. Here is the first title.

   "Gas-Assisted Powder Injection Molding (GAPIM)"[1]

Based on its title, is this paper specific or general?

Scientist: *Hmm, you are right, I know nothing about powder injection molding. The title seems halfway between being specific and being general. "Powder Injection Molding" by itself would be general, may be a review paper. But, this title is a little more specific. It says "Gas-assisted", which tells me there are other ways to do powder injection molding.*

Author: You are right. GAPIM is used to make hollow ceramic parts. People in that field would be quite familiar with powder injection molding and its PIM acronym. What would have made the title more specific?

Scientist: *The author could have mentioned a new specific application for GAPIM.*

Author: Good, How do you feel about the use of the GAPIM acronym in the title?

Scientist: *I am not sure it is necessary. I have seen acronyms in titles before, but they were used to launch a name for a new system, a new tool, or a new database. The acronym was usually more memorable than the long name it replaced. Unless GAPIM is so well known, people have it memorized and it has become a search keyword, I don't think it should be in the title.*

Author: Thank you. How about this second title: general or specific?

   "Energy-Efficient Data Gathering in Large Wireless Sensor Networks"[2]

Scientist: *This title is very specific and its scope is well-defined: it is not sensor networks, it is wireless sensor networks, more precisely large*

---

[1] Dr. Li Qingfa, Dr. Keith William, Dr. Ian Ernest Pinwill, Dr. Choy Chee Mun and Ms. Zhang Suxia, Gas-Assisted Powder Injection Moulding (GAPIM), ICMAT 2001, international conference on materials for advanced technologies, Symposium C Novel and Advanced Ceramic materials, July 2001.

[2] KeZhing Lu, LiuSheng Huang, YingYu Wan, HongLi Xu, Energy-efficient data gathering in large wireless sensor network, second International Conference on embedded software and systems (ICESS'05), Dec 2005, pp. 327–331.

wireless sensor networks. And the paper is only looking at data gathering in these networks. Its contribution, "energy efficient" is placed where it should be, right at the beginning of the title. "Energy efficient" gives me a hint that data gathering is not energy efficient in large networks. By the way, I am wondering whether "large" is the correct descriptor here, maybe "sparsely populated" would be better.

Author: You are perfectly entitled to logically infer that from the title. Actually, all readers generate hypothesis and expectations from titles. How about these two titles: are they both claiming the same thing?

> *"Highly efficient waveguide grating couplers using Silicon-on-Insulator"*
> *"Silicon-on-Insulator for high output waveguide grating couplers"*

Scientist: *The second title seems to introduce a new technology — Silicon-on-Insulator — to make waveguide grating couplers of the high output kind. Come to think of it, I'm not clear as to what "high output" means. It may be a type of coupler but it may also be a benefit, meaning that other technologies can only deliver lower output. You can tell I'm not an expert in this field either! Now let's look at the first title. What comes first in the title is usually the author's contribution, so this paper seems to be more concerned about making the whole system more efficient, using existing silicon-on-insulator technology. In my opinion, these are the titles of two different papers. The first paper published was the one introducing Silicon-on-insulator.*

Author: Bravo! You are doing fine. Now look at the following titles. Besides the use of an em dash or a colon to introduce the benefit of Web Services, these two titles are equivalent: which one do you prefer?

> *"Web services — an enabling technology for trading partners community virtual integration"*[3]
> *"Web services: integrating virtual communities of trading partners"*

Scientist: *Um… this is a difficult one. The long five word modified noun in the first title is difficult to read, yet I am attracted by the catchy term 'enabling Technology' even though it isn't a search keyword. The second*

---

[3] Siew Poh Lee, Han Boon Lee, Eng Wah Lee: Web Services — An Enabling Technology for Trading Partners Community Virtual Integration. ICEB 2004: pp. 727–731.

title is easy to read. It is shorter, clearer, more dynamic, and purposeful. But is it necessary to put a colon after "Web Services"? The second part of the title does not really explain or illustrate Web services. Could the title be changed to "Integrating virtual communities of trading partners through Web Services"? In this way, what is new comes at the beginning of the title. I don't think that web services are really new.

Author: The title could be changed to what you propose. You are right; the second title is more dynamic. The use of the verbal form "Integrating" makes it so. Many papers have a two part title separated by an em dash or a colon. The net effect of that separation is to create two places of emphasis in the title: before and right after the punctuation. Otherwise, only the first part of the title is emphasized. Now regarding your remark about the novelty of web services, if web services were indeed new, you could have a two part title. You are doing very well. Only two more titles. Is the following title for a British or an American journal?

"Vapor pressure assisted void growth and cracking of polymeric films and interfaces"[4]

Scientist: *Vapor with an "o". It is for an American journal, isn't it? If it had been for a British paper, they would have written "vapour". One has to be careful with keyword spelling nowadays, even if search engines are getting better. Fortunately, the title contains many keywords, so it would be easy to find.*

Author: Actually, in this field there is an alternative keyword frequently used also, but by the British. Instead of "vapor pressure-assisted", they use "moisture-induced". How would you make sure that both British and American scientists find your paper?

Scientist: *Hum,… let me think. Maybe I would use one keyword in the title, and the other one in the abstract or in the list of keywords. Since the search is often done on title and abstract, the paper has a chance to be found by both communities. Actually, now that I've read the title again, I find it somewhat ambiguous.*

---

[4] L.Cheng, T.F. Guo, 2003. Vapor pressure assisted void growth and cracking of polymeric films and interfaces, *Interface Science*, **11**(3): 277.

Author: In what way?

Scientist: *This title contains two "and"s. Are there two contributions in this paper:* **vapor-pressure assisted void growth** *being one, and* **cracking of polymeric films and interfaces** *being the other, or is there only one:* **vapour assists both** *void growth and cracking? The second "and" is just as ambiguous: does the adjective "polymeric" apply to films and interfaces, or only to films? I am sure an expert would find the title unambiguous, but non-experts like myself lack the knowledge to disambiguate.*

Author: Excellent observation. Titles have to be clear to all, experts and non-experts. Besides *and* and *or*, other prepositions can also be quite ambiguous in titles. For example, the preposition *with* could mean *and* as in "coffee with milk", or it could mean *using* as in "stir the coffee with a spoon".

The time has come for our last title. It is somewhat tricky. Can you identify the author's contribution?

"A new approach to blind multi-user detection based on inter symbol correlation"

Scientist: *Other researchers are already doing research in this field, and the author is following the pack with a new approach. Personally, I don't like the word 'approach' : it is vague. I would use 'method,' 'technique,' 'system,' 'algorithm,' or 'technology' instead. They are more specific. And I also don't like titles that start with 'a new' something. 'New' does not specify what is new or what makes it new. Regarding the contribution of this paper, I must say I am at a loss. The inter symbol correlation could be new, but if that is the case, why is it at the back of the title? It should be upfront. "Inter Symbol correlation for blind multi-user detection" is clear. It may also be - and I suspect this is the case - that inter symbol correlation is not new, but the author modified the method. That would explain the use of 'based on'. In that case, why doesn't he tell us the benefit of the modified method? Something like… "Modifying Inter Symbol Correlation to increase accuracy of blind multi-user detection"… It would be more informative and more compelling.*

Author: You are quite good at this. Thank you so much for assisting me in this dialog.

Scientist: *Not at all!*

Less time than you think

Vladimir often liked to Google his one published paper to see who cited it, or what other people wrote about it. But, that day, instead of typing the whole title, he typed only two of its main keywords, thinking it would be enough. Aghast, he looked at the first page. It was filled with other people's titles, and there were another ten pages following that one. He started scanning down the list of titles, barely spending a couple of seconds on each one. And then the thought hit him. That was what other researchers did when searching for interesting papers!

"This is terrible," he shouted.

Ruslana heard him. She asked "What is terrible, darling?"

"Good grief! I spent 9 months doing research, 2 full weeks writing my paper, but if its title does not catch the eyeballs of googling readers, my paper won't even be read, and I can kiss goodbye to citations and career growth!"

Ruslana helped as helpless people usually do, by stating the obvious.

"Well, you'd better write an attractive title then!"

Vladimir responded:

"Thanks a lot, and how am I supposed to make a title attractive, Mrs Toldoff?"

She escaped by jesting:

"Try lipstick, Darling."

Actually, Ruslana made an interesting point. The title is the face of your paper. But how do you recognize a face amongst hundreds?

- First, look at a face, face on, not from the back (see technique 1 below).
- Smile. A smiling face is more memorable than a bland one. It is engaging, alive, dynamic (see techniques 2 and 3).
- Do not hide the salient characteristic features by which people recognize your face behind scarves, glasses, or veils (see techniques 4 and 5).
- Find a way to enhance the attractiveness and uniqueness of the face through make-up, special hairstyle, mustache, etc (see technique 6).

But whatever you do, don't be an impostor and use facial cosmetic surgery to turn into another Michael or Elvis. Each face is unique, and it should remain so.

Why don't you identify how these solutions also apply to the crafting of a great title? You may even have your own solutions. Here are six techniques that I have found effective.

Six Techniques for Improving Titles

1-Placement of contribution upfront in a title

As seen in the six titles reviewed so far, readers expect to see the contribution of a paper at the same place in each title: its beginning. The contribution is followed by the scope, the method, or the application (as in the example hereafter)

> *Self-installing substation for insular tidal-energy farms*

The majority of the titles in print are incomplete sentences lacking a conjugated verb. In certain fields however (life sciences being one), the title is a full sentence with a conjugated verb.

☀ When a title contains a conjugated verb, the contribution starts with the verb and continues until the end of the sentence (the sentence stress); otherwise, the contribution is stated upfront in the title.

> *"Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation"*[5]

The contribution is not that thymocyte apoptosis is induced by Glucocorticoid, This is an already known fact. The contribution is the association between apoptosis and endogenous endonuclease activation.

2-Addition of verbal forms

A title without a conjugated verb lacks energy. Nouns do not have the action-packed strength of a verb.

---

[5] A. H. Wyllie Glucocorticoid-induced thymocyte apoptosis is associated with endogenous endonuclease activation, *Nature* **284**: 555–556 (10 April 1980).

☀ The gerundive and infinitive verbal forms add energy to a title without a conjugated verb.

*"Data learning: **understanding** biological data"*[6]

*"Nonlinear Finite Element Simulation **to Elucidate** the Efficacy of Slit Arteriotomy for End-to-side Arterial Anastomosis in Microsurgery"*[7]

3-Adjectives and numbers featuring qualitative or quantitative aspects of the contribution

Adjectives and adverbs are also used to attract — *fast*, *highly efficient* or *robust*. They enhance the contribution. Since adjectives are subjective, replacing them with something more specific is always better. A "100 MHz DCT processor" is clearer than a "fast DCT processor"; and while in twenty years "fast" will make a liar out of you, "100 MHz" will not. And by the way, avoid using '*new*,' '*novel*,' or '*first*,' which are not informative because a contribution should be new, novel and a first, anyway.

A 100 MHz 2-D 8 × 8 DCT/IDCT processor for HDTV applications[8]

4-Clear and specific keywords 💬

Specific keywords attract the expert. The specificity of a paper is proportional to the number of specific keywords in its title. Beware of keywords buried in long modified nouns whose clarity is inversely proportional to the length of the modified noun. To clarify such nouns, add prepositions. You may lose in conciseness, but you certainly gain in clarity — the gain is greater than the loss.

---

[6] Brusic, V., Wilkins, J. S., Stanyon, C. A. and Zeleznikow, J. (1998a). Data learning: understanding biological data. *In:* Merrill G. and Pathak D.K. (eds.) Knowledge Sharing Across Biological and Medical Knowledge Based Systems: Papers from the 1998 AAAI Workshop pp. 12–19. AAAI Technical Report WS-98–04. AAAI Press.

[7] Reprinted from journal of biomechanics, vol 39, Hai Gu, Alvin Chua, Bien-Keem Tan, Kin Chew Hung, nonlinear finite element simulation to elucidate the efficacy of slit arteriotomy for end-to-side arterial anastomosis in microsurgery, pages 435–443, copyright 2006, with permission from Elsevier."

[8] Madisetti, A. Willson, A.N., Jr., "A 100 MHz 2-D 8 × 8 DCT/IDCT processor for HDTV applications", IEEE Transactions on Circuits and Systems for Video Technology, Apr 1995, Vol. 5, no. 2, pp158–165.

"Transient model for kinetic analysis of electric-stimulus responsive hydrogels" (unclear)

*"Transient model for kinetic analysis of hydrogels responsive **to** electric stimulus" (clear)*

5-Smart choice of keyword coverage

Even published, an article has little impact if not found. Readers find new articles through online keyword searches; that is why choosing effective keywords is vital.

Keywords are divided into three categories (☛1).

General keywords ('*simulation,*' '*model,*' '*chemical,*' '*image recognition,*' '*wireless network*') are useful to describe the domain, but they have little differentiating power precisely because they frequently appear in titles. They do not bring your title closer to the top of the list of retrieved titles. Intermediary keywords are better at differentiating. They are usually associated with methods common to several fields of research ('*fast Fourier Transform,*' '*clustering,*' '*microarray*') or to large sub-domains ('*fingerprint recognition*'). But, for maximum differentiation, specific keywords are unbeatable ('*Hyper surface,*' '*hop-count localization,*' '*non-alternative spliced genes*'). For a given journal,

Keyword type

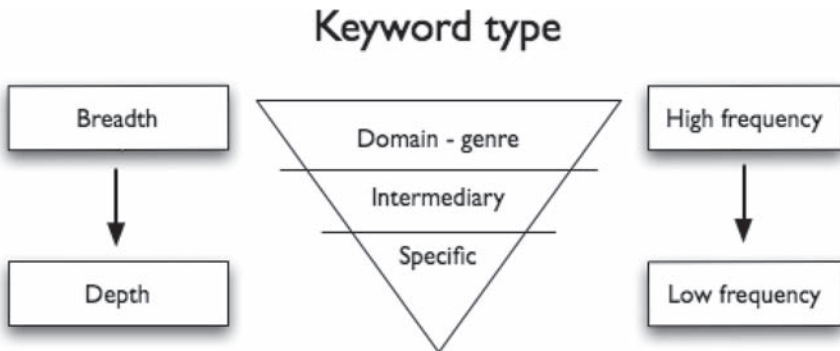| Breadth | Domain - genre | High frequency |
| | Intermediary | |
| Depth | Specific | Low frequency |

Figure ☛1

Specialized keywords are at the pointed deep end of the inverted triangle. General keywords are at the broad top end of the triangle. The general to specific scale is correlated to the frequency of use of a scientific keyword. Depth and breadth of a keyword are not intrinsic qualities. They depend on the frequency of use of these words in the journal that publishes the paper. The reader's knowledge also influences the perception of keyword levels: the less knowledgeable the reader is, the more general keywords seem specific.

or for domain experts, the category of a keyword is well-defined. It changes from journal to journal, or from experts to non-experts. *Polymer* would be an intermediary keyword in the journal Nature, but it is definitely generic in the journal of Polymer Science.

Make sure your title has keywords at more than one level of the triangle. If too specific, your title will only be found by a handful of experts in your field; it will also discourage readers with a sizable knowledge gap. If too general, your title will not be found by experts. Decision on the keyword choice is yours. Base it on who you are writing your paper for.

If you are a known author pioneering the research in your field, someone readers follow through your announcements of publications on twitter, or someone whose articles are retrieved simply by searching by author name, citations, or references… If you are that someone, do not worry excessively about other people's titles: you are the leader. But if that is not the case (yet), make sure your title has more than one keyword. People need keywords to find it.

6-Catchy attention-getting schemes

Catchy Acronym. The BLAST acronym is now a common word in bioinformatics. It started its life as five words in a title, "Basic Local Alignment Search Tool," published in the 1990 Journal of Molecular Biology. The author built a fun and memorable acronym, and everyone remembered it. Acronyms provide a shortcut to help other writers refer to your work succinctly.

> *"Visor: learning **VI**sual **S**chemas in neural networks for **O**bject **R**ecognition and scene analysis"*[9]

The title above is that of the doctoral thesis of Wee Kheng Leow. Other researchers mentioning his work could, for example, write "In the VISOR system [45]". The acronym provides a convenient way for others to refer to his work. Notice that both BLAST and VISOR are memorable. Acronyms like GLPOGN are doomed to fail.

---

[9] Wee Kheng Leow (1994). VISOR: Learning Visual Schemas in Neural Networks for Object Recognition and Scene Analysis, PhD Dissertation; Technical Report AI-94–219, June 1994.

Question. The question makes a mighty hook. But journal editors rarely allow them.

"Software acceleration using programmable logic: is it worth the effort?"[10]

Words out of the expected range. Here is a catchy and intriguing title that relies on a simile.

"The Diner-Waiter pattern in Distributed Control"[11]

"Distributed control" is not usually associated with the interaction between a restaurant waiter and a customer. What the title gains in interest, it loses in retrievability: it only has one general domain keyword, (*"distributed control"*) and researchers in this domain are unlikely to even think of *"diner-waiter"* as a search keyword. The next title has words out of the expected range that do not work cross-culture.

"The inflammatory macrophage: a story of Jekyll and Hyde"[12]

Would you understand this title if you were a researcher from mainland China? Dr. Jekyll never published any papers, as far as I know. This title only makes sense if you have read the 1886 novel by Robert Louis Stevenson.

Besides these six popular techniques to improve titles, listed here without example are three more.

- Some keywords become buzzwords that carry the passion of our times. Seeing them in titles attracts the reader keen to keep up-to-date with the happenings in Science.
- A shorter title is more attractive than a long one, and a general title more attractive than a specific one.
- Words that announce the unexpected, the surprising, or the refutation of something well established, all fuel the curiosity of the reader.

---

[10] Martyn Edwards, Software acceleration using programmable logic : is it worth the effort, Proceedings of the 5th international workshop on hardware/software co-design, IEEE computer society, p.135, March 1997.

[11] H. He and A. Aendenroomer, "Diner-Waiter Pattern in Distributed Control", Proceedings of 2nd International Conference on Industrial Informatics, (INDIN'04), Berlin, Germany, 12–16 June 2004, Vol. 2, pp. 293–297.

[12] Duffield JS., The inflammatory macrophage: a story of Jekyll and Hyde, Clin Sci (Lond). 2003 Jan; 104(1): 27–38.

To make a title catchy, there is one rule only: catchy, yes; dishonest, no.

Purpose and Qualities of Titles

Purpose of the title for the reader

- It helps the reader decide whether the paper is worth reading further.
- It gives the reader a first idea of the contribution: a new method, chemical, reaction, application, preparation, compound, mechanism, process, algorithm, or system.
- It provides clues on the paper's purpose (a review, an introductory paper), its specificity (narrow or broad), its theoretical level, and its nature (simulation, experimental). By the same means, it helps the reader assess the knowledge depth required to benefit from the paper.
- It informs on the scope of the research, and possibly on the impact of the contribution.

Purpose of the title for the writer

- It allows the writer to place enough keywords for search engines to find the title.
- It catches the attention of the reader.
- It states the contribution in a concise manner.
- It differentiates the title from other titles.
- It attracts the targeted readers, and filters out the un-targeted readers.

Qualities of a title

UNIQUE. It differentiates your title from all others.

LASTING. Do not to use *new* in a title that outlives you.

CONCISE. Remove redundant or filler words ('study of'). If your title is already unique and easy to find without the details, remove them.

CLEAR. Avoid long modified nouns. They bring imprecision and misunderstanding.

HONEST and REPRESENTATIVE of the contribution. It sets expectations and it fulfills them.

CATCHY. The writer has one chance and 2 seconds to interest the reader.

EASY TO FIND. Its keywords are carefully chosen.

What do you think of your title? Does it have enough of the qualities mentioned here? Is your contribution featured at the head of your title? It is time you take a closer look.

Title Q&A

Q: When do I write my title?

A: There is no set time. You may write a tentative title as soon as you identify something is novel and useful enough to merit publication. An early title gives focus to your work. As your research progresses, you may decide to change your title to include new findings. But you may also discover new elements that, on their own, represent sufficient contribution to merit their own paper. How would you know which route to take? Two papers may be better if your title gets longer when you try to bring in everything together. But if your title

becomes shorter, and more general, you may have a better, more attractive paper. General or generic titles should not be abused, however. It is better to have a string of published short papers than to have an unpublished paper with a generic title covering a shapeless amalgam of new research tidbits.

Q: Should I look at other titles in my list of references before or after I write my final title?

A: Your final title should have all the qualities mentioned in this chapter, and not be influenced by other people's titles — apart from making sure your title is unique. This said, while looking at the titles of related papers (and relevant titles on www.SciTopics.com), you may discover that the most highly cited papers use a particular keyword, whereas you use a synonymous keyword also often found. It may be advantageous to use instead the one found in the title of the highly cited papers so that when their titles are retrieved, your title is retrieved alongside theirs.

Q: When two keywords are synonymous, which one do I choose for my title?

A: From what we have seen before, the most frequent one should be chosen. When two different keywords with the same meaning appear with the same frequency in titles, choose one for the title and leave the other one for the abstract. That way, search engines will find your paper, regardless of the keyword used for the search

Q: Is it appropriate to mention the impact of the contribution in a title?

A: *"Multiple reflections of Lamb waves at a delamination"* is the title of a paper by T. Hayashi and Koichiro Kawashima.[13] After reading the paper, scientists who attended the writing skills seminar suggested to rewrite the title to depict more precisely what the paper contains. Here is their proposed title:

> *Multiple reflections of Lamb waves for rapid measurement of delamination.*

---

[13] T Hayashi, Koichiro Kawashima, 2002. Multiple reflections of Lamb waves at a delimitation, *Ultrasonics*, **40**(1–8): 193–197.

Note the addition of *'rapid measurement'*. What attracted the readers, the research impact, was absent from the title. Yes, the title is longer, but also more attractive.

Q: Isn't it better to keep the title general so that it attracts a maximum of readers?

A: The following title of a paper by B. Seifert *et al.*, published in the artificial organs journal (Volume 26 Issue 2, Pages 189–199) is a typical title with two parts separated by punctuation. It emphasizes the name of the polymer and its novelty in membrane-forming.

> *"Polyetherimide: A New Membrane-Forming Polymer for Biomedical Applications"*

The terms *'biomedical applications'* is general, maybe too much so for a journal dealing with biomedical applications anyway. The indefinite article *'a,'* is general and non-descriptive: it means *'one of many'*. *'New'* is going to be obsolete very soon. *'Membrane'* is also a general term. Membranes are of many types and have many properties. Here is how the readers of this paper proposed to change its title to make it more specific and yet more attractive:

> *Polyetherimide: A biocompatible polymer to form anti-fouling membranes*

They added a hot keyword *'Biocompatible,'* and they characterized the membrane property with *'anti-fouling'*. The title now includes a dynamic verb (*'to form'*) enhancing its appeal.

Q: I agree that a short title is better than a long one, but how does one go from a long title to a short title?

A: The following title of a paper by Cook BL, Ernberg KE, Chung H, Zhang S, published in the PLoS One. 2008 Aug 6;3(8) is long:

> *"Study of a synthetic human olfactory receptor 17–4: expression and purification from an inducible mammalian cell line"*

Look out for words such as *'study of,'* or *'investigation of'*. You can safely remove them. *'a'* is not attractive, and misleading because in this case, there is only one such receptor. Readers of the article rewrote the title to make it concise by removing what they considered an unimportant and unproductive search keyword (the Mammalian cell line).

Purifying the synthetic human olfactory receptor 17–4 in milligram quantities

The readers placed the outstanding contribution right upfront in the title, in verb form to make the title more attractive, and they added the outcome of the contribution at the end of the title.

Q: What are the consequences of changing the keywords in my title?

A: Here are two alternative titles for a same paper written by Dr. Linda Y.L. Wu, A.M. Soutar and X.T. Zeng, and published in Surface and Coatings Technology, 198(1–3), pp. 420–424 (2005).

"Increasing hydrophobicity of sol-gel hard coatings by chemical and morphological modifications"

"Increasing hydrophobicity of sol-gel hard coatings by mimicking the lotus leaf morphology"

The second title is quite catchy. Only one keyword describing the methodology was lost ('*chemical modification*'). '*Lotus leaf*' is unexpected. That keyword may attract scientists outside the domain of manufacturing technology, and even journalists writing for widely distributed science magazines. The author of this paper, Dr. Linda Wu, decided to keep the first title even though the second title was indeed attractive. Why? The journal she targeted with her title was "Surface and Coating Technologies". Its readers have a chemical, and material science background, not biology. Had she chosen the '*lotus leaf*' title, she would have attracted a more general audience to her paper. As a result, she would have had to rewrite the whole paper, change its structure (emphasis on morphology and biomimicry), and simplify the vocabulary (simpler terms understandable by a larger group of non-experts).

Q: If I submit a paper to different journals, do I need to change the title?

A: First, if you write a paper, it is for *one* journal. Look at the web page of the journal you are targeting. It specifically mentions the type of contents the journal covers. Each journal attracts different types of readers in terms of expertise. Therefore, each paper is written for

the specific readers of a specific journal. A paper cannot possibly be written the same way for two different journals. Not only should your title change, but also the contents of the paper.

Also, one should not send the same paper to two journals at the same time. Submission should be a serial process. Pick one journal. If that journal decides not to publish you or if you feel that the journal's lengthy review process (due in part to a reviewer's attitude towards your paper) delays important results from being published, submit it to another journal — after reworking its contents and title to match the readers' interests. Read the interesting recommendations of the US department Office of Research Integrity (ORI) on the subject.[13]

Q: In the role of the title for the writer, you write "[the title] catches the attention of the reader **targeted** by the writer". Why should the writer target anyone? Isn't the reader the one identifying the target according to his or her needs?

A: Your impact will be negligible if you do not have in mind the reader who can make use of your contribution. What keywords is this reader looking for? Are they in your title? These keywords usually describe the impact or the application domain. In a precedent title, the words *'in milligram quantities'* were added to the title by the readers themselves, as they thought this was very significant and opened many new applications. In another title the words *'rapid measurement'* were added by readers, again reflecting the same concern to present real solutions to real problems faced by the scientists in this field. Put yourself in the shoes of your reader, and if space allows, find a way to mention the outcome of your contribution in the title of your paper.

Q: In a previous chapter you mentioned that adjectives make claims and are therefore dangerous to use in a scientific paper. But in this chapter you write that adjectives are attractive. Isn't there a contradiction here?

A: Precisely, the eye-catching adjectives make claims. But so does any title. It should not be astonishing to find an adjective such as *'robust,' 'efficient,'* or *'fast'* in a title. Once the claim is made, the writer is expected to provide early justification for that claim, i.e in the abstract, even if the justification is only partial.

---

[13] http://ori.dhhs.gov/education/products/plagiarism/14.shtml

Q: When one writes a title, what are the advantages and drawbacks for being first in a field?

A: The advantage is clear. If you are a pioneer in your field, the choice of words is entirely yours. Since you are the first to write in this field, you need not worry about titles that may already have been used. Think about it. Imagine being the first to write about dialogue in speech recognition. Finding a title is easy: *"Dialogue in speech recognition"*. Nice and short. Now imagine you are writer 856 with a paper in this crowded field. You have to be much more specific to differentiate your title from all the others. As a result, you might have to settle for a long specific title like *"semantic-based model for multi phase parsing of spontaneous speech in dialogue systems"*

Being first gives you the opportunity to have a shorter title… that won't be found! You are a pathfinder. At first, people may not find you through the possibly new keyword you introduce (the name you gave to a new polymer, for example), because they are not yet aware of it. This is why you want to make sure your title is found by including other well-known keywords. Alternatively, if you are already famous, you need not care. People find you by name, not through the keywords of your papers.

Title Metrics

✓(+) The words representative of your contribution are upfront in a verb-less title.

✓(+) The title has 2+ search keywords, in different sections on the inverted pyramid.

✓(+) Your title has attractive words (non search keywords)

✓(+) No noun phrase exceeds three words.

✓(+) All your title search keywords are found in your abstract.

✓(+) Your title is read in less than two seconds and is clear at first reading.

✓(+) No search keyword present only in the abstract appears with a frequency higher than any title search keyword found in the abstract.

✓(+) The impact or outcome of your contribution is identifiable in the title.

✓(+) Your title clearly sets the scope of the research.

✓(−) The words representative of your contribution are back and front, or at the back.

✓(−) All your keywords are in the same inverted pyramid section.

✓(−) The title has only one search keyword.

✓(−) Your title has two prepositions (and, or) or one (with).

✓(−) Your title has no attractive words.

✓(−) Your title contains *'a,' 'an,' 'study,' 'Investigation'*.

✓(−) Nouns phrases have 3+ words (e.g. Metastable transition metal nitride coatings).

✓(−) Some title search keywords are missing from your abstract.

✓(−) Your title requires more than two seconds to read.

✓(−) Search keywords repeatedly found in the abstract are not in the title.

✓(−) Your title creates diverging expectations about its contents.

✓(−) Your title does not set the scope of the research, or does so partially.

AND NOW FOR THE BONUS POINTS:

✓(+++) The title creates only one expectation about its contents, and fulfills it.

This page intentionally left blank

Chapter 11

Abstract: The Heart of Your Paper

The heart plays an essential role in the human body. Similarly, the essence of an article is its abstract. *The heart has four chambers.* The abstract is also composed of four easily identifiable parts.

> **Visuals in Abstracts?**
>
> *Never say never! I used to think that abstracts had no visuals, but it looks as though I was mistaken. The table of contents of some journals (e.g. Advanced Materials, Journal of the American Chemical Society) now include one key visual alongside an abridged abstract. Is this a preview of the shape of things to come for all journals? I believe it is. A good figure far exceeds plain text in illustrating and explaining a contribution efficiently and concisely. Therefore, take note and prepare yourself. Which visual will be 'the one' to choose for your abstract?*

The abstract dissected here is at the crossroads between surgery and computer science. It comes from a paper on slit arteriotomy. The easiest way to explain arteriotomy is to visualize anastomosis — the surgical connection of two tubes (here, arteries). Normally the surgeon cuts an elliptic hole (with removal of material) in the recipient artery and then stitches the donor artery over the hole. In this case, however, only a slit is cut in the side of the recipient artery before the donor artery is stitched over it. No material is removed.

Does slit arteriotomy work as well as hole arteriotomy? Even if the answer is yes, surgeons are conservative. If an established procedure (the hole arteriotomy) works, why is there a need to replace it with a new one (slit arteriotomy)! Initial statistics that establish the equivalence of both techniques are not enough. What will happen to the slit

in ten years when arteries age or when the patient's blood pressure rises? To find out the safety and efficacy of the new technique over time, the inventor surgeon asked for the help of computer-modeling scientists. The technique was modeled, and a paper[1] was published in the *Journal of Biomechanics*. The readers of this journal come from diverse horizons: life science, engineering science, and computer science. When they glanced at the table of contents of volume 39 of the journal, they saw the following title:

> *Nonlinear Finite Element Simulation to Elucidate the Efficacy of Slit Arteriotomy for End-to-side Arterial Anastomosis in Microsurgery*[2]

The title has two parts: contribution and background. If you were to insert a dividing bar | between these two parts, where would you place it? The answer will come later, after you read the abstract. Note that the words in bold are common to both the abstract and the title.

> *"[Part 0–20 words]* **The slit arteriotomy for end-to-side arterial Microanastomosis** *is a technique used to revascularize free flaps in reconstructive surgery. [Part 1–41 words] Does a slit open to a width sufficient for blood supply? How is the slit opening affected by factors such as arterial wall thickness and material stiffness? To answer these questions we propose* **a non-linear finite element** *procedure to simulate the operation. [Part 2–10 words] Through modeling the arteries using hyperelastic shell elements, our* **simulation** *[Part 3–112 words] reveals that the slit opens to a width even larger than the original diameter of the donor artery, allowing sufficient blood supply. It also identifies two factors that explain the opening of the slit: blood pressure which is predominant in most cases, and the forces applied to the slit by the donor artery. During simulation, when we increase the donor artery thickness and stiffness, it is found that the contribution of blood pressure to the slit opening decreases while that of the forces applied by the donor artery increases. This result indicates that sometimes the forces by the donor artery can play an even more significant role than the blood pressure factor.*

---

[1] Reprinted from journal of biomechanics, Vol 39, Hai Gu, Alvin Chua, Bien-Keem Tan, Kin Chew Hung, nonlinear finite element simulation to elucidate the efficacy of slit arteriotomy for end-to-side arterial anastomosis in microsurgery, pages 435–443, copyright 2006, with permission from Elsevier.

[2] H. Gu[a], A.W.C Chua[b], B. K. Tan[b], K.C. Hung[a].

[a] *Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore.*

[b] *Department of Plastic Surgery, Singapore General Hospital.*

*[Part 4–28 words] Our simulation **elucidates the efficacy** of the slit arteriotomy. It improves our understanding of the interplay between blood pressure and donor vessel factors in keeping the slit open. [Total: 211 words].*

Where does the bar "|" fall, in other words, where is the separation between the contribution and the context of the contribution?

"Nonlinear Finite Element Simulation to Elucidate the Efficacy of Slit Arteriotomy | for End-to-side Arterial Anastomosis in Microsurgery"

In our sample abstract, if one locates the contribution based on the word count in each part, it seems that part 3, the elucidation of the efficacy, covers the contribution (112 words). Part 2, the non-linear finite element analysis, plays an incidental role (only 10 words), yet it comes right upfront in the title. The title could have been the following:

Elucidating the Efficacy of Slit Arteriotomy | for End-to-side Arterial Anastomosis in Microsurgery using non linear finite element simulation

However, after examining the structure of the paper (headings and subheadings), it appears that the contribution is indeed the nonlinear finite element simulation. The title tallies with the structure, less so with the abstract. One concludes that the abstract is aimed at surgeons who care little about the technical details of the finite element simulation. They may never read the paper, and be content reading the abstract. Had the paper targeted computer scientists, the methodology part would have been longer and the results part shorter.

The Four Parts

Each of the four parts in the abstract (separated by word count) answers key reader questions.

Part 1: What is the problem? What is the topic, the aim of this paper?

Part 2: How is the problem solved, the aim achieved (methodology)?

Part 3: What are the specific results? How well is the problem solved?

Part 4: So what? How useful is this to Science and to the reader?

You may have noticed that our sample abstract has a **Part 0**. It is optional, and not recommended, apart from situations such as the one arising in our sample abstract. The writer, anticipating that the meaning of a keyword in the title may be obscure to non-surgeons, provides just-in-time background — in this case, a functional definition of the surgical procedure.

A four-part abstract should be the norm, but many have only three parts: the fourth one (the impact) is missing. Why?

1) Was the maximum number of words allowed by the journal reached too quickly because a long rambling start justified the importance of the aim, thereby forcing the author to keep silent on the benefits of having achieved the aim?

2) Did the author (mistakenly) consider that the results speak for themselves? Again, it is worth repeating that the writer must have a reader in mind when writing a paper. Who is likely to benefit from the research? The impact of the paper is what convinces a reader to download your paper. And that reader may not have enough knowledge to determine the impact (outcomes) made possible by the results (outputs).

3) Could it be that the author was unable to assess the impact as a result of the myopia caused by the atomization of research tasks among many researchers?

4) Could it be that the author was unable to mention any impact because the contribution is only a small improvement over a previous result, not enough to claim significant impact?

Whatever the reason, having less than four parts reduces the informative value of the abstract and, therefore, its value to the reader. Since the reader decides whether to read the rest of your article or not based on the abstract, its incompleteness reduces your chances to be read… and cited.

Read your abstract and locate its various parts. Does your abstract have its four essential parts? Are the parts with the largest number of words, those corresponding to the contribution? Are you still using adjectives and remaining vague when you should be precise?

Coherence Between Abstract and Title

A rapid keyword count will determine whether the abstract is coherent with the title. The title contains 9 search keywords: [Slit, Arteriotomy, End-to-side, Arterial, Anastomosis, Microsurgery, non-linear, finite-element, simulation]. In this count, articles (a, an, the, etc.), prepositions (to, of, for, in), and non-search keywords (elucidate, efficacy) are not taken into account.

In our sample abstract, 6 words are both in the title and in the first sentence of the abstract (*66%*).

> (TITLE) *Nonlinear Finite Element Simulation to Elucidate the Efficacy of* **Slit Arteriotomy** *for* **End-to-side Arterial Anastomosis** *in Micro-**surgery*** *(title)*
> (FIRST SENTENCE) *The* **slit arteriotomy** *for* **end-to-side arterial** *Micro-**anastomosis** is a technique used to revascularize free flaps in reconstructive* **surgery***.

This percentage is good. Why? The reader, having just read the title, expects to know more about it as soon as possible. Can you imagine the first sentence of your abstract disconnected from the message announced by the title? It is unimaginable. Coherence between title and abstract is achieved through the repetition of keywords. Percentages outside the 30%–80% range should be examined more closely.

0%–20% There is a problem. The first sentence deals with generalities loosely related to the topic of the paper. It contains two title words or less. It sets the background to the problem. If it briefly explains one or two unusual title keywords, this is fine, as long as sentences 2 and 3 mention most of the other title words. Otherwise, the background is too long and, as a result, the abstract lacks conciseness.

80%–100%. Idyllic percentage? Not necessarily. The first sentence repeats the title with just a verb added. Why repeat! The first sentence should expand the title. However, if that sentence contains many more words than the title, then 80–100% may be acceptable.

To summarize, your title merely whets the appetite of your readers; they expect to know more about your title in your abstract. You should satisfy their expectation and rapidly provide more precise details. The first sentence of your abstract should contain at least one-fourth of the words in your title.

First count the total number of significant (search) words in your title (In your count, do not include non search words such as *on, the, a, in*, or adjectives). Let's call that number T.

Then, see if your first sentence contains any of the T words. If you find some, underline them IN THE TITLE. Modified forms (a noun changed to a verb or vice versa) are acceptable but synonyms are not. For example, simulation would be considered the same as simulated, but abrasion would not be the same as corrosion.

Count the number of words underlined in your title. Let's call that number U.

Calculate the percentage $100 \times U/T$.

What is your percentage? Between 20 and 80%, you are doing fine. Outside this range, investigate.

A second count will help you identify the strength of the cohesion between abstract and title. Are ALL title keywords also in the abstract? They should be.

In our sample abstract, 6 out of the 9 words were in the first sentence, three were still missing. But they were in the rest of the abstract (*we propose **a non-linear finite-element** procedure to **simulate** the operation*). All title search keywords were in the abstract.

Think about it. You give high visibility to a word by giving it title status — the highest status in a paper. Why would title words be missing in the abstract? It may be for the following reasons:

- The title word is not important. Remove it from the title to increase conciseness.
- The title word missing in your abstract is really important. Find a place for it in your abstract.
- It may also be that your abstract frequently refers to a keyword not found in the title. Rewrite your title to incorporate that keyword.
- You used a synonym to avoid repetition. Don't. Repeating a title word in the abstract increases the relevance score calculated by search engines for that keyword. As a result, your title will be brought up towards the top of the list of titles retrieved.

You have already counted, T, the number of significant words in your title. Read your abstract and see if any of the important title words are missing. If some are, ask yourself why. Are your title claims are too broad? Is your title not concise enough? Are you using synonyms that dilute the strength of your keywords and confuse the reader? Decide which reason applies, and modify title or abstract if necessary.

You now have three techniques to gauge the quality of your abstract.

A) Abstracts have four parts (and one optional part for just-in-time explanation of obscure title words as seen in our sample abstract). The part that represents your contribution should be the most developed.

B) Abstracts repeat all title words. A possible exception to this recommendation is when you use alternative keywords because a particular concept is expressed by two equally probable keywords and you want your paper to be found. You then use one keyword in the title, and the other equally probable keyword in the abstract.

C) Abstracts expand the title in the first one or two sentences because the reader expects it.

Tense of Verbs and Precision

So far, you have managed to bring the reader past the title barrier and into your abstract. Congratulations! But if the reader stops there, what are your chances to increase your citation count? You need to use your abstract as a launch pad from which the reader will propel himself or herself inside your paper by clicking on the *download pdf file* button. For some, it even means entering their visa card details for the online payment if their library does not subscribe to the journal which publishes your paper.

How does one turn an abstract in a launch pad? In two ways, by writing it in the tense used in ads — the present tense, and by convincing the reader to read more of your paper through a brief but precise account of your main accomplishments.

Accomplishments are better communicated with the dynamic present tense than with the dull past tense. There are advantages to choosing the present tense for the abstract. The present tense is

vibrant, lively, engaging, leading, contemporary, and fresh. Even if the reader discovers your paper ten years after it has been published, it will feel as fresh and current as if it had just been written. The past tense is passé, déjà vu, a thing of the past, gone, stale, unexciting, and lagging. It feels like reading old news. It could even introduce ambiguity. For example, the phrase *was studied* creates a doubt: did the writer publish this before?

If ever you needed one more reason to be convinced that the present tense is preferable, first check whether the journal targeted by your paper is agreeable to such modern practices or not. Following which, if there is no journal-imposed past tense directive, think how often you jump straight from the abstract to the conclusions when you read a new paper. Since the conclusions are also written in the past tense, to the reader, reading the conclusions feels like reading the abstract all over again (i.e. boring). Let's now move to the second element of our launch pad: precision.

Precision is convincing. The title is too concise to satisfy the reader eager to get precise information on your contribution. The abstract continues to sustain the initial excitement created by the title by stating your main accomplishment(s) with numerical precision, or with a brief but precise account of the novel steps if your contribution is a method.

The following title contains a couple of attractive words: '*clinically distinct*'. How *distinct*, though? The reader wants to know with precision before deciding whether your paper is worth reading further.

> "*A gene expression-based method to diagnose* **clinically distinct** *subgroups of diffuse large B Cell Lymphoma*"

The abstract does answer the question by justifying the claim made in the title.

> "**The GCB and ABC DLBCL subgroups** *identified in this data set* **have significantly different 5-yr survival rates after the multiagent chemotherapy (62% vs. 26%;** $P = 0.0051$**)**, *in accord with analyses of other DLBCL cohorts. These results demonstrate the ability of this gene expression-based predictor to classify DLBCLs into biologically and clinically distinct subgroups irrespective of the method used to measure gene expression.*"[3]

---

[3] George Wright, Bruce Tan, Andreas Rosenwald, Elain Hurt, Adrian Wiestner, and Louis M.Staudt, a gene expression-based method to diagnose clinically distinct

The sample abstract provided in this chapter is entirely written in the present tense. Its results, are not numerical, but the main results are described with precision.

> *"The slit opens to a width even larger than the original diameter of the donor artery, allowing sufficient blood supply. It also identifies two factors that explain the opening of the slit: blood pressure which is predominant in most cases, and the forces applied to the slit by the donor artery. […] the contribution of blood pressure to the slit opening decreases while that of the forces applied by the donor artery increases."*

US$ 35

Vladimir had located what looked like a really interesting paper. He had asked the librarian to download it. Instead of receiving in the internal mail the paper he expected, he received an email from the librarian informing him that the research center did not subscribe to the journal that published his requested paper. However, she offered to download the paper and charge his department the download fee of US$ 35 as long as he secured his manager's approval.

"Thirty-five bucks!" Vladimir exclaimed!

His friend in the next cubicle heard him across the partition. He shouted: "Not enough, but I'll settle for fifty!"

"Oh, shut up John. The librarian is charging us thirty-five bucks for a reprint of someone's paper. I could buy a book for that price!"

John had stood up and was now looking at Vladimir across the cubicle divider.

"You must be joking! That much?" He said.

"How much?" That was Cheng Jia, another colleague who was sitting back to back with John in the same cubicle. She appeared next to John.

The commotion attracted the attention of Professor Popov who, at that instant, happened to look away from his desk, past his opened office door, towards the group. He joined in the conversation, listened to the complaint, and advised:

"Vlad, read that abstract again. If it contains enough detail to convince you that you could really benefit from reading this paper, then I will authorize the purchase."

subgroups of diffuse large B Cell Lymphoma, Proceedings of the national Academy of Sciences, August 19, 2003 Vol 100 No.17, www.pnas.org, copyright 2003 National Academy of Sciences, U.S.A.

Purpose and Qualities of Abstracts

Purpose of the abstract for the reader

- It clarifies the title.
- It provides convincing details on the writer's scientific contribution.
- It helps the reader decide whether the rest of the article is worth buying or downloading for further reading.
- It helps the reader rapidly gather competitive intelligence.

Purpose of the abstract for the writer

- Because it has more keywords than the title, the abstract allows the paper to be found more easily.
- It states the writer's contribution in more precise detail than the title to persuade the reader to read the rest of the paper.
- When written early in the life of a paper, it guides the writing and keeps the writer focused.

The abstract is NOT to be used for the following:

1) To mention the work of other researchers, except when your paper is an extension of a (single) previously published paper, yours or that of another author.
2) To justify why the problem you have chosen is significant: the significance of your impact on the problem is the one that matters the most.

Qualities of an abstract

TIED TO TITLE. All title words are in the abstract.

COMPLETE. It has four parts (what, how, results, impact).

CONCISE. Not longer than necessary, as a courtesy to the reader. Research is justified through significance of results, not significance of problem.

STAND-ALONE. It lives by itself in its own world: databases of abstracts, journal abstracts. It may include the main key visual in the future.

REPRESENTATIVE of the contribution of the paper. It sets expectations. It encourages the reader to read the paper.

PRECISE when describing contribution.

PRESENT. Real. News.

What do you think of your abstract? Does it have enough of the qualities mentioned here? Is the contribution you mention in your abstract consistent with that claimed by the title? A quality abstract makes a good first impression. Spend some time reviewing it.

Abstract Q&A

Q: What is an extended abstract?

A: An extended abstract is not an abstract. It is a short paper that a reader reads in less than an hour — usually a fifth of a regular paper. Unlike the abstract, it contains a few key visuals (space allowing) and references. It is short on details. For example, even though the extended abstract has an introduction, it does not paint the context of the contribution in great details, but focuses instead on the immediate related works that are essential to understand and value your contribution. Likewise, its conclusion (if there is one) skips the future

work. In some ways, writing an extended abstract is more difficult than writing a full paper because one has to make hard choices when it comes to details: what to leave out, and what to keep in to interest and convince.

Q: Can I submit to a journal a paper that corresponds to the extended abstract I prepared for a conference?

A: The short answer is no. There would be no added novelty, just additional details on the same contribution. This is an issue that journals deal with explicitly in their guides to authors, so do read your journal guidelines on this matter. When you submit your paper to non open-access journals, you sign a legally binding submission declaration assigning the copyright of your material to the journal. Therefore, reusing in the journal paper paragraphs recycled from the extended abstract would be considered self-plagiarism.

The extended abstract is often nothing more than the paper corresponding to the first research step. Additional steps are required for a journal paper. Its title would be different, its visuals also, and much more.

It is always good to declare the existence of your extended abstract to the journal editor when submitting your paper, and to explain the differences between it and your paper (different methodology, increased scope, new findings…).

Q: Do all abstracts have four parts?

A: Not all abstracts have four parts, sometimes with good reason. A review paper that covers the state of the art in a particular domain has only one or two parts. Short papers (letters, reports) have one or two lines. "Extended" abstracts are written prior to a conference, in some cases well before the research is completed; as a result, their parts 3 and 4 are shallow or missing. But, apart from these special cases, all abstracts should have four parts.

Q: Some journals limit the length of an abstract to 300 words. Is it best to use all 300 words?

A: Think of the busy reader! If you can write your abstract in less than 300 words, and still have a complete abstract (4 parts), then be brief! Brevity comes through selection.

Q: Abstracts should be written in the past tense. Everybody does it this way. Where do you get this present-tense nonsense from?

A: Never say never! I used to think that abstracts were only written in the past tense because they refer to work that is completed. It looks as though I was mistaken. More journals now accept and even recommend the use of the present tense for the abstract. Even NASA scientists are advised to use the present tense in their abstracts (in Mary McKaskill's grammar).

Before you select the tense for your abstract, read the journal recommendations to the author. If the journal does not forbid the use of the present tense, consider using it.

Q: Does an abstract contain text only?

A: Actually, the visual abstract is slowly making its way into some chemistry, material science, or even biology journals. It helps the reader understand the contribution visually and with few words. Pathways, chemical formulas, or microstructures are more effectively understood visually. If your journal demands a visual for its table of contents, be ready to pick the visual the most representative of your contribution.

Q: My paper is a review paper. Do the principles detailed in this chapter also apply to this kind of paper?

A: A review paper is quite different. Look at the resources provided in the final book chapter to find guidelines for review papers.

Abstract Metrics

✓(+) All search keywords in your title are also in the abstract.

✓(+) Your abstract has all four main parts (what, how, results, impact).

✓(+) The part that contains the contribution has the greatest word count.

✓(+) Your abstract does not contain background or justification of problem importance.

✓(+) Your abstract is written using verbs at the present or present perfect tense only.

✓(+) Your abstract mentions the main result(s) with precision or the key method steps if contribution is a method.

✓(+) By revealing the main outcome of your results, your abstract targets the reader who stands to benefit the most from your research.

✓(−) One or more search keywords in your title are missing in the abstract.

✓(−) The first sentence in your abstract is more or less a repetition of the title.

✓(−) The first sentence in your abstract contains none or just one of the title keywords.

✓(−) Your abstract is missing one of the main parts.

✓(−) The part that contains the contribution does not have the greatest word count.

✓(−) Your abstract is written using the past tense only, or a mix of various tenses.

✓(−) Your abstract remains vague and lacks precision when mentioning the main result(s) or the key method steps if contribution is a method.

✓(−) The abstract remains vague on part 4 (the impact) because no reader is targeted.

AND NOW FOR THE BONUS POINTS:

✓(+++) The reader is able to figure out the title of your paper, just by reading your abstract.

Chapter 12
Headings-Subheadings: The Skeleton of Your Paper

The skeleton gives a frame to the body. With it, the reinforced body takes shape; without it, the human would be a jellyfish. The skeleton of a paper is its structure. *The skeleton supports the various parts of the body according to their functional needs.* Composed of headings and subheadings set in a logical order, the structure reinforces the scientific contribution. *The skeleton is standard but it allows for variations in shape and size.* Headings are generally the same from one article to the next (introduction, discussion, conclusions) but subheadings differ. *The most sophisticated parts of the skeleton are also the most detailed (backbone, metacarpus, metatarsus).* The most detailed part of a structure contains the largest amount of contributive details.

> ### The Scientific Paper: 300 years of History
>
> *In an article published in The Scientist entitled "What's right about scientific writing," authors Alan Gross and Joseph Harmon defend the structure of the scientific paper against those who claim it does not represent the way science "happens". The structure, refined over more than 300 years, has enabled readers to evaluate the trustworthiness and importance of the presented facts and conclusions. The authors praise the standard narrative. They also observe that today, as a result of the increased role played by visuals, it is necessary to go beyond the interpretation of linear text.*

Four Principles for a Good Structure

A structure that plays its role follows these four principles:

1. Contribution guides its shape.

2. Headings and subheadings detailing the contribution are grouped.
3. Title words reflective of the contribution (not necessarily all title words) are repeated in its headings and subheadings.
4. It tells a story clear and complete in its broad lines.

Studying the structure of your paper allows you to identify important problems, such as an imperfect title, or a paper which is too complex, too detailed, too premature or too shallow.

Let us review the structure of the paper on slit arteriotomy. The words in italic are common to both title and structure.

Non-linear Finite Element Simulation to Elucidate the Efficacy of Slit Arteriotomy for End-to-side Arterial Anastomosis in Microsurgery[1]

1. Introduction
2. Mechanical factors underlying *slit* opening
3. Methodology for computer *simulation*
   3.1 Reference configuration for the *finite element* model
   3.2 Geometry details and boundary conditions of *the finite element* model in the reference configuration
   3.3 Hyperelastic material for the *arteries*
   3.4 *Simulation* procedure for the operation
4. Results and discussion
5. Conclusions

Recall that a title has two parts: the front part represents its contribution; and the back part, its context.

"Non-linear Finite Element Simulation to Elucidate the Efficacy of Slit Arteriotomy [Contribution] for End-to-side Arterial Anastomosis in Microsurgery [Context]"

---

[1] Reprinted from journal of biomechanics, Vol 39, Hai GU, Alvin Chua, Bien-Keem Tan, Kin Chew Hung, nonlinear finite element simulation to elucidate the efficacy of slit arteriotomy for end-to-side arterial anastomosis in microsurgery, pages 435–443, copyright 2006, with permission from Elsevier."

Principle 1: the contribution guides the shape of a structure

In the example above, three headings are standard: *Introduction, Results & discussion,* and *Conclusions.* Standard headings are disconnected from titles; they contain no title word. They are simply marks that show the location and the function of a part. In contrast, headings 2 and 3 are meaningful: they contain nearly half the title words, and relate directly to the contribution.

Heading 3 dominates this structure. With four subheadings, it provides much detail on the contribution. The subheadings organize the details in a logical order. All of this is to be expected, is it not? A structure should be the most detailed where the author has the most to write about, namely the scientific contribution of the paper. The structure has to expand to match the level of detail by offering more subheadings to help organize these details in a logical order, for the benefit of the reader and for the sake of clarity (☞1).
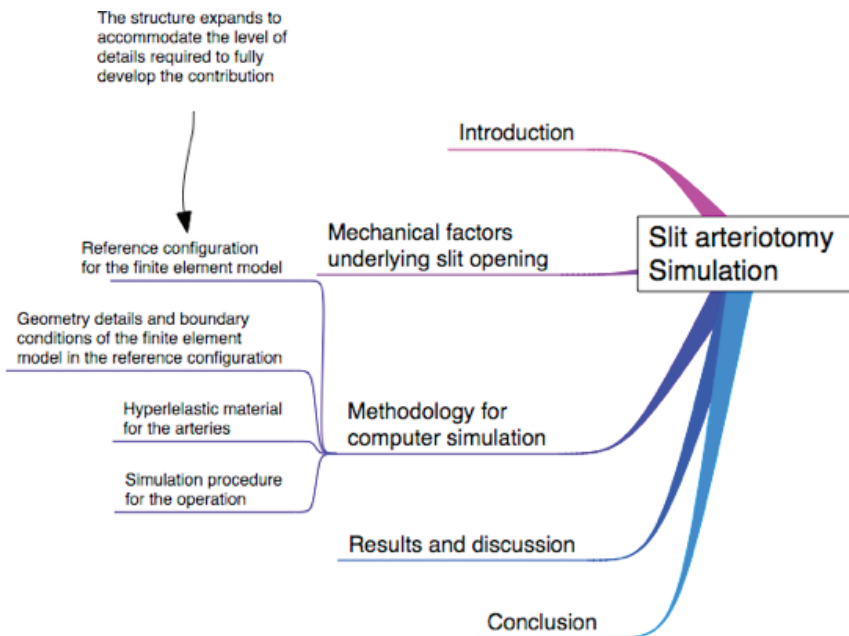

The structure expands to accommodate the level of details required to fully develop the contribution

Introduction

Reference configuration for the finite element model

Mechanical factors underlying slit opening

Slit arteriotomy Simulation

Geometry details and boundary conditions of the finite element model in the reference configuration

Hyperlelastic material for the arteries

Methodology for computer simulation

Simulation procedure for the operation

Results and discussion

Conclusion

Figure ☞1
Contribution is often found under the heading which has the deepest level of indentation, and the largest number of subheadings.

This first principle has a corollary: when excessively detailed parts do not contain much contribution, the structure lacks balance.

- A secondary part may be too detailed. Simplify or move details to appendix, footnotes, or supplementary material.[2]
- The knowledge level of the reader is underestimated. Remove details and provide references to seminal papers and books (☛2).
- Subheadings are "sliced and diced" too small. When a section with only one or two short paragraphs has its own subheading, it should be merged with other sections.
- The top-level structure is not divided into enough parts. For example, the background section is merged with the introduction. As a result, many subheadings are necessary within
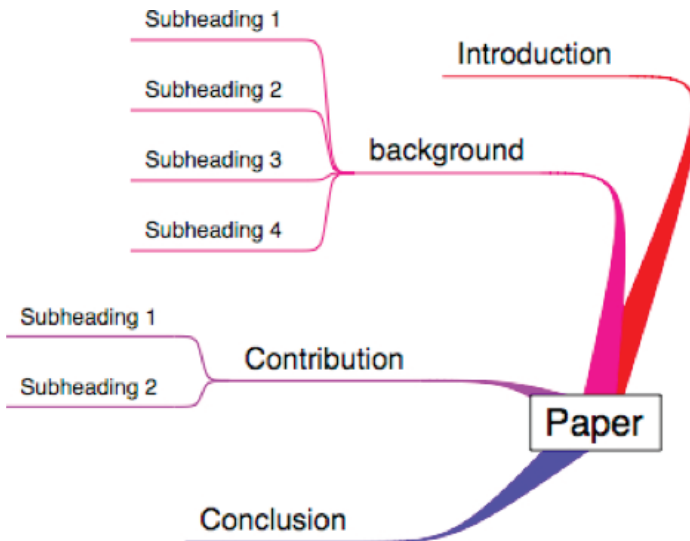
Figure ☛2
This structure points to one or several of the following problems: (a) background is too detailed; (b) contribution is small, therefore writer fills up paper with background; (c) Writer underestimates the knowledge level of the reader.

---

[2] The supplementary material is not published. It is sent to the journal alongside your paper to enable the reviewer to check your data at a higher level of detail than that used in the paper, to provide additional proof which you left out, or to verify a formula which you could not expand in detail for lack of space and ease of reading.

the introduction. Add headings at the top-level of your structure to reduce the number of subheadings.

- The paper has a multifaceted contribution that requires a large background and an extensive structure. Rewrite it as several smaller papers (☞3).
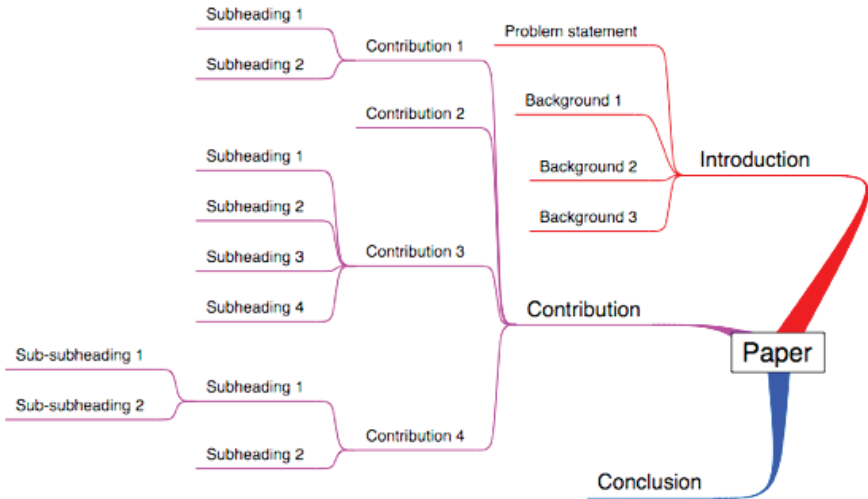

Figure ☞3
This structure points to one or several of the following problems: (a) the top-level structure has too few headings; (b) the contribution is too large for one single paper; (c) subheadings need to be merged.

Principle 2: Headings and subheadings detailing the contribution are grouped

When the branches containing the contribution are scattered instead of being grouped, the structure lacks singleness of focus (☞4).

- There may be more than one contribution in the paper; To make sure the paper is accepted, the writer stuffs his paper with contributions. The titles of such papers are very difficult to write! It is better to focus one paper around one contribution and have another paper (or letter) cover the additional contribution.
- The paper may not yet be ready for publication. The contribution is scattered in disconnected parts. The paper lacks unity and conciseness: repetition is unavoidable in such cases.
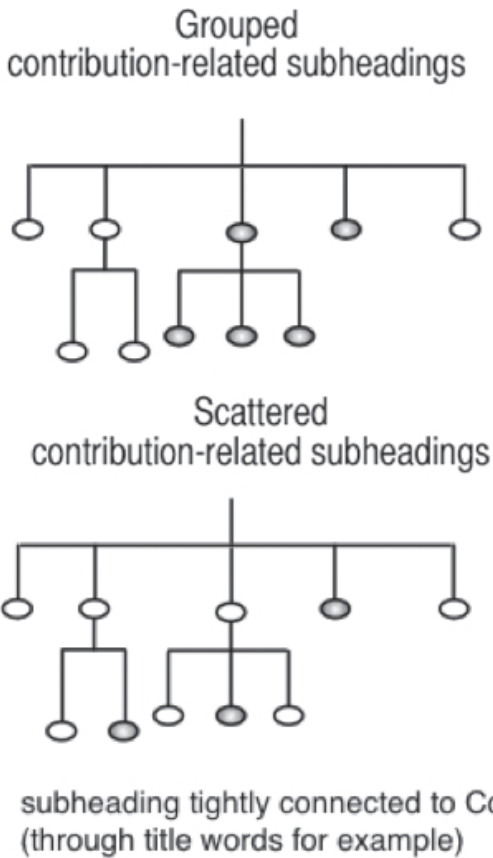
Grouped
contribution-related subheadings

Scattered
contribution-related subheadings

⬮ = subheading tightly connected to Contribution
(through title words for example)

Figure ☛4
Grouped headings and subheadings show that the contribution is well identified and unique. When headings and subheadings covering the contribution in detail are dispersed throughout the structure, the structure has problems.

- The writer may be unable to identify the main contribution of the paper, or unable to establish a priority between major and minor contributions. The paper and its title probably lack focus.

Principle 3: Title words describing the contribution are repeated in the headings and subheadings of a structure

A structure disconnected from its title is either unhelpful or indicative of a wrong title. Since the role of a structure is to help the reader

navigate inside your paper and identify where your contribution is located, a structure should have its headings and subheadings con- nected to the title (☛5).
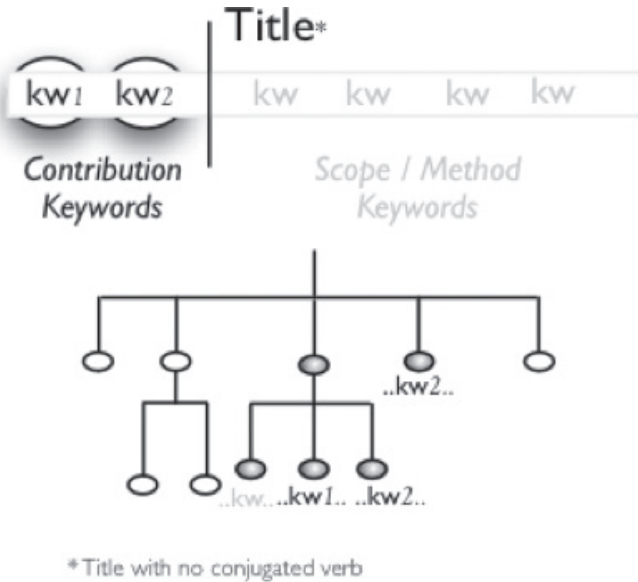
Figure ☛5
Title keywords reappear in structure headings and subheadings to allow direct reader access to key parts of the paper.

Let us apply the third principle to our sample structure and con- sider headings 2 and 3.

2. Mechanical factors underlying **slit** opening

Heading 2 contains '*slit*,' a title word found in the first part of the title describing the contribution. Heading 2 seems to be written for surgeons, whereas, heading 3 is definitely written for people on the IT side of mechanical engineering. Recall that the paper is published in the journal of Biomechanics, read by scientists from two different words — life sciences and engineering sciences — who may have difficulties understanding each other's work for lack of background knowledge. Therefore, it does make sense for the structure to address both types of readers.

Heading 2 takes us through the surgery steps of slit arteriotomy and the mechanically induced stresses and deformations observed

during the surgery. At this point, surgeons stop reading but mechanical engineers read on and find in heading 3 the model for the steps described in heading 2.

3. Methodology for computer **simulation**
   3.1 Reference configuration for the **finite element** model
   3.2 Geometry details and boundary conditions of the **finite element** model in the reference configuration
   3.3 Hyperelastic material for the arteries
   3.4 **Simulation** procedure for the operation

Heading 3 and its four subheadings contain '*simulation*' and '*finite element*,' two words located in the front part of the title (contribution part). They confirm that this heading, like heading 2, belongs to the contribution-related section of the paper. The author could have added '*nonlinear*' to strengthen the coherence between title and structure. The specificity of the words in heading 3 and its subheadings conveys to the surgeons the message that this section of the paper is not written for them.

This third principle has a corollary: When headings and subheadings are disconnected from the title of a paper, the structure OR the title may be wrong.

- The structure reflects the contribution better than the title. For example, a structure where the word '*trajectory*' appears in three of the five headings, and yet does not even appear once in the title, has an imperfect title.
- The structure is too cryptic. Its headings and subheadings are too generic, brief, or tangential. They do not give enough information on the contents. Revise the structure and reconnect it to the title.
- The structure contains synonyms of the title words, or specific keywords replacing the title's generic keywords or even an acronym made of title words. Having lost homogeneity and coherence, the article is less clear. Return to the original keywords, or change the title to make it more specific.

Principle 4: A structure tells a story that is clear and complete in its broad lines

According to this fourth principle, someone unfamiliar with the domain of computer simulations should be able to see the logic of the story after reading the title, the abstract, and the successive headings and subheadings.

Is this story clear?

1. Introduction
2. Mechanical factors underlying slit opening
3. Methodology for computer simulation
    3.1 Reference configuration for the finite element model
    3.2 Geometry details and boundary conditions of the finite element model in the reference configuration
    3.3 Hyperelastic material for the arteries
    3.4 Simulation procedure for the operation
4. Results and discussion
5. Conclusions

Heading 2 takes the reader into the operating theater, to observe the surgeon cut and stitch the arteries. Under the sharp blade of the scalpel, the arteries open; they deform under the pressure of fingers and the pull of stitches. Once the surgery is completed, the reader can imagine the blood flowing through the arteries, opening the slit wider.

Heading 3 provides details on the simulation.

Subheading 3.1 defines the initial state of the simulated objects.

Subheading 3.2 gives details on the model parameters (arteries, slit) and defines their limits.

Subheading 3.3 describes how the arteries, key objects in the simulation, are modeled.

Subheading 3.4 makes the simulation steps correspond to the steps of the actual surgery.

The story is coherent with what the title announces, but it is incomplete. There is no link between the model and the result (elucidation). This could easily have been achieved by replacing the standard heading "*Results and discussion*" with a more informative heading such as "*Elucidation of the efficacy of slit*

arteriotomy," thus establishing a direct connection between the model and its results, and encouraging the surgeons to read this section also.

The story may not be clear to surgeons unfamiliar with finite element modeling. This is clearly seen by the expert vocabulary used in headings 3.1 and 3.2. Six of their words (*reference, configuration, geometry, details, boundary, conditions*) are not even found in the abstract, whereas the words used by all other subheadings are found in the abstract.

The fourth principle has a corollary: when headings or subheadings reviewed in sequence tell a nonsensical story, the structure has holes or the title is wrong.

- The paper could be premature: its structure has not yet reached clarity. More work is needed until the structure falls into place.
- The story is nonsensical because it is not the story of the title, but another story. Change the title or rewrite the paper. You have the wrong face for the right body, or vice versa.
- The headings and subheadings are too concise or too cryptic — possibly because of the use of acronyms, synonyms, or highly specific keywords understood only by experts. Write more informative and understandable headings and subheadings.
- Key subheadings or headings are missing. Subdivide the obscure heading/subheading to reveal the missing connectors, or insert a new heading or subheading in the structure.

Syntactic Rules for Headings

Michael Alley, an author I designated as one of the giants who enabled me to see further afield in the world of scientific writing, advocates parallel syntax in structures. It does make the structure easier to read. Three styles are commonly found in headings and subheadings: Noun phrase such as '*Parameter determination*', verb phrase such as '*Determining the parameters*,' and (in life science papers mostly) full sentence such as '*The parameters are determined statically*'. To help the reader rebuild a story from its structure, headings at the same

indentation level or subheadings under the same heading should adopt a parallel syntax. In the sample structure, headings 2 and 3 are noun phrases. Within heading 3, all subheadings are also noun phrases.

In the following structure, the syntax is not parallel.

1. Introduction
2. Interference mechanism
3. Design rules
4. Proposing a solution
    4.1 Three layer prediction algorithm
        4.1.1 Algorithm classification
        4.1.2 Layer prediction comparison
5. Proposed recognition
6. Simulation studies
7. Discussion
8. Conclusions

This is not a good structure for many reasons. Focusing solely on the lack of consistency, one cannot miss the "one parent and only one child" problem: heading 4 has one subheading only 4.1 (no 4.2). The syntax also lacks consistency at the same heading level: headings 1, 2, 3, 5, 6, 7, and 8 are all single-noun phrases; but heading 4 starts with a present participle '*Proposing*,' thus breaking parallelism in syntax.

Purpose and Qualities of Structures

Purpose of the structure for the reader

- It makes navigation easy by providing direct access to parts of the paper.
- It helps the reader rapidly locate the section of the paper related to the author's contribution.
- It allows the reader to quickly grasp the main story of the paper by making a logical story out of the succession of headings and subheadings.
- It sets reading time expectations through the length and detail level of each section.

Purpose of the structure for the writer

- It reinforces the contribution by repeating Title/Abstract keywords in the headings or subheadings.
- It helps the writer divide the paper into informative, non-redundant, and logically connected sections that support the contribution.

Qualities of a structure

INFORMATIVE. No empty signposts should be found outside of the expected standard headings. Reading the structure headings and subheadings should be enough for readers to understand the title or to immediately identify a place of interest.

TIED TO TITLE AND ABSTRACT. Keywords from title and abstract are also in the structure. They support the contribution.

LOGICAL. Going from one heading or subheading to the next, the non-expert reader sees the logic of the order chosen by the writer. There are no logical gaps.

CONSISTENT at the syntax level. Each parent heading has more than one child subheading. Syntax is parallel.

CLEAR AND CONCISE. Neither too detailed nor too condensed.

Here is a very simple and productive method to ascertain the quality of your structure. "Flatten" your structure on a blank piece of paper. By this, I mean, write the title at the top of the page, and then write ALL headings and subheadings in the order they appear in your paper. Once done, underline the words structure and title have in common. Do you detect any discrepancy here? Are Contribution keywords from the title missing in the structure? Is it because you are using synonyms? Change your structure to bring it closer to the title. Are you using acronyms that are defined in the text and make your headings obscure? If so, remove them or give their full name. Do you have frequent structure keywords which are absent from your title? Shouldn't they be part of the title? Could it be that your structure is right and your title is wrong? If so, change the title.

Once you have examined how well the structure matches your title, have someone else read your flattened structure and explain to you what he or she thinks your paper contains. The less this person knows about your work, the better. Ask this person if the logic is visible in the succession of headings or subheadings. If the person is largely puzzled, you are not quite ready to publish yet. Rework your structure and your paper.

When the story is clear, give a quick syntactic check. Is the syntax of your headings parallel? Are subheadings orphans?

When the volunteer reviewer asks questions, do not start explaining! Remember that you will not be there to explain your structure to your readers once your paper is published. Just take note of the questions, and adjust your structure or title accordingly.

Structure Q&A

Q: The structure is like an outline. Shouldn't we start a paper by creating its outline first?

A: Some writers use the structure as a framework for writing. They create the structure in bullet point form or in outline form, and then expand the points. This method has value. It gives focus to a paper. If the story flows well at the structure level, it will probably flow well at the detailed level also. The writer may still change the structure later, but it will mostly be to refine the headings or to create more subheadings, not to totally restructure the flow of the paper. Actually, as a side comment, when asked if they know that Microsoft Word

includes an outlining program, most scientists say that they are not aware of such a capability! It is right there at your fingertips, an item under the *view* menu. It comes with its own set of tools. Play with it. I discovered outliners on the Apple II (a program called "*More*"). Life has never been the same for me since then. Even this book started its life as an 11 page outline!

Q: To keep my headings short, can I use acronyms?

A: Using acronyms in headings or subheadings, unless they are well-known, is troublesome for the majority of readers. Obscure acronyms prevent readers from making a story out of the subheadings. But this is true also of highly specific expert words. Write your headings and subheadings for non-experts. And if you absolutely need to use an expert word in a subheading, define it right away in the first sentence that follows that subheading. Doing so respects the principle of just-in-time background mentioned in the first chapter.

Q: What makes a good subheading?

A: A good subheading is an informative noun phrase whose key-words appear with high frequency in the paragraphs that follow it. It fits neatly and logically in between other subheadings. And it is closely related to the contribution, often by including words from the title and the abstract. A bad subheading contains acronyms, syn-onyms of the title keywords, or specific instances of title keywords (for example the title contains '*memory*,' and the subheading con-tains '*register*,' a specific keyword not even used in the abstract). The computer science expert knows the connection between memory and register, but the non-expert with a knowledge gap cannot make the connection.

Q: How many paragraphs come under a subheading?

A: That will depend on the length of your paragraphs, and on the flow of your paper. For example, if you describe a succession of steps, but each step is covered in one paragraph of two to four sentences, it may make more sense to group or collapse several steps under a more comprehensive subheading name. Let the specificity of the key-words in the subheading guide you. If the subheading is so specific

that its keywords do not even appear in the abstract, reconsider its usefulness or its wording.

Q: My journal dictates the labeling of the subheadings! Can I change them?

A: To enable the reader to reproduce your results, some journals (chemistry or life science journals come to mind) demand that you follow their imposed structure. If a journal dictates the number and types of subheadings, you have no choice but to comply.

Q: I wrote a letter, not a paper. My letter does not have headings, should I include some?

A: Not all papers have an explicit structure. When the paper is short, a letter for example, there is no need to add a structure: the structure is implicit. The "Introduction" heading is absent, but the first paragraph of the letter introduces. The "conclusions" heading is absent, but the last paragraph concludes. However, if you want to benefit from the tests and the metric in this chapter, I recommend that you artificially create and write down such a structure based on the contents of your paragraphs, even though it won't be part of your letter.

Q: My structure has standard headings (intro, results, ...). Should I include more subheadings?

A: If your paper has enough pages, you should at least add two subheadings (or more) under the heading that contains the bulk of your contribution. It facilitates reading.

Q: I have read the structure of my friend's paper, and found it very difficult to connect it with the title of his paper. What is wrong: the title, or the structure?

A: It could be both. If you were not able to retell the story of the title just by browsing the headings and subheadings of your friend's paper, you may trace the problem back to five main sources: (1) synonyms (title and structure use different keywords meaning the same thing), (2) Acronyms (defined elsewhere), (3) uninformative headings/subheadings (hollow markers such as experiment, (4) logical gaps, or (5) unintelligible expert keywords.

Q: My paper is a review. Do the principles detailed here also apply to this kind of paper?

A: A review paper is quite different, but principle 4 (the story line) should still be respected. Look at the resources provided in the final book chapter to find guidelines for review papers.

Structure Metrics

✓(+) All contribution-related search keywords in the title are also in the headings/subheadings.

✓(+) The contribution is grouped under successive headings.

✓(+) The structure contains informative subheadings

✓(+) Not one heading/subheading could change place without compromising the structure of the paper.

✓(+) The structure does not contain acronyms, synonyms, or keywords only understood by experts only.

✓(+) The majority of your structure words are in the abstract.

✓(+) The structure does not contain orphan headings or subheadings.

✓(−) Contribution-related search keywords from title are missing in your structure.

✓(−) The contribution is scattered throughout the structure.

✓(−) The structure does not contain any subheadings, or any informative headings or subheadings.

✓(−) Headings/subheadings could change place without compromising the structure of the paper.

✓(−) The structure contains acronyms, synonyms, or specific expert keywords.

✓(−) Less than 50 % of your informative structure words are in the abstract.

✓(−) The structure contains orphan headings or subheadings.

AND NOW FOR THE BONUS POINTS:

✓(+++) Even a non-expert could figure out the title of your paper, just by reading the structure.

This page intentionally left blank

Chapter 13
Introduction: The Hands of Your Paper

Extended hands welcome and invite to enter. They guide someone unfamiliar with a new place. The introduction of a paper plays a similar role. It greets, provides guidance, and introduces a topic not familiar to the reader. *Hands point to something worthy of attention, and invite the eyes to follow.* The introduction also points to the related works of other scientists and to your contribution.

For many, the introduction is a necessary evil, something more difficult to write than the methodology or results section. Therefore, to ease the burden, the scientist usually keeps it brief. Alas, brevity is only appreciated by the few experts in the field already familiar with the introduction material. The *many* readers (a reasonable 40 %) with a significant knowledge gap will not be satisfied. Reviewers could even be among them. Ask a reviewer how many papers he or she accepts to review without full expertise in the matter being reviewed, and be ready to be surprised! Therefore, write an introduction that bridges their knowledge gap, otherwise they may not be able to evaluate your paper correctly. Remember that they have veto power over the selection of your paper for publication.

Writing Against All Odds

Vladimir, rejected paper in hand, was reading the online comments from the three people who had reviewed his paper. From the remarks, Vladimir could tell that two of them were knowledgeable in his field. It was the remarks of the third reviewer that bothered him. Clearly, that was not someone from his field. He had presumed all of them would be experts.

Popov, his supervisor, was passing by. Vladimir called him.

"Hey, boss, aren't you supposed to be an expert in the field before you are invited to review papers?"

"Often the case. May be 80 % of the time."

"You mean you have reviewed papers that are outside of your direct field of expertise?"

"Sorry, Vlad, no time to chat, have to run off to a meeting."

Left alone, Vladimir did a quick mental calculation. Let's see. If I have three reviewers, what is the probability that I get at least one reviewer who is not an expert if each reviewer's probability to be an expert is 80 %... Hum, that would be... point eight times point eight, that's point sixty-four times point eight again, and that's...sixty times eight, four hundred and eighty, plus four times eight, thirty-two, that makes ...51.2 % chances that all reviewers are experts. So I have one chance in two that at least one of my three reviewers is not an expert!

On his way back shortly after, Popov stopped in front of Vladimir's cubicle. "Meeting cancelled," he said. What is it you wanted to know?"

"Are you sure about your 80 %?" asked Vladimir.

"80 % what?" His boss had already forgotten.

"80 % chance to be a direct expert in the topic of the paper you are asked to review."

"That sounds about right. It depends on the journal of course. For large journals, the percentage may be even lower."

"And out of three reviewers, how many have to recommend your paper for it to have a chance to be published?"

"Again, that will depend on the journal, but for prestigious journals, I'd say, all three."

"Great! Vladimir mumbled. That means that I have to write my paper to be understood by amateurs!"

"Are you calling me an amateur, Vladimir, ... because in front of you stands the amateur who has reviewed many papers where he did not considered himself to be an expert in everything that was written."

"Oh, No, Boss... You're a pro, right John?"

From across the cubicle partition, John's voice shouted:

"A real Pro, Boss!"

The Introduction Starts Fast and Finishes Strong

The introduction engages. Do not keep the reader's brain idling for too long. Your reader is eager to go places. How do you know for sure? Well, the impatient reader having gone past two filters, the title (coarse grain) and the abstract (fine grain), decided to click on the download button. Interest must be high! Do not bore or delay the reader. Do not disappoint with a false start that slows the pace of your introduction.

The vacuous false start

> *In the age of genomes, large-scale data are produced by numerous scientific groups all over the world.*

> *Significant progress in the chemical sciences in general, and crystallography in particular, is often highly dependent on extracting meaningful knowledge from considerable amount of experimental data. Such experimental measurements are made on a wide range of instruments."*

> *Because of the long-term trend towards smaller and smaller consumer goods, the need for the manufacture of micro components is growing.*

Was there anything in these examples you did not already know? Catch and ruthlessly exterminate these cold starts, these hollow statements where the writer warms up with a few brain push-ups before getting down to the matter at hand. By the way, if you are such a writer, do not feel bad. Many of us are. It takes a little while for our brain to output coherent and interesting words when we face a blank page or a blank screen! Since we have a propensity to do brain push-ups at such times, let's make sure to remove the sentences they produce.

Here is another type of false start. At first glance, there seems to be nothing wrong. After all, the writer tries to conjure up excitement by showing how massively important the current problem is.

The considerable false start

> *There has been a surge, in recent times, towards the increasing use of …*

> *There has been considerable interest in recent years in this technology, and, as trends indicate, it is expected to show continuing growth over the next decade …*

169

In this type of false start, the author sees the heat of a research field as being sufficient to warm the reader to his contribution. The words used to raise the temperature are *'exponential,' 'considerable,' 'surge,' 'growing,' 'increasing,'* and other ballooning words. They lead an important class of readers, the reviewers, to suspect a "me-too" paper. There may have been a recent surge but the writer is obviously running behind the pack of researchers who created the surge. People who are pioneering research make the past, they don't catch up with it!

Reviewers may also suspect that the writer is attempting to influence their judgment by equating importance of problem to significance of solution. If many people consider the problem important, does that make the contribution an important one? There is no necessary cause and effect relationship here. If I met the President of the United States, does that make me a great man? If my neighbor has been arrested for robbery, does that make me a felon?

The right start

It is best to start with what readers expect: they want more details. Your title is sketchy. It draws the silhouette of a face. Your abstract puts a bright but narrow frontal spotlight on the face, giving it a flat, even look and making everything else recede into the shadows. The introduction is the soft filling light. It adds dimension, softens the shadows, and reveals the background. To use a metaphor familiar to photographers, the abstract privileges speed of capture whereas the introduction emphasizes depth of field.

At the start of the introduction, since the face is still in full sight of a reader, it is best to remain close to the title and frame it in a context. The context, be it historical, geographical, and even lexical (definitions), should not lose its relationship with the face. Returning to the photo metaphor, the reader should constantly see the face in the viewfinder even when the perspective widens. Too many papers paint a landscape so vast in the introduction that the reader is unable to place the title in it.

Here is a fast and focused start combining definition and historical perspective.

Name Entity Recognition (NER), an information extraction task, automatically identifies named entities and classifies them into predefined classes. NER has been applied to Newswires successfully [references]. Today, researchers are adapting NER systems to extract biomedical named entities — protein, gene or virus — [more references] for applications such as automatic build of biomedical databases. Despite early promising results, NER's ability to apply to such entities has fallen short of people's expectations.

After reading this paragraph, the reader expects the writer to explain why success is limited, and to bring an answer to the main question: what adaptations to the original NER would enable biomedical named entities to be extracted more successfully.

The dead end

At the Doorstep — The Script Hollywood Rejected

Steve Wilkinson is an insurance agent for a large insurance company. He is also Vladimir's neighbor. For weeks, he has been chatting across the fence with Vladimir's wife, Ruslana (about insurance of course), while she prunes her rose bushes in the garden. She finally tells Vladimir who, being quite good at maths and curious to find out about the benefits of insuring his family, decides to accept his neighbor's invitation to hear about his company's insurance products. A date is fixed: next Thursday after work.

Thursday evening. Vladimir shows up at the front door of Mr. Wilkinson and rings the bell. Steve goes to open the door, and immediately enters into a monologue at the doorstep before letting Vladimir in.

"Recently, an increasing number of people buy insurance because of global warming. Global warming may lead to life and property threatening weather. AIE, and PRUDENTA offer insurance schemes in this domain. A different scheme is examined today. After four years of insurance premiums, the GLObal WARming BLanket Insurance Scheme, GLOWARBLIS, provides a 5.6% yearly interest on the insurance premiums, assuming the global warming index remains stable four years in a row. A 1-point fluctuation in the global warming index during this period will decrease the yearly interest rate by a corresponding amount. After entering this house, go through the corridor leading to the living room where the prospectus will be examined. Following this, proceed to the office to discuss the signing of the contract. After signature, return to this door to leave with a package presenting other insurance schemes from the company."

Table-of-content type endings have no place in an introduction where readers can just flip a few pages and discover the whole structure by rapidly scanning the headings and subheadings. Therefore, don't end your introduction as seen below.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 presents the technology and shows how our approach is conducted using our scheme. Section 4 presents the results of our experiments and shows how the efficiency and accuracy of our approach compares with others. Finally, we offer our conclusions and discuss limitations.

The rest of this paper is organized as follows. Section 2 describes some related works, in particular similar work that has been done. Following that, the proposed approaches are discussed in section 3, with the implementation details being discussed in section 4. Section 5 evaluates the performance and compares the proposed approaches to a baseline model. Finally, we draw conclusions and outline future work in section 6.

When a place is small and more or less standard — for example the house of the insurance agent in the Vladimir story — there is no need to describe the various rooms and tasks that lay ahead. But when the place is large and not standard (the White House in Washington, the Imperial Palace in Beijing, or the Versailles Castle in France), it is wise for the guide to give a brief overview of the visit (schedule, route followed…) before entering. A Ph.D thesis, or a book is large enough to deserve an introduction that helps the reader anticipate what are the main parts and what they accomplish.

The strong finish

At the Doorstep — The Script Hollywood Accepted

Steve Wilkinson is an insurance agent for a large insurance company. He is also Vladimir's neighbor. For weeks, he has been chatting across the fence with Vladimir's wife, Ruslana (about insurance of course), while she prunes her rose bushes in the garden. She finally tells Vladimir who, being quite good at maths, and curious to find out about the benefits of insuring his family, decides to accept his neighbor's invitation to hear about his company's insurance products. A date is fixed: next Thursday after work.

Thursday evening. Steve sees Vladimir approaching his house. He goes to open the door even before Vladimir has a chance to ring the doorbell, and warmly greets his neighbor.

"Vladimir, I'm so glad you could come. Today is a day you will remember as the day you found peace by sheltering your family from the financial worries that global warming brings with its devastating tornadoes and countless other financial calamities. Come in, come in."

The best ending to an introduction is in this second Hollywood-approved scenario: Vladimir is told about the outcome of signing the insurance contract, and then moves right in. End your introduction with the expected post-contribution outcome of your research to keep reader motivation high.

Read the first paragraph of your introduction. Is it vacuous or considerable? If it is, delete it. Is the last paragraph redundant with the structure? If it is, delete it.

The Introduction Answers Key Reader Questions

What is the main question of your paper? It is the question that you answer by stating your contribution. If you cannot phrase this question, you are not ready to write your paper because you do not yet have a clear idea of your contribution. To help you determine the main question, practice on the following familiar titles:

"Non-linear Finite Element Simulation to Elucidate the Efficacy of Slit Arteriotomy for End-to-side Arterial Anastomosis in Microsurgery"

Main question:

Having observed the efficacy of slit-arteriotomy in terms of blood flow and robustness of anastomosis for patients of various age and blood pressure conditions, can this efficacy be explained by modeling the mechanical stresses sustained by the arteries during and after the operation?

"Energy-Efficient Data Gathering in Large Wireless Sensor Networks"

Main question:

How can a sensor node be chosen to forward data in a large wireless sensor network so that total energy consumption for data forwarding in the network is minimized?

Read your title and abstract. Write the main question they answer. Is that question clearly stated in your introduction? If there is more than one question, you may have a paper with multiple contributions, and possibly a paper that could be divided into multiple papers. Or, you may not yet clearly perceive your contribution.

As soon as you identify the main question answered by your paper, include it in your introduction. And here is an "Oh, by the way" remark that may turn out to be the only help your introduction needs to be close to perfect. ***"Oh, by the say, have you thought of putting this question to the reader in visual form?"*** As you know by now, nothing is better than a visual to demonstrate and convince. So make sure you spend the time to clarify first in your mind for yourself, and then on paper for the reader, that "mother of all questions".

From that main question comes a mother lode of other questions: Why this? Why now? Why this way? Why should the reader care (how is your work relevant to the reader's needs)?

> ### The Questionable Cake
>
> *That afternoon, Vladimir Toldoff received a call from his wife Ruslana as he was finishing an experiment in the lab.*
>
> *"I am coming with a cake, the cake knife, and enough plates and cutlery for four or five," she announced.*
>
> *He answered, "What? Wait! First, what is the occasion? And why now, can't it wait until tonight? And by the way, what cake is it, and why do you want to cut it in the lab? You know that crumbs are not welcomed here."*
>
> *The rapid fire of questions did not faze Ruslana. She knew her Vladimir. A full-fledged scientist. She paused and rephrased his questions succinctly.*
>
> *"All right, let me see. You would like to know why a cake, why eat it now, why its mouth-watering taste should make you shout 'Darling come right away,' and why I should slice it in the lab instead of at home, am I right?"*
>
> *Vladimir grinned. He was quite impressed with his wife's listening skills.*
>
> *"That's right, Mrs Toldoff," he responded.*
>
> *Ruslana then uttered four words that had him shout for joy: "Your Medovik birthday cake".*

Through these *why* questions, the reader expects you to justify your research goals, your methodology, as well as the timeliness and value of your contribution.

The following example taken from a life science paper illustrates how the writer answered these questions; it contains acronyms or jargon you may be unfamiliar with, but that should not deter you! The paper's title is the following:

> *"A gene expression-based method to diagnose **clinically distinct** sub-groups of Diffuse Large B Cell Lymphoma" (DLBCL)*

Why now? In this case, because recent studies present diverging results.

> *"We were curious to see whether we could resolve the discrepancy between these gene profiling studies by using our current understanding of the gene differences between GCB and ABC DLBCL."*[1]

---

[1] George Wright, Bruce Tan, Andreas Rosenwald, Elain Hurt, Adrian Wiestner, and Louis M. Staudt, a gene expression-based method to diagnose clinically distinct

Why this? In this case, because there is a need for labs to deliver a coherent clinical diagnosis regardless of the platform used to measure gene expression.

> *"As was pointed out (3), it is a challenging task to compare the results of these profiling studies because they used different microarray platforms that were only partially overlapping in gene composition. Notably, the Affymetrix arrays lacked many of the genes on the lymphochip microarrays …"[2]*

Why this way? In this case, because the method works on different microarray platforms.

> *"For this reason we developed a classification method that focuses on those genes that discriminate the Germinal Centre B-cell like (GCB) and the Activated B-Cell like (ABC) Diffuse Large B Cell Lymphoma (DLBCL) subgroups with highest significance."[3]*

Why should the reader care? In this case, because it predicts survival regardless of the experimental platform.

> *"Our method does not merely assign a tumor to a DLBCL subgroup but also estimates the probability that the tumor belongs to the subgroup. We demonstrate that this method is capable of classifying a tumor irrespective of which experimental platform is used to measure gene expression. The GCB and ABC DLBCL subgroups defined by using this predictor have significantly different survival rates after chemotherapy."[4]*

To these reader questions, the reviewer adds other questions. Even though they overlap, they differ in important ways.

1) Is the problem clear, and is solving it useful?
2) Is the solution novel, and is it much better than others?

Therefore, you should have both reader and reviewer in mind when you write your introduction. It is up to you to convince them that the problem is real, and that your solution is original and useful.

---

subgroups of diffuse large B Cell Lymphoma, Proceedings of the national Academy of Sciences, August 19, 2003 Vol. 100 No.17, www.pnas.org, copyright 2003 National Academy of Sciences, U.S.A.

[2] *ibid.*

[3] *ibid.*

[4] *ibid.*

The Introduction Frames Through Scope and Definitions

The writer's intellectual honesty is demonstrated in many ways. One of them is a clear and honest description of the scope. Readers need to know the scope of your work because they **want** to benefit from it, and therefore need to evaluate how well your solution might work on their problems. If the scope of your solution covers their area of need, they will be satisfied. If it does not, at least they will know why, and they may even be encouraged to extend your work to solve their problems. Either way, your work will have been helpful.

Scope

In essence, the scope or frame of your contribution is set by the methodology, the data, the time frame, and the application field. Establishing a frame around problem and solution enables you to claim with some authority and assurance that your solution is "good" within that frame. Some writers leave the framing until later in the paper. I believe that a reader informed on the scope early is better than a reader disappointed by late disclosures restricting the applicability and the value of your work. Therefore, establish the scope early in your paper.

Use the answer to the "Why this way" to make the scope more precise while enhancing the attractiveness of your paper and highlighting its novelty.

> *Our method does not need a kernel function, nor does it require remapping from a lower dimension space to a higher dimension space.*

> *Our dithering algorithm does not make any assumption on the resolution of pictures nor does it make any assumption on the color depth of the pixels.*

Not all assumptions in your paper affect the scope. Those that do not, are best presented just in time in the paragraphs in which they apply (often in the methodology section). Justify their use, or give a measure of their impact on your results, as in the following three examples: (1) *Using the same assumption as in [7], we consider that...; (2) Without loss of generality, it is also assumed that ...; (3) Because we assume that the event is slow varying, it is reasonable to update the information on event allocation after all other steps.*

Define

Another way of framing is by defining. In the following example, the authors define what an *'effective'* solution is. They do not let readers decide on their own what *'effective'* means.

> *"An effective signature scheme should have the following desirable features:*
>
> - *Security: the ability to prevent attacked images from passing verification;*
> - *Robustness: the ability to tolerate ...*
> - *Integrity: the ability to integrate ...*
> - *..."*[5]

When you define, you frame by restricting the meaning of the words to your definition. Demonstrating a solution is *effective* because it fulfills predefined criteria is easier than demonstrating a solution is *effective* when the evaluation criteria are up to the reader.

Scope and definitions ground the readers' expectations and set strong foundations for the credibility of your results. In the next chapters two other credibility-enhancers will be considered: citations and precision.

The Introduction Is a Personal Active Story

Personal

Your contribution is woven into the work of other scientists (the referenced related work). Clearly identify what is yours and what belongs to others by using the personal pronouns *'we'* or *'our'*.

Naturally, using personal pronouns changes the writing style. Thankfully, there is not **one** scientific writing style, as many believe. There are two: the personal and the impersonal style. There are two occasions in a business meeting when a more personal style is appropriate: at the beginning of the meeting and at the end. Both start and end with a bow or a handshake. In a scientific paper, the introduction

---

[5] Shuiming Ye, Zhicheng Zhou, Qibin Sun, Ee-chien Chang and Qi Tian, a quantization-based image authentication system. ICICS-PCM, Vol. 2: 955–959. © I.E.E.E 2003

and the conclusions correspond to these occasions. The direct style supports the role each part plays. The introduction reinforces the motivation of the reader to read the rest of the paper, by being low on difficult content, and high in relationship with the active voice and personal pronouns. The conclusions reinforce the motivation of the reader to put to use what was discovered in the paper, again by being low on difficult content simply because all the details have been presented earlier in the paper.

The Story of Vladimir Toldoff

"Vladimir!"

The finger of Popov, his supervisor, is pointing at a word in the third paragraph of Vladimir's revised introduction.

"You cannot use 'we' in a scientific paper. You are a scientist, Vlad, not Tolstoy. A scientist's work speaks for itself. A scientist disappears behind his work. You don't matter Vlad. 'The data suggest'… you cannot write "our data". It's THE data, Vlad. Data do not belong to you. They belong to Science! They speak for themselves, objectively. You, on the other hand, will only mess things up, introduce bias and subjectivity. No Vlad, I'm telling you: stick to the scientific traditions of your forefathers. Turn the sentences around so that you, the scientist, become invisible. Write everything in the passive voice. Am I clear?"

"Crystal," Vladimir responds, "But I was only taking the reviewer's comments into account."

And with that, he hands out a printed copy of the comments he had received from the reviewer.

Popov grabs the paper.

"What kind of nonsense is this?", he says.

(reading the letter aloud)

…Your related work section is not clear. You write, "The data suggest". Which data? Is it the data of [3], or is it your data? If you want me to assess your contribution fairly, you should make clear what YOUR work is and what the work of others is. Therefore, if it is your data, then write, "our data suggest". Also, if I may make a suggestion, I feel that your introduction is somewhat impersonal and hard to read. You could improve it by using more active verbs. That would make reading easier….

"Ah, Vladimir! No doubt, this comes from a junior reviewer. What is happening to Science!"

Even if you are the first author of the paper, research is, for the most part, a collaborative effort. It is therefore appropriate to use personal pronouns such as *we* or *our*, leaving the personal pronoun *I* for later in your professorial career when you write papers alone. Some object to the use of personal pronouns, and advocate the passive voice to give the paper a more authoritative disembodied voice, but I will argue that it is difficult to welcome a reader into the body of your research this way. For a fun read on the scientific writing style, do read the bonus book chapter entitled "Scientists Anonymous" on the website: www.scientific-writing.com

> ### The Story of the Passive Lover
>
> *Imagine yourself at the doorstep of your loved one. You are clutching, somewhat nervously, a beautiful bouquet of fragrant roses behind your back. You ring the doorbell. As your loved one opens the door and gives you a beaming smile, you hand out the bouquet of flowers and utter these immortal words:*
>
> *"You are loved by me."*
> *What do you think happens next?*
> *(a) You eat the flowers; or*
> *(b) You ring the doorbell again and say the same thing, this time, using the active voice.*

Active story

Avoid the passive voice when it makes the subject disappear causing the reader to no longer know who does what (authorship disappears). In the introduction, the passive voice has a damping effect. Stories don't rely on the passive voice.

The introduction frames your contribution, and answers questions through the story of the four 'Why's. It is a story, not a report. And indeed, the active voice is the voice used in story telling.

> *"**We** were **curious** to see whether **we** could resolve the discrepancy between these gene profiling studies by using **our** current understanding of the gene differences between GCB and ABC DLBCL."*[6]

---

[6] *ibid.* 1, at 175

See how abstract and introduction differ in writing style.

Abstract.

"The GCB and ABC DLBCL subgroups identified in this data set had **significantly different 5-yr survival rates after the multiagent chemotherapy (62% vs. 26%; P = 0.0051),** *in accord with analyses of other DLBCL cohorts. These results demonstrate the ability of* **this gene expression-based predictor** *to classify DLBCLs into biologically and clinically distinct subgroups irrespective of the method used to measure gene expression."*[7]

Introduction.

*"**We** demonstrate that this method is capable of classifying a tumour irrespective of which experimental platform is used to measure gene expression. The GCB and ABC DLBCL subgroups defined by using **this predictor** have **significantly different survival rates after chemotherapy**."*[8]

The abstract is more precise than the introduction when it comes to the key numerical results. But, the factual abstract does not tell a personal story: '*these results demonstrate*' is impersonal whereas '*We demonstrate*' is active and personal. The passive voice is quite acceptable in the rest of your paper where knowing who does what does not really matter. But in the introduction, the active voice rules!

Did you use the pronoun 'we'? Did you answer all the whys? Identify where each why is answered. Do you bridge the knowledge gap maybe with some definitions? Did you build interest in the story? If not, why not? Did you cut and paste text between your abstract and your introduction? If you did, rewrite. To identify whether you adequately scoped your problem and solution, simply underline the sentences that deal with scope, assumptions and limitations. Are there enough of them? Are they at their proper place? Finally, did you mix introduction with technical background? If you did, place your technical background after the introduction under a separate heading. The introduction captures the mind; the technical background fills it. These two functions are best kept separate.

---

[7] *ibid.*

[8] *ibid.*

This page intentionally left blank

Chapter 14
Introduction Part II: Popular Traps

The introduction helps the reader understand the context from which your research originated. Do you borrow or adapt the work of other scientists to reach your objective, or do you follow a completely different research path? The reader wants to know. Positioning your work on the research landscape is a perilous exercise because it is tempting to justify your choices by criticizing the work of others. It is tempting **not** to compare your work with the work of others so as not to upset anyone. It is equally tempting to borrow other people's ideas and conveniently forget to tell the reader where these ideas came from. Temptation is plaguing the scientist writing a paper simply because the stakes are high: publish **or *perish***.

Four traps are laid in the path of writers: the trap of the story plot, the trap of plagiarism (using someone else's words without proper quotes and acknowledgment), the trap of imprecision, and the trap of judgmental words.

The Trap of the Story Plot

The introduction tells the personal story of your research. All good stories have a story plot to make them interesting and clear.

> A story
>
> *I'm so excited to share this great story with you. My father [1] is on the front lawn cleaning the lawn mower. My sister [2] is in the kitchen making a cake. My mum [3] has gone shopping, and I am playing my electric guitar in my bedroom.*
>
> *Do you like my story? No? It is a great story. What's that you're saying? My story has no plot? Of course there is a plot! See, it describes my family's activities, starting with my father. We all have something in common: family ties, living under one roof…*

If this story left you cold, the analogous story found in scientific papers will also leave the reader cold. In short, the story says: in this domain, this particular researcher did this; that research lab did that; in Finland, this other researcher is doing something else; and I'm doing this particular thing. The problem with this type of story is that the relationship between their work and your work is not stated. In symbolic graphical form, the story plot would look like ☛**1a**. The pieces are juxtaposed, not linked.

Father Mother Sister Me
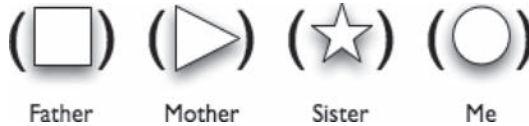
Figure ☛**1a**
All story elements are juxtaposed, disconnected: father, mother, sister, brother.

Contrast the first story with this one:

> *I'm so excited. I'm going to tell you a great story. My father [1] is on the front lawn cleaning the lawn mower. And do you know what that means? Trouble! He hates it. He wants everyone to help bring this or bring that to make himself feel less miserable. When that happens, we all run away, not because we refuse to help him, but because he wants us to stand there and watch idly while he works. So my sister [2] takes refuge in the kitchen and plunges her hands in flour to slowly make a cake. My mum [3] suddenly discovers that she is missing parsley and rushes to the supermarket for an hour or so. As for me, I escape to the upstairs bedroom and play my electric guitar with the volume cranked up to rock concert levels.*

The difference is striking, isn't it? In an interesting story plot, all parts are connected as in ☛**1b.**

Mother

Me
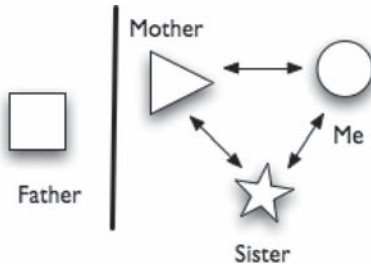
Father

Sister

Figure ☛**1b**
Three story elements (mother, sister, me) share a common bond. This bond isolates the square element (father).

Here is a second story based on a story plot often found in scientific papers.

> **A terrible story**
>
> *I'm so excited. I'm going to tell you my second best story. A red Ferrari [1] would take me to the house of Vladimir Toldoff in 5 hours. It is fast. However, it is very expensive [2,3,4]. A red bicycle [5] is much less expensive and quite convenient for short trips. So, if Vladimir Toldoff comes to live near my house, it will be quite cost effective [6]. However, bicycles require accessories like mudguards or bicycle clips [7] to keep trousers clean. Red athletic shoes [8], however, require no accessories, and are as good a solution as a bicycle to travel over short distances [9]. However, their look is easily degraded [10] by adverse weather. On the other hand, brownish open plastic sandals [11] do not have any of the previous problems: they are cheap, weatherproof, convenient, and require no accessories. Furthermore, they are easily cleaned, and fast to put on. However, contrary to the Ferrari, they reflect poorly on the status [12] of their owner. Therefore, I am working on a framework to integrate self-awareness into means of transportation, and will validate it through the popular SIMs2 simulation package.*

Yes, I have exaggerated (only a little), but you get the point. The *however* plot, after taking readers through a series of sharp *however* turns, completely loses and confuses them. The seemingly logical connection between the elements is tenuous, as in ☛**2**.

Figure ☛2
4 shapes: a sun, a star, a cross, and an ellipse. The first element is compared with the second, the second with the third, and so on. At the end, the final element is connected back to the original element, thus completing the loop. Yet, the sun is never compared with the cross and the star is never compared with the ellipse. For the chain of comparisons to be meaningful, the comparison criteria must be identical for all elements, and all elements must be compared.

On the way to the last proposal (the writer's contribution), a long list of disconnected advantages and disadvantages is given; by the time readers get to the end of the list, they innocently (and wrongly)

assume that the final solution will provide all the advantages and none of the disadvantages of the previous solutions. Unfortunately, the comparison criteria continuously vary, and therefore, nothing is really comparable.

Both plots, the juxtaposed story plot and the meandering story plot, are often found because they are convenient from a writer's perspective. There is no need to spend hours reading the papers referenced, reading their titles is enough (and you can find these in the list of references of other papers anyway), at a push the writer will read their abstracts, no more.

Are there better plots? Assuredly, but giving examples would fill too many pages, so here they are in schematic form ☛3.

Taking someone's work, validating and extending it

Ironing out the quirks of someone's work to make it work better (new method may be)

Combining work to increase applicability and reduce individual shortcomings

Completing the missing piece to increase usefulness

Making past work more general by removing constraints and increasing applicability

Independently confirming (or not) someone else's experimental or theoretical results.

Isolating a subset of data for which an unsuccessful method is now successful.

Figure ☛3
Various schematic story plots that work.

I have found that a popular movie plot is also useful in scientific writing. The author shows you how the story ends even before it starts. When readers have the full picture, they are better able to situate your work in it. They understand how and with whose help you will achieve your results. In addition, they know the scope of your work and are clear about the future work. Graphically, it is represented in ☛4.
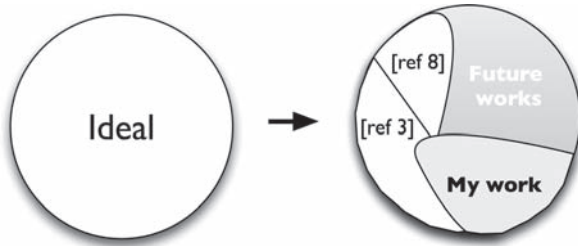
Figure ☞ 4

First the ideal system or solution is depicted (here represented by the circle). Then the story tells how this ideal picture comes together: what the author contributes, what others have already contributed [referenced work], and what remains an open field of research (future works). Everything is clear, everything fits nicely, and the reader is more easily convinced of the worth of your contribution.

Identify your story plot. Does it look like a "however" meander or a series of juxtaposed disconnected elements? Is your story easy to follow? Does it flow logically: from past to recent, from general to specific, from specific to general, from primitive to sophisticated, from static to dynamic, from problem to solution, or from one element in a sequence to the next in line?

The Trap of Plagiarism

Plagiarism exists when someone else's words are found in your paper without proper quotes **and** references. Senior researchers, whose names often appear as the third or fourth author in a paper, do not need to be told. Their reputation is at stake. They know only too well the hefty price one pays when caught. They have heard the tale of the faculty dean high up in the research ladder who had to resign because someone found out that he had plagiarized in a paper written 20 years earlier when he was still a junior researcher.

Vladimir Toldoff told off again

"Vladimir!"

The finger of Popov, his supervisor, points to a sentence in the third paragraph of Vlad's introduction in the paper published three months earlier in a good journal.

"Yes, anything wrong?"

> *"The English in this paragraph about Leontiev's algorithm is too good. These are not your sentences."*
>
> *"Um, let me see. Ah, yes, it is rather good, isn't it! I must have been in great shape that day. I remember noticing how well I had worded that paragraph when I cut and pasted it into my paper from my reading notes."*
>
> *"Would it be too much to ask you to bring your reading notes?"*
>
> *"You have access to them already. I emailed you the files after the review meeting last month."*
>
> *"Oh yes. That's right. Give me a moment… Ah! Here are your notes on Leontiev's work, and here is that sentence. Now let me retrieve Leontiev's paper from the electronic library. Just a minute. Here it is. Let me copy a sentence from your paragraph and do a string search on Leontiev's paper and…well, well, well! What do we have here?! An exact copy of the original!"*
>
> *"Oh NO!" Vladimir turns red. But he recovers quickly and smiles widely. "It's fine! Look! I put a reference to Leontiev's work right at the end of the paragraph. A reference is the same as a quote, isn't it? After all, Leontiev should be happy. I am increasing his citation count. He will not come and bother me by claiming that these words are his, not mine."*
>
> *Popov remains silent. He retrieves from the top of his in-tray basket what looks like an official letter and reads it out loud. "Dear Sir, one of my students brought to my attention that a certain Vladimir Toldoff who works in your research center has not had the courtesy to quote me in his recent paper, but instead claimed my words to be his (see paragraph three of his introduction). I am disappointed that a prestigious Institute like yours does not carefully check its papers before publication. I expect to receive from your Institute and from Mr Toldoff a letter of apology, with a copy forwarded to the editor of the journal. I hope this will be the last time such misconduct occurs.*
>
> *Signed: Professor Leontiev."*

Plagiarism comes in two flavors: unintentional and intentional. Vladimir's story illustrates a case of accidental plagiarism. It is often due to a less than perfect process to collect and annotate background material. Keeping relevant documentation about the information source when capturing information electronically is simply good practice. If you are unsure that a sentence is yours, just cut and paste it into the Google search window to see if it belongs to someone else. Another source of accidental plagiarism comes from writing a paragraph

too soon after reading related ideas in someone's paper. What you read is still undigested, unprocessed, raw; parts of the sentences read may still be lingering in memory. There has been no integration, no transformation, and no enrichment. Unintentional plagiarism may also come from sheer ignorance of the concept. I remember reading a paper where one paragraph was so much better written than the others that I asked where it came from. The researcher said that she found the paragraph describing the Java function on a website and it was so well written, so concise, that she could not have written it better herself. She simply assumed that being freely accessible on the web made it legal to copy without quote and attribution.

Intentional plagiarism is stealing for gain. It could be to gain time (no time to write a paraphrase or to retrieve the original paper where the sentence comes from). It could be to gain consideration by enlarging one's contribution with the uncredited contribution of others. Intentional plagiarism is avoiding to give credit where credit is due. Sir Isaac Newton (a contemporary of Pascal) wrote : *"If I have seen further it is only by standing on the shoulders of Giants"*. Newton and Pascal recognized that much of what they achieved is due to the people who came before them. Pascal was critical of prideful people who claimed to have accomplished everything on their own.

"Certain authors, speaking of their works, say: 'My book,' 'My commentary,' 'My story,' etc. They are just like middle-class people who have a house of their own on main street and never miss an opportunity to mention it. It would be better for these authors to say: 'Our book,' 'Our commentary,' 'Our story,' etc., given that frequently in these, more belong to other people than to them." (Pascal, Thoughts)

Plagiarism covers more than plain cut and paste. It includes word substitution. In this case, the plagiarism is definitely intentional. The writer knows the sentence is from someone else. To evade suspicion of plagiarism, the writer changes a word here and there in the plagiarized sentence. Literature has a term for this bad practice: "patchwork plagiarism".

Intentional plagiarism also includes the complete rewrite of a sequence of ideas. The writer reads the original text (possibly written in a foreign language) and rewrites it sentence by sentence using different words. This is wrong. What the law protects is more than the expression of ideas (copyright), it is also their sequence. If I translated

a paragraph from this book in French[1] or Chinese[2] (by the way, this book is also available in Chinese and French), all words would be different, but I would still be plagiarizing. The ideas expressed in successive sentences would exactly be the same.

You will find examples and an excellent write-up on *"Quoting, paraphrasing, and acknowledging sources"* in a PDF document from the writing center of the University of Wisconsin.[3]

Even subtler is plagiarism of oneself (self-plagiarism). You might think that it is unnecessary to quote a sentence or paragraph from one of your earlier publications. But, in all likelihood, you assigned your copyright to the journal, in which case the reproduction rights of your article no longer belong to you. Copying large chunks of your past publications (including visuals) would constitute a breach of copyright unless it is authorized. Good news is, all publishers have someone dealing with permissions. So if you want to reuse a visual or table from another paper, especially if it has been published in a different journal, ask for permission. If you are the author, permission is rarely refused, but the publisher may impose some restrictions such as obligatory mentions, etc.

Copyright may still be yours if you published your paper in an open access peer-review journal. To retain this right, writers pay a publication fee per article published. The average publication fee for an article published in Public Library of Science Journals (PLoS) was US\$ 2300 for the year 2010–2011. However, the use of open access entails attribution. Open Access journals such as PLoS adopt the Creative Commons Attribution License. This license[4] allows people to download, reuse, reprint, modify, copy, and distribute, *as long as original author and source are mentioned.* For a good review of self-plagiarism issues, see the article posted by Jef Akst, in his September 9 2010 blog entry[5] entitled *"When is self-plagiarism O.K.?"*

Plagiarism has become such a problem that most journals are now using plagiarism-detection software (Turnitin, Copyscape, Ithenticate, Crosscheck) to identify scientists who plagiarize. Research institutes or journals which value their reputation often require researchers to

---

[1] http://livres.edpsciences.org/ouvrage.php?ISBN = 978-2-86883-904-6
[2] http://www.sciencep.com/e_single.php?id = 17266
[3] http://students.wisc.edu/pdf/Plagiarism.pdf
[4] http://creativecommons.org/licenses/by/2.5/
[5] www.the-scientist.com/blog/display/58676

check their paper against plagiarism before submission. It is only a matter of time before such checks are conducted retroactively. Woe to the researchers found plagiarizing, even twenty years ago.

Plagiarism is just one of the crimes of unethical writing. For more crimes such as selective reporting, ghost authorship, and questionable citation practices, do read the document[6] written by Dr. Miguel Roig and sponsored by the Office of Research Integrity (ORI). It is entitled: "*Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to Ethical Writing*".

Now that you are aware of the dangers of plagiarism, let me encourage you to quote. Quoting is good practice. It demonstrates your intellectual honesty. The list of advantages of quoting does not end here. Quotes are proof that you have read the whole paper, not skimmed its abstract. You are seen as someone authoritative, someone who does not take shortcuts. When you give credit where credit is due, you have everything to gain and nothing to lose. Quoting scientists who have been published, particularly if they are authoritative, adds credibility to your own work. If you do not share their views, quoting what you object to cannot be disputed. You do not interpret; you quote.

Observe how Professor Feibelman quotes others.

> "*In apparent support of the half-dissociated overlayer, Pirug, Ritke, and Bonzel's x-ray photoemission spectroscopy (XPS) study of H2O/Ru(0001)* **"revealed a state at 531.3 eV binding energy which is close to [that] of adsorbed hydroxyl groups"** *(28).*"[7]

Note how skillfully he quotes from another paper to hint (with the word '*apparent*') that the support is not there at all. Indeed, the next sentence (not shown here) starts with '*However*' to confirm the lack of support.

The Trap of Imprecision

Imprecise reference

How many references do you list at the end of your paper? 20, 30, 50? Did you read all the corresponding papers? Where do these references come from?

---

[6] http://facpub.stjohns.edu/ ~ roigm/plagiarism/Index.html
[7] Reprinted excerpt with permission from Peter J. Feibelman. "Partial Dissociation of Water on RU(0001)", SCIENCE Vol 295:99–102 © 2002 AAAS.

Let us just take one of the references you added to the reference section of your paper. Where did it come from? A paper you wrote? A review article? A citation index? Let's imagine it comes from the list of references you found at the end of a paper you read, and let's follow the trail of that reference like a bloodhound. Where did the writer find that reference? In an older paper they wrote, a review article, a citation index, or in the reference section from another paper? Do you get my drift? The reference that you are using is as reliable as the process used to capture the reference by all the writers who used that reference before you. This process may have been manual or electronic. If the reference was extracted from an older printed paper through OCR (optical character recognition), it may contain errors. If the error occurred during manual entry, it is then propagated electronically, unless someone got hold of the original paper, spotted, and corrected the reference error. But why bother to check? Can't we trust what comes to us electronically?

The news story of Philip Ball in the December 12 2002 issue of Nature created quite a stir when he reported the findings of the paper entitled: *"Paper trail reveals references go unread by citing authors,"* by Mikhail Simkin and Vwani Roychowdhury. In the original Simkin paper,[8] entitled *"Read before you cite!"*, the authors claimed that "only about 20 % of citers read the original." The percentage may be too low and the conclusion hasty because one cannot logically infer from the copy of an erroneous reference that the writer has not read the paper. But the problem is real.

The point is this: Even though one cannot infer that the writer has not read the paper, the reviewer may still have that suspicion. If the writer is suspected of taking shortcuts in his references, then the list of reference is no longer indicative that the writer is knowledgeable. The overall credibility rating of the writer goes down.

Imprecise referencing

But the problem of bad references does not stop there. Often times a reference is wrong not because it is incorrect, but because it is at the wrong place in a sentence. Imagine this.

---

[8] http://arxiv.org/pdf/cond-mat/0212043v1

Paper [6] proves that the boiling temperature of water is *100°* Celsius.

And paper [7] proves that water freezes at *0°* Celsius.

Author John Smith writes the following sentence:

> *"It has been proven that water boils at 100° Celsius and freezes at 0° Celsius [6]."*

Of course, this is wrong. Paper [6] does not even mention the freezing temperature of water. You would perpetuate the error if you wrote the following in your own paper which references John Smith's paper:

> *"The freezing and boiling temperatures of water are known [6]."*

Had John Smith placed the reference at the correct place, the problem would not have occurred.

> *"It has been proven that water boils at 100° Celsius [7] and freezes at 0° Celsius."*

The guideline for proper referencing is given in the "scientific Style and Format," a manual for authors, editors, and publishers now in its seventh edition (Council of Science Editors). It is very clear: "A reference immediately follows the phrase to which it is directly relevant rather than appear at the end of long clauses or sentences." Anything else may introduce imprecision, or worse, error in attribution.

Checking the references takes time. Can't we just trust that other people check them first? Can't we just trust the citation tools in Endnote, or Wizfolio? Well, rubbish in, rubbish out applies in this case also. In short, check all your references and read the original papers you reference.

Imprecise words

Under the pressure of a conference- or manager-imposed deadline, you may be tempted to prepare the related work section from abstracts, not from the full text of the papers you had no time to read. Abstracts do not contain *all* the results, they do not mention assumptions or limitations, and they do not justify the methods used. As a result, your sentences will resemble this one: "*Many people have been working in this domain [1–10], and others have recently improved*

what their predecessors did [11–17]." Reviewers will see through the smokescreen. Stuffing the reference brackets with reference numbers that exceed three, not only points to abstract skimming, but it may also be symptomatic of insufficient knowledge, shoddy science, or unsound methodology.

Abstract skimming, or dotting your paper with the references of articles you have not read, will hurt you in many ways.

- Errors will creep into your paper.
- Because they find your domain knowledge too superficial, reviewers are tempted to lower the value of your contribution.
- Your research will be disconnected from other research efforts.
- Your story will lack detail, and therefore, interest.
- Readers are usually quick to detect authors who write with authority from the level of detail and precision in their paper. If your words lack precision and assurance, readers and reviewers will doubt your expertise, and question your credibility.

If any of the words from table ☛5 are found in your introduction, you may have fallen into the trap of imprecision. But if these words are immediately justified (*'Several technologies, such as…'*), they are fine.

Read your introduction, and circle the words you find in the list of imprecise words and other words which you feel are imprecise. Do you need them? How authoritative are you? Can you delete them, or replace them with more specific words or numbers to increase precision? Have you checked your references by going back to their source?

Typically	*A number of*	*Several*	*Many*	*Most*
Generally	*The majority of*	*Less*	*Others*	*A few*
Commonly	*Substantial*	*Various*	*More*	*Usually*
Can & may	*Probably*	*Frequent*	*Often*	*…*

Figure ☛5
Words that are potential indicators of a lack of precision in scientific writing.

The Trap of Judgmental Words

Some adjectives, verbs, and adverbs are dangerous when used in the related work section of your paper. The danger comes from their use in judgmental comparisons. Adjectives such as '*poor,*' '*not well,*' '*slow,*' '*faster,*' '*not reliable,*' '*primitive,*' '*naive,*' or '*limited*' can do much damage. Verbs like '*fail to,*' '*ignore,*' or '*suffers from*' are just as judgmental. Negations like '*may not*' or '*potentially unable to*' cast doubt without a shred of evidence. They make your work look good at the expense of the work of others who came before you. Sir Isaac Newton did not write: *"If I have seen further, it is because they were all as blind as bats"*. The people whose work you judge will one day read what you wrote about them, and be understandably upset.

Does it mean that all adjectives are bad? No, they are just dangerous. Every adjective is a claim; and in science, claims have to be justified. How would you explain and justify the adjective '*poor*'?

What adjectives (if any) are to be used? Adjectives that compliment (with reason) the authors or their work, adjectives that reflect undisputed public knowledge, adjectives supported by data, visuals, or quotes, and adjectives you define.

Here are eight ways to avoid direct judgment on the findings of a particular paper:

- **State agreement or disagreement** of your results/conclusions with another paper's results/conclusions, or state that your results/conclusions are coherent with, in accord with, or at variance with those found in another paper.
- **Use facts and numbers** to justify your claims. Do make sure to be fair and compare apples with apples.
- **Define your uniqueness**, your difference (nothing is comparable to what you do — maybe because you are exploring an alternative never tried before).
- **Quote another peer-reviewed paper** that independently supports your views (a review paper maybe), or quote the limitations as stated by the authors of the paper you are comparing yours with.
- Show you **improve or extend someone's work**, not destroy it.

- **Bring balance to your views**: recognize the value of an existing method in a sentence's main clause while mentioning its limitation in a subordinate clause as in '*Although this method is no longer used, it helped kick-start the work in this field*'. Avoid inverting the clauses as in '*Although this method helped kick-start the work in this field, it is no longer used.*'
- **Compare visually**, thus avoiding the use of judgmental words. Be respectful of people's earlier work.
- **Change the point of view** and the evaluation criteria. Show that with the new criteria (that you have justified), the method you are comparing yours with is no longer as effective.

Scientists of old were very gracious. Let's learn from Pascal, Benjamin Franklin, and Santiago Ramón y Cajal.

Pascal

Blaise Pascal, is not only a great scientist but also a great Christian philosopher, and a man with the right attitude. Here is a translation of one of his thoughts dealing with correcting people's mistakes, followed by a similar recommendation from Benjamin Franklin:

> Thoughts #9 of Pascal "*When one wishes to correct to one's advantage, and reveal how mistaken someone is, one must observe from which angle that person is looking at things, because, usually, from that angle, things look right, and openly admit this truth, but present the other angle from which the same things now look wrong. The one who is corrected is satisfied for no mistake was made, it was simply a matter of not being aware of other perspectives.*"

Benjamin Franklin

Extract from chapter 8 of Benjamin Franklin's autobiography: "I made it a rule to forbear all direct contradiction to the sentiments of others, and all positive assertion of my own. I even forbid myself, agreeably to the old laws of our Junto, the use of every word or expression in the language that imported a fixed opinion, such as certainly, undoubtedly, etc., and I adopted, instead of them, I conceive, I apprehend, or I imagine a thing to be so or so; or it so appears to me at present. When another asserted something that I thought an error, I denied myself the pleasure

of contradicting him abruptly, and of showing immediately some absurdity in his proposition; and in answering I began by observing that in certain cases or circumstances his opinion would be right, but in the present case there appeared or seemed to me some difference, etc. I soon found the advantage of this change in my manner; the conversations I engaged in went on more pleasantly. The modest way in which I proposed my opinions procured them a readier reception and less contradiction; I had less mortification when I was found to be in the wrong, and I more easily prevailed with others to give up their mistakes and join with me when I happened to be in the right."

Santiago Ramón y Cajal

In his book « Reglas y Consejos sobre Investigación Cientifica: Los tonicós de la voluntad », Santiago Ramón y Cajal recommends indulgence because methodology is the source of many errors. He never doubts that the author has talent, commenting that, if the author had had access to the same equipment he used, he or she would have arrived at the same conclusion. In any case, the author's work was published, and his own efforts contributed to the advancement of science, whether they were crowned with success or not.

Read your introduction, and underline the adjectives, verbs, or adverbs you find a little too judgmental or gratuitous. Replace them using one of the eight recommended techniques.

The Deadly Outcome of the Sum of All Traps: Disbelief

Stephen D. Senturia, MIT professor and Senior editor of JMEMS wrote in the June 2003 edition of the journal, an excellent article entitled: "How to Avoid the Reviewer's Axe: One Editor's View". He writes: *"A paper is written in order of decreasing believability"*. Therefore, everything in the introduction has to be believable. If the reviewer doubts the origin or sincerity of your words, the accuracy of your sources and numbers, the validity of your claims, the extent of your knowledge, or your fairness in character, then disbelief sets in. It is the fly in the ointment, the tipping point that moves the reviewer's first impression of your paper from neutral to negative.

Why would reviewers trust results or the interpretation of results if they can't even believe introductory statements! Nothing written in the introduction should be perceived as deliberately partial (outdated or omitted reference to research papers from rival groups). And nothing should be perceived as speculative. Naturally, we all know the words that express speculation: '*possibly*,' '*likely*,' '*probably*,' etc; But Professor Senturia adds a few unexpected words to that list: '*obviously*,' '*undoubtedly*,' and '*certainly*,' and other strong words of assurance behind which hide mere speculations.

The reader scientist is critical and suspicious for good reasons. Research is expensive and time-consuming. Before taking on board new ideas from other people's papers, scientists want to be sure that these ideas will answer their needs. All they have to get that assurance, are the words of the writer, the peer-review process, and their own experience. Drawing from their prior knowledge, they examine the writer's work, and since they cannot verify everything presented in the paper, at some point, they need to decide whether to trust the writer or not.

In this decision process, you realize that the reviewer of the paper plays a critical role. Good reviewers have developed a sixth sense after reviewing many papers. They know that some writers desperate to be published lie by omission (that is why the judge asks the witness to swear to tell the **whole** truth). Such writers omit mentioning the reference to a paper too close to theirs for comfort. They omit mentioning the known (and often crippling) limitations of their method or results. They omit data. They omit to show the results that do not support their hypothesis. Some of these omissions will only be known after readers discover them while trying to reproduce the research results.

The introduction is a good place for reviewers to deploy their antennas in order to pick up any signal pointing to a lack of knowledge or a lack of intellectual honesty. I remember reading an article on presentation skills than claimed that if only one side of an issue is presented, then believability is in the low 10 %; but if both sides are presented, believability is in the high 50 %. The title of the slide was "fairness". In Science, it would have been "intellectual honesty".

> The drug info sheet
>
> *To be really scared, don't watch a horror movie. Instead, go into your medicine cabinet, and read the piece of paper folded in eight sandwiched between the two strips of aluminum holding the precious pills that may cure your headache. Take the time to read the microscopic text to build up some really unhealthy anxiety. The warnings are so overwhelming that if the pills don't cure you, they might just as effectively lead you straight to the emergency room.*
>
> *If the pharmaceutical companies disclose these limitations, it is to avoid lawsuits and to help doctors prescribe the right medicine. Not stating limitations in your scientific paper won't kill anyone, but it might damage your credibility, and your chances of getting published!*

Purpose and Qualities of Introductions

Purpose of the introduction for the reader

- It brings the reader up to speed and reduces the initial knowledge gap.
- It poses the problem, the proposed solution, and the scope in clear terms.
- It answers the Why questions raised by the title and the abstract.

Purpose of the introduction for the writer

- It gives the writer a chance to loosen the tie, unbutton the collar, and write in a personal way to the reader.
- It enhances the reader's motivation to find out more by reading the rest of the paper.
- It features the writer's expertise in communication skills, scientific skills, and social skills.
- It enables the writer to strengthen the contribution.

Qualities of an introduction

MINDFUL. The author makes a real effort to assess and bridge the knowledge gap. He respects other people's work and is not judgmental.

STORY-LIKE. Its plot answers all the "why" questions of the reader. It uses the active voice and includes the writer ('we').

AUTHORITATIVE. References are accurate, comparisons are factual, related works are closely related, and imprecise words are absent.

COMPLETE. All "whys" have their "because". The key references are mentioned.

CONCISE. No considerable, vacuous starts, no table-of-content paragraphs, no excessive details.

Introduction Q&A

Q: In their introduction, some writers present their main results, others just present their main goal. What is the best way?

A: Follow the journal guidelines on how to write the introduction. Some ask the writer to present the main results and some recommend not to state the results in order to keep the introduction short, to avoid repetition, and to focus on the goals. So there are indeed two ways to write an introduction, and people have strong preferences for or against each way. Each side has convincing supportive arguments. Here are those proposed by the side supporting repetition of results.

(1) The most frequent argument hinges around the famous quote: *"Tell them what you're going to tell them. Tell them. Then tell them what you told them."* This argument may apply to an easily distracted audience, but can we assume that readers are equally distracted?

(2) Another argument, this one more convincing, is that some journals no longer require a conclusion. The last section of any paper is the discussion. In that case, it may be worth repeating the main result in the introduction.

(3) Some say that many readers only read the introduction of the paper anyway, and therefore you had better mention your results there too… just in case. While this may be true, it is doubtful that readers with this behavior would skip the conclusions (if available).

Amongst those who suggest an alternative to the repetition of the results, Michael Alley advocates "*mapping the document in the introduction*". He gives the example of a journal article where the author successfully manages to present in story form an overview of the methodology, thereby answering the "Why this way" question. The story reveals the problem and the method used to solve it. It remains silent on the results, but mentions the impact of being able to solve the problem.

My personal view is that the introduction should keep things moving and interesting by mentioning the *expected* results and their *foreseen* impact. If you do have to mention the *actual* results, do so without much detail, and make them part of the *why* story. In all cases, whether you mention the results specifically or not, end your introduction with the main expected outcome of your work.

Q: Can I cut and paste sentences from introduction to abstract?

A: The reader and the reviewer easily detect such shortcuts. Do not give readers the impression that you are in a hurry. The abstract is not written in the same way as the introduction, and the introduction is not written in the same way as the body of your paper. The verb tenses are different, the style is different, and the role of each part is different. Cutting and pasting extends beyond the mere transport of words. It carries writing styles, precision levels, and verb tenses that are fine in their original settings, but not necessarily fine in their new settings.

Q: Can I paraphrase sentences from the introduction of some earlier papers since the context of my new paper is the same?

A: Doing so contributes to boring papers, and you may be found guilty of self-plagiarism. The reason introductions feel repetitive is

often because the writer rewrites a paper for different journals, or because there is not much of a knowledge gap filled between two successive papers from the same writer. To avoid such problems, think of each paper as a unique piece of communication to a unique reader. Do not rewrite, write afresh.

Q: When does one write the introduction of a paper?

A: In his excellent little book "A Ph.D is not enough," Professor Feibelman gives sound advice:

> *"Virtually everyone finds that writing the introduction to a paper is the most difficult task. […] My solution to this problem is to start thinking about the first paragraph of an article when I begin a project rather than when I complete it."*[9]

When you write the introduction of a paper early in your project, you still have the excitement of the journey that lies ahead to energize your writing: the tantalizing hypothesis, the supportive preliminary data, and the fruitful methods. Yet, some argue that the introduction should be written at the end of the paper, once the contribution is clearer. So when does one write the introduction?

Let the content of your introduction dictate the timing. If you write a teaser introduction that states the goals, the context of your work, but only the expected results and impacts, you can indeed write your introduction early, assuming that your work keeps its original focus. Writing the introduction can also be done early if you prefer to write a succession of short papers, because your limited focus is unlikely to change.

But if your paper is the result of collaboration between many researchers spanning over a few years, it may be impossible to write the introduction early. You would then have to rely on your excellent writing skills to recapture the essence of your earlier goals and motivations to keep the reader interested along a good storyline.

Q: How long is the introduction of a paper?

A: One of my friends, now CTO in a National library, used to systematically reject any paper in which the introductory segments made less

---

[9] Copyright 1993 by Peter J. Feibelman, "A PH.D is not enough: a guide to survival in Science" published by Basic books.

than 30% of the whole paper. To him, these introductory segments are necessary to bridge the reader's knowledge gap. They include the introduction, but also the technical background section that immediately follows it. In the writing seminars I conduct, I rarely see papers with such a thorough background. The majority of papers I see have introductions that represent 10 to 15% of the paper. Occasionally, the sum of introductory segments is slightly above 20%, and for short letters, the introduction is only one paragraph.

So, how long is the introduction? To answer this question without thinking of the reader is unreasonable. The introduction is for the non-expert reader. The writer must have a precise idea of the type of non-expert reader likely to benefit from his or her paper. How much does that scientist know? How much background does that scientist require to use the contribution in whole or in parts? After answering these questions, and after considering the number of pages available for the paper, the writer can then set the length of the introduction. Note that a paper with multiple contributions is likely to sacrifice the technical background to fit within a given number of pages. That is why it is better to write multiple papers, each with a single contribution and with an appropriate gap-reducing introduction.

Q: What can one tell about a scientist from reading the introduction of their latest paper?

A: Amazingly, one learns a great deal. If the introduction is well developed, and yet easy and interesting to read, the writer displays good communication skills. If the introduction contains targeted references (as opposed to bulk references) and few or no imprecise words, the writer displays good scientific skills. If the introduction contains no judgmental words, the writer displays good social skills. Communication, scientific, and social skills are essential qualities. Besides reviewing the CV or publication record, managers would be well advised to read the introduction from the latest paper written by their potential hire.

Q: How can I give a hint to the reader that my results are not able to reproduce the results claimed in some of the related works?

A: Use the past tense when referring to possibly erroneous findings. Use the present tense to show that, as far as you are concerned,

you are in no doubt that the information is correct as shown in this sentence:

Tom et al. identified a catalyst that increases the yield at high temperatures [7].

What sentence comes next, (1) or (2)?

(1) Slinger et al. subsequently reported that the increased yield is not due to the catalyst [8].
(2) Slinger et al. subsequently reported that the increased yield was not due to the catalyst [8].

The correct answer is (2). It allows you to contradict Slinger's findings with the following sentence:

We found evidence that the catalyst does increase the yield.

Let's change the first sentence by expressing doubt on the findings of Tom *et al.*

Tom et al. identified a catalyst that increased the yield at high temperatures [7].
Slinger et al. subsequently showed that the increased yield is not due to the catalyst [8].
We also found evidence that the yield increase at high temperatures is not linked to the catalyst but to...

Q: In your list of four justifying questions (*why this, why now, why this way, and why should the reader care*), I noticed that the question *"why you"* was missing. Do I need to answer that question also?

A: The *why you* question is not answered directly. You have four ways to answer it. (1) Your track record — the trail of references from your previous papers. (2) The acknowledgements — if people fund your research, it is because they are confident that you can deliver something of value. (3) The most senior author mentioned in your list of authors, a well-known and well-cited author acts as a warrant. (4) The reputation of your research institution and its track record in well-cited research papers in your field.

But if you are new to the field, an outlier with no research sponsor, no academic heavy weight on your list of authors, and no glamorous university or research center to attach your name to, do not despair, even if the odds seem against you, what matters in the end is the quality of your writing and the timeliness and impact of your research.

Do not forget that besides the four reader questions you mention, an additional two sets of questions on problems and solutions come from the reviewer: (1) Is the problem real and is it a useful problem to solve; and (2) Is the solution novel and better than other solutions.

Q: Should I always place the bracket containing a reference at the end of a sentence?

A: The reference must be unambiguous. For example in the following sentence, where should the reference be placed to avoid ambiguity when referring to the review paper, in * or in **?

*Three speech recognition technologies * are prevalent today: Hidden Markov Models, Neural Networks, and statistical methods such as Template Matching or Nearest Neighbor **.*

The correct answer is * because ** would be ambiguous as it may only refer to the nearest neighbor statistical method, not to the review paper that covers all three technologies. In short, the reference should be placed *immediately* after the information it references. If each technology has its own reference, then the following scheme would apply:

Three speech recognition technologies are prevalent today: Hidden Markov Models [1], Neural Networks [2], and statistical methods such as Template Matching [3], or Nearest Neighbor [4].

Q: What references should I put in the reference section?

A: Multiple types of references: (1) recent references because they show that you are keeping up-to-date with what is happening in the field; (2) references to papers published in the journal you are targeting because the readers of that journal will have easier access to them; (3) references to review papers when the information you use can only be found in them, but (4) references to the original paper and not the review paper when the review only points to the original paper without adding value; (5) references to papers YOU HAVE READ; (6) references to all the papers that have directly contributed results, data, or methods to your paper; and (7) References to the papers from the leader in your field.

Q: What references should I NOT put in the reference section?

A: References simply cut and pasted from one paper to the next, references to papers you have not read, references loosely connected to what you are doing, references to mildly relevant papers published by one of the following classes of people: your researcher friends, your manager, people from the same university or institute, or the unknown researchers who cited your paper — just to return the favor.

Introduction Metrics

✓(+) The introduction starts fast, without warm-ups.
✓(+) The introduction finishes with the anticipated outcome of the research.
✓(+) The introductory segments enable the non-expert reader to benefit from the paper. They represent more than 15 % of the paper.
✓(+) References are never in groups exceeding three, the majority are single.
✓(+) All four "why" questions are answered explicitly: why this, why now, why this way, and why should the reader care.
✓(+) The introduction is active, personal, and story-like.
✓(+) Methods, data, and/or application field correctly frame the scope of the paper.
✓(+) The rare imprecise words found in the introduction are qualified immediately after their use ("several… such as")
✓(+) Judgmental words are never used and the story plot is well connected.
✓(+) Background is provided for each specific and intermediary title keyword.
✓(−) The first sentences of the introduction are known to the journal readers, or attempt to warm the reader by referring to the hot research topic.
✓(−) The introduction does not finish with the impact of your contribution.
✓(−) Bridging the knowledge gap has not been considered. Introduction size is below 10 % for a regular scientific paper.
✓(−) References are presented in groups exceeding three.
✓(−) The answer to one or more "why" questions is missing
✓(−) The mostly passive voice introduction uses less than three personal pronouns.
✓(−) The scope of the paper is mentioned in parts only, and not easily identified.
✓(−) Imprecise words are sprinkled throughout the introduction.
✓(−) Judgmental words are found in the introduction, or the story plot does not include comparisons or does not relate the paper to past papers.
✓(−) Background is missing for some specific and intermediary title keywords.

AND NOW FOR THE BONUS POINTS:
✓(+++) The introduction contains one visual or (+ + +) the average number of words per sentence in your introduction is 22 words or less.

This page intentionally left blank

Chapter 15
Visuals: The Voice of Your Paper

A voice attracts attention; it announces, it warns. It is a substitute to writing: one can read a book or listen to a recorded version of it. Likewise, photos, tables, diagrams, and graphics attract attention even without words. They are worth a thousand words. *The voice gets out of the body. It is not necessary to see the body to hear its voice.* Visuals inform readers independently, even before they start reading the first paragraph. *A voice has its own language, a universal and wordless language, like the one used by the child who babbles, laughs, and cries.* Visuals have their own language, the universal language of scientific graphics. They tell a story directly and quickly with minimum text. *Voice intonation reinforces the message expressed by the body.* Visuals also reinforce the main message of the text, and are in synergy with it.

Just observe the title of this chapter for a few seconds and then bring your eyes back to this mark ☛

Title, headings and subheadings shout, don't they? They are so authoritative in their bold font suit. Framed by white space, nothing crowds them in their spacious surroundings. They are understood at a glance.

Tables and diagrams speak just as much as photos. Guided by a grid of vertical and horizontal lines, bold font, and arrows, the reader captures a large volume of information in little time and easily extracts trends and relationships between the visual elements. The visual story is told in few words.

Visuals excel at comparisons such as *before* versus *after*, or *with* versus *without*. More talents are featured in ☞ 1

Table ☞ 1
Typical use for visuals

To represent complexity	To summarize	To reveal sequence
To classify	To reveal patterns	To establish relationships
To compare and contrast	To give precision and detail	To provide context

Readers prefer a visual to text for several reasons:

- In linear text, eyes walk like ants along the narrow path created by lined-up words. In visuals, eyes leap like crickets from one place of interest to another, probing with silent questions. Readers enjoy the speed and the freedom of self-guided exploration.
- Because the text part (title and caption) of a visual is redundant with its graphical part, readers understand more easily.

Visuals have a loud and convincing voice, but only if you can make them speak. Their language is based on a special grammar that describes the correct use of fonts, blocking, kerning, framing, white space, line and color. This visual language is well understood by graphic designers. They can make a visual shout, whereas most of us can only make it whisper or croak. This chapter is not about graphic design, it is about the correct use of visuals in a scientific paper. It is also about principles that will help you design visuals that have an impact from a scientific perspective, even if the lines are a little thin, the white space is not quite well-distributed, or the kerning is an abomination. You may not get an Oscar in a design competition, but you will have visuals that do more than whisper or croak.

However loud and clear a voice may be, if it babbles or utters gibberish, it is not helpful. Visuals need to deliver a loud, clear, intelligible, and convincing message.

Seven Principles for Good Visuals

After reading hundreds of papers, I have detected consistent error patterns. A bad visual breaks one or more of the following principles.

- A visual does not raise unexpected questions.
- A visual is custom-designed to support the contribution of only one paper.
- A visual keeps its complexity in step with readers' understanding.
- A visual is designed based on its contribution, not on its ease of creation.
- A visual has its elements arranged to make its purpose immediately apparent.
- A visual is concise if its clarity declines when a new element is added or removed.
- Besides title and caption, a visual requires no external text support to be understood.

Principle 1: A visual does not raise unexpected questions

As soon as a visual appears, the eyes of the reader probe the visual. They are on a fact finding mission. Wouldn't it be nice to track their path! In Finland, at the university of Joensuu where I conducted a scientific writing skills seminar, the usability lab has an eyeball-tracking machine. I ran the test that I will ask you to do shortly, and discovered a few interesting things in the process. Four participants were asked to look at a particular visual for five seconds (they did not know what to expect), and then to look at it again, this time in order to ask as many questions as the visual raised.

So, if you do not mind, I would love for you to do the same. Without the eyeball-tracking machine, you will have to remember where your eyes travel in the first five seconds. I am confident that, within that time, your brain will predictably direct your eyes to key parts of the visual; probing, evaluating, and asking silent questions. The first questions will be "what-am-I-looking-at" questions.

It is now time to put your eyes to the test, but do not look for information in the title and caption of the figure because I have removed them to help you focus on the visual itself. After five seconds, read the next paragraph. Look now at ☞2 **for five seconds** then read the next paragraph.

In the first five seconds many things happened. The first one was the discovery of the overall image: that of a class of visuals you know
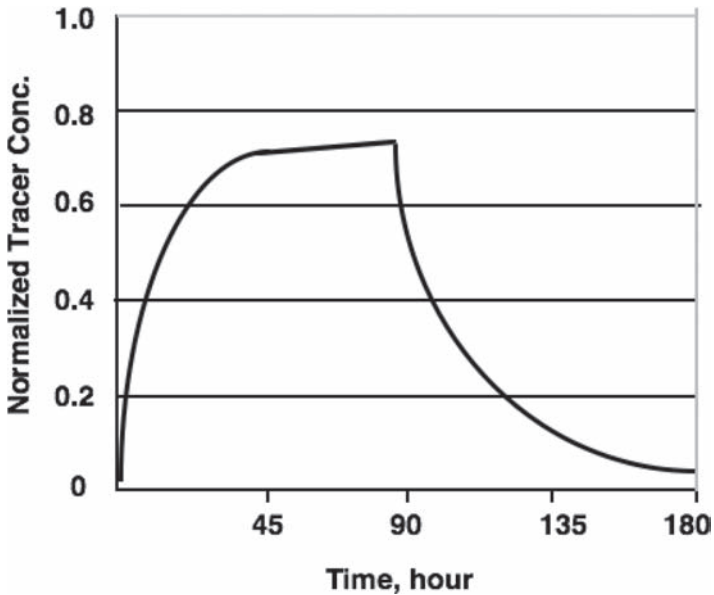
Figure ☞2

about: The X-Y Diagram with its singular point of change which is meaningless unless you know the meaning and value of its coordinates. Your eyes then moved to one of the two axes. Some went to look at the X axis, maybe looked at the end value, then at the unit (hours), and then looked at the Y axis. The others looked first at the Y axis label and tried to decipher the obscure abbreviation, then moved their eyes down to the unit of the X axis (Time), and they moved past the 90 hours to go to the point of change on the curve, and then looked at the Y value for that point. Interestingly, not all went to the Y axis first, even though it is supposed to contain the dependent variable. Those who looked at the X axis first were not driven by logic but by the ease of reading. Horizontal information is easier to read than vertical information.

Look at the visual again, this time without a time limit and identify all the questions this visual raises. Before you look again at ☞2, let me answer the first two questions: What is the Conc. Abbreviation, and why is time so slow? The graph represents the evolution in concentration of a fluorescent tracer in a large tank containing some muddy toxic sludge. This explains the use of hours. Now it is your

turn. What other questions do you have? When you are done with your questions, read the next paragraph.

This curve ☛**3** looks like a quadratic curve but it isn't. The differences need to be explained.
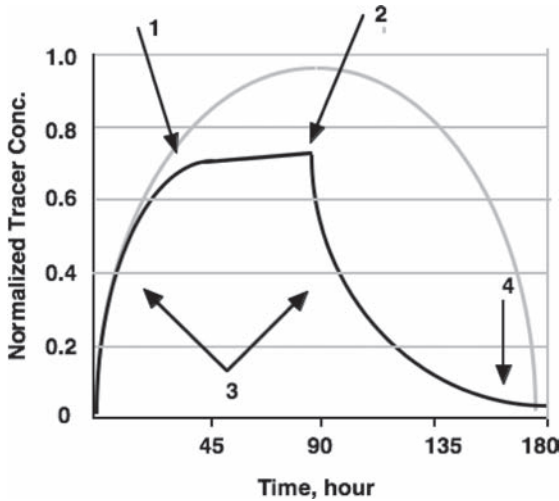
Figure ☛**3**
I have superposed a regular parabola on top of figure 2. This graph 'asks' five questions. Each arrow corresponds to one question. Can you guess what the fifth question is?

- Question1: Why is the top of the curve linear between points 1 & 2?
- Question 2: What happens at point 2 to change the behavior of the phenomenon so drastically?
- Question 3: Why is the curve convex on the way up and concave on the way down?
- Question 4: Why is the curve asymptotic for high values of time?
- Question 5: What is the concentration normalized to, and why doesn't the curve reach 1?

The readers will look for answers to these questions in the caption. If they are left unanswered, the readers will be frustrated **because the visual raises more questions than the writer is willing to answer.**

Chief among the visuals that raise unneeded questions, is the screenshot. It is widely used to illustrate papers because a mouse click is all it takes to capture it ☛**4.**
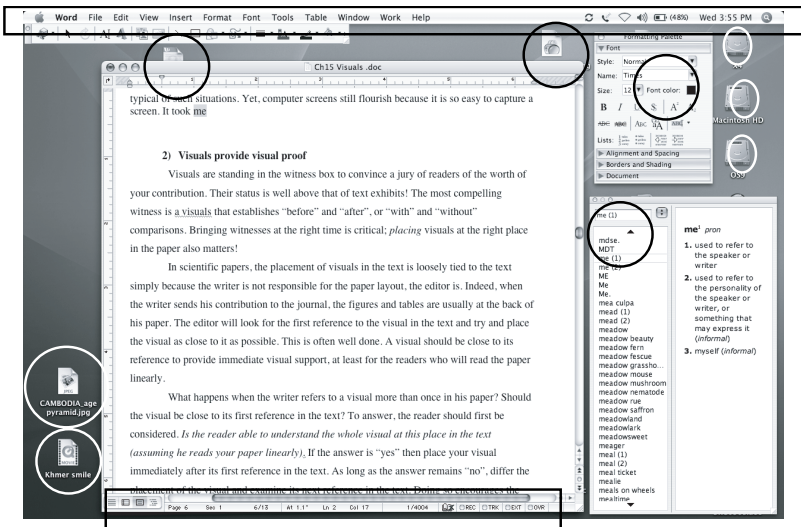
Figure ☛ 4

Visual gallery of errors exhibit. The cluttered screen 'dump'. If my objective is to show the content of the center window, what are the extra windows, circled files and folders and framed menus doing on this visual?

The screen capture includes all the artifacts of the software application: menu items, windows, icons, tool palettes, and other distracting elements that raise questions.

Naturally, to the writer who captured the screen, everything on the screen is clear and familiar. But is it also clear to the non-expert reader that you have in mind, the one who stands to benefit from your work? There are only two ways to know: ask a reader, or look at each visual of your own paper **as if for the first time** and pretend to be that reader.

Readers come up with the most unexpected questions! What is at their source?

You will often find that they are caused by the reader's lack of knowledge. You see, the writer knows too much! That is why it is useful to ask someone else to do this exercise. You may be able to identify by yourself other sources of questions such as unfamiliar acronyms and abbreviations for example. Once you have identified all questions the reader might have (or has), you have five choices:

- **Leave the visual as it is** but answer the unanswered reader questions in the visual title, its caption, or in the body of the paper.

- **Add to the visual**. Add clarifying visual elements (boxes, arrows, links…)
- **Take away from the visual.** Remove whatever raises distracting questions to keep the focus on the essential point made by the visual.
- **Divide the visual**. Split it into two or more independent visuals that are less complex.
- **Modify the visual** elements (shape, size, order, font, etc) to clearly reveal the purpose of the visual.

The only choice **you do not have,** is to ignore these questions.

What are the questions raised by your visuals? Choose the key visual in your paper (the one which is the most representative of your contribution), and show it to one or two colleagues. HIDE the caption; show only the visual and its title. Ask them which questions the visual raises. As they tell you, write these questions down. When they have no more questions, DO NOT answer the questions. Uncover the hidden caption. Ask if their questions are answered in the caption or in the part of the paper referring to the visual. Once you have identified the problems, ask the reader whether the caption contains new information not illustrated (apart from the description of the visual's context or the visual's interpretation). If the caption does contain visually unsupported information, remove it; otherwise, it would be a claim unsupported by visual evidence.

Show your modified visual to new colleagues and verify that the caption now answers ALL questions, and does not itself raise unwanted questions.

Principle 2: A visual is custom-designed to support the contribution of only one paper

Do you remember this particular visual you worked on for hours? It was a work of art. Using Photoshop, you or the graphic artist had spent much time to make it look perfect. It had been admired at an internal technical presentation or at a previous conference. Part of it does illustrate a key point in your paper, but that part cannot easily be extracted from your original masterpiece without serious rework. You are tempted to reuse the whole drawing/diagram as it is. As a result, the visual includes information irrelevant to

your purpose: names, curves, numbers, or acronyms foreign to the reader. All raise questions. Hence, the second principle: a visual is custom-designed to support the contribution of only one paper.

Redrawing is a small price to pay for a custom-designed visual in view of the benefits, not just to the readers, but also to you because (1) your contribution, unclouded by irrelevant details, is easier to identify; and (2) you do not have to ask for the permission to reuse the original if the reproduction rights belong to a journal.

Principle 3: A visual keeps its complexity in step with readers' understanding

Visuals are your star witnesses. They stand in the witness-box to convince a jury of readers of the worth of your contribution. Their placement in your paper is as critical as the timing lawyers choose to bring in their key witness. To know where to place a visual in a paper, take into account the level of understanding of your reader. Complex visuals are more logically placed towards the end of the paper when the reader has filled the knowledge gap. Simpler visuals or self-contained visuals can be placed anywhere.

What happens when you refer to a visual more than once? Assuming readers read your paper linearly from introduction to conclusion, the first time they are asked to look at your visual, they may find it too complex because they have yet to acquire the knowledge that will make the visual totally understandable. A single visual explained here and there in the paper breaks principle #1. It will always raise more questions than the writer is willing to immediately answer.

So ask yourself, why is it necessary to refer to the visual more than once? Is it because you are making multiple points in one large or complex visual? If this is so, divide the one complex visual into several parts (a), (b), (c) to reduce its complexity. Next, make sure readers have enough information to understand *everything* in Figure 1(a) when reaching Figure 1(a) in the text, and do likewise for parts (b) and (c). But if after dividing the visual into parts you realize there is no value in having visuals 1(a), 1(b), and 1(c) next to one another, say for comparative reasons, then divide the visual into separate visuals and have them appear in a just-in-time fashion next to their respective references.

When you send your paper to a journal, its figures and tables are usually at the back of the paper after the references (unless you submit the paper as a PDF file). The person in charge of the page layout will look for the first reference to the visual in your text, and try to place the visual as close to its reference as possible. This is often well done. But a large visual creates page layout problems. For this reason, if you want your visual to be properly placed, 1) design it so that its width matches an integer number of the journal column width — use a template if provided by the journal; and 2) avoid using small serif fonts (Times for example) that cannot be made smaller without a significant reduction in readability, thus limiting the range of options for the placement of your visual. Use fonts without serifs (also called sans serif fonts) like Helvetica, or Arial because of their uniform thickness. When shrunk, their thicker lines do not disappear as fast as the thin lines always present in serif fonts.

Principle 4: A visual is designed based on its contribution, not on its ease of creation

Information with visual impact requires creativity, graphic skill, and time. Because most of these are in short supply, software and hardware producers have provided skill-enhancing, and time-saving tools: statistical packages that crank out tables, graphs, and cheesy charts in a few mouse clicks; digital cameras that, in one click, capture poorly lit photos of experimental setups replete with noodle wires (I suppose the more awful they look, the more authentic they are); and screen capture programs that effortlessly lasso and shrink your screen to make it fit in your paper. The ease of creation of visuals contributes to their abundance—mouse-produced becomes mass-produced. When I show you the photo of a keyboard with the caption "keyboard on which this book was typed," as in ☞**5,** does it contribute greatly to the usefulness of this book?

Such irrelevant photos are found in scientific papers. They are not useful. They only prove that the writer conducted an experiment using real equipment. To ensure that each visual is critical to your paper, ask yourself whether it replaces much text or strongly supports your contribution. Conciseness applies to both text AND visuals.

Figure ☞5
Gallery of errors exhibit. A Qwerty keyboard, but who wants to know. This photo speaks volumes, doesn't it? It tells you that I use a Macintosh Powerbook with a titanium casing, that I do not use a French keyboard even though I am French, that my right SHIFT key is broken in two, and finally that I am not much of a photographer! What does this have to do with the book itself? Nothing much.

The abundance of visuals is a source of unpleasant side effects, the main one being the reader's inability to identify the key visual in a paper. When I ask researchers to read a paper and designate the one *key* graph/figure/table that represents the contribution of the paper, they often disagree with one another and even disagree with the author of the paper. Surely, it should not be so! All should agree on which visual is the most important. Why is there disagreement? It may be due to the author's inability to make his contribution clear visually, but the cause may also lie elsewhere. The more visuals you have, the more your contribution is hard to find, and the more difficult it is for the reader to grasp your whole contribution succinctly. This unpleasant side effect hides another. Removing a visual will weaken your contribution if your contribution is dispersed (diluted) across many visuals.

To summarize, if weak (diluted) visual circumstantial evidence dominates your paper at the expense of succinct but detailed convicting evidence, your article loses clarity and conciseness.

How many visuals do you have in your paper? Could you identify the one that encapsulates the core of your contribution? Could other people? Are you verbose or concise when it comes to visuals? What do your readers think?

≈	PostAtl5(liquid)		PostAt30(liquid)		PostAt30(solid)		PostAt30(solid)	
	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$
B4	0.5323+17.4%	0.5323+18.9%	0.4225+19%	0.4254+20.0%	0.2157+9.6%	0.2185 + 11.1%	0.1493+9.3%	0.1501+8.4%
B6	0.5323+17.4%	0.5373+18.9%	0.4202+18.0%	0.4254+20.1%	0.2156+9.5%	0.2171 + 10.8%	0.1493+9.0%	0.1500+8.5%
B8	0.5324+17.5%	0.5373+18.9%	0.4189+17.7%	0.4255+20.0%	0.2156+9.5%	0.2182+10.9%	0.1492+9.1%	0.1496+8.5%
BJI	0.4720	0.4720	0.3706	0.3706	0.2997	0.2997	0.2380	0.2380

Table 1 Statistics on ß – 1 or 20, LG = 3

≈	PostAtl5(liquid)		PostAt30(liquid)		PostAt30(solid)		PostAt30(solid)	
	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$
B4	0.5456+23.4%	0.5423+23.9%	0.4324+26.0%	0.4341+27.0%	0.2192+12.6%	0.2232+14.9%	0.1544+13.3%	0.1554+14.8%
B6	0.5440+23.4%	0.5473+23.9%	0.4332+25.0%	0.4341+27.1%	0.2193+12.5%	0.2235 + 14.1%	0.1544+13.3%	0.1550+14.6%
B8	0.5424+23.5%	0.5473+24.0%	0.4299+24.7%	0.4341+27.0%	0.2126+12.6%	0.2231 + 14.7%	0.1543 + 13.0%	0.1556+14.0%
BJI	0.4720	0.4720	0.3706	0.3706	0.2997	0.2997	0.2380	0.2380

Table 1 Statistics on ß – 1 or 20, LG = 4

≈	PostAtl5(liquid)		PostAt30(liquid)		PostAt30(solid)		PostAt30(solid)	
	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$	$\beta = 1$	$\beta = 20$
B4	0.5400+22.1%	0.5323+18.9%	0.4195+27,2%	0.4254+28.5%	0.2198+13.2%	0.2241 + 15.0%	0.1545+13.8.%	0.1580+15.2%)
B6	0.5401+22.5%	0.5349+18.9%	0.4242+28.0%	0.4254+28.5%	0.2198+13.1%	0.2140+15.8%	0.1546+14.0%	0.1550+15.1%
B8	0.5397+22.7%	0.5373 + 18.9%	0.4201+27.7%	0.4255+28.5%	0.2130+13.2%	0.2235+14.9%	0.1546+13.8%	0.1567+15.7%
BJI	0.4720	0.4720	0.3706	0.3706	0.2997	0.2997	0.2380	0.2380

Table 1 Statistics on ß – l or 20, LG = 5

Figure ☛6

Visual gallery of errors exhibit. The visual salvo. A large number of visuals are set side by side; each one is marginally different from the one that precedes, so much so that the eye can barely see the difference between them. In this case the visuals are tables, but they could be graphs or images also.

Principle 5: A visual has its elements arranged to make its purpose immediately apparent

The visual salvo is a popular classic in the gallery of errors leading to complex visuals (☞**6**).

Visual ☞**6** is impressive, but the reader is not impressed. The principle — a visual has its elements arranged to make its purpose immediately apparent — is certainly not applied here.

Clarity of purpose on the writer's side is essential to achieve clarity of meaning on the reader's side. Clouding the meaning is easy: simply bury the key information in the midst of other data so that its purpose does not stand out. For example, writers could inadvertently camouflage key information by presenting the data in the wrong sorting order, by aggregating it in classes that absorb the interesting variance, or by placing it in the midst of loosely related data.

Key data must stand out. They are your crown jewels, but they need a crown; and the role of the crown is not to hide the jewels, but to make them stand out by giving them the appropriate structure. And the structure is the arrangement of data that makes clear a purpose. So when visuals fail to reveal their purpose, it is for several reasons:

- The writer has no specific purpose — the visual is a data dump.
- The writer has a purpose, but leaves to the expert the work of identifying that purpose from the data presented. The writer has not considered that the non-expert reader is not able to identify the purpose without his or her help.
- The writer has a clear purpose (it is even mentioned in the caption), but the arrangement of the data does not support the purpose, and the reader struggles to reconnect purpose and data.

To highlight the main point of a visual, its elements need to be organized. In table ☞**7**, the caption asks the reader to compare one-step with two-steps methods. Even though the arrangement of the table elements is not haphazard with all one-step methods grouped at the end of the table, the reader is lost. What is to be compared? The average value of one-step methods to the average value of two-step methods? Each one-step method with its corresponding two-step

Table ☞7

Visual gallery of errors. The table without a clear message. Comparison of all combinations of one — two step methods

Methods	True positive rate (%)	False positive rate (%)
BN&BN	22.0	1.3
BN&MO	24.9	1.9
BN&MSV	39.2	0.2
PSY&BN	27.1	2.6
PSY&MO	27.0	2.7
PSY&MSV	66.9	0.3
COR&BN	23.0	1.9
COR&MO	25.8	2.5
COR&MSV	38.1	0.2
BN	21.8	1.2
MO	24.8	1.9
MSV	35.9	0.2

method where it reappears as the second step (for example, MO compared to BN&MO, COR&MO, and PSY&MO)? The best one-step method with the best two-step method?

The table, filled with acronyms defined elsewhere in the paper, is unfriendly to the memory. What point did the writer want to make? What do you think? Look again at ☞7, and then return to the next paragraph.

Most readers thought the writer wanted to show that one combination of two steps, PSY and MSV, has the highest true positive and

Table ☞8

The table's message is clearer. Best method for all combinations of one — two step methods

Methods	True positive rate (%)	False positive rate (%)
BN	21.8	1.2
BN&BN	22.0	1.3
COR&BN	23.0	1.9
MO	24.8	1.9
BN&MO	24.9	1.9
COR&MO	25.8	2.5
PSY&MO	27.0	2.7
PSY&BN	27.1	2.6
MSV	35.9	0.2
COR&MSV	38.1	0.2
BN&MSV	39.2	0.2
PSY&MSV	66.9	0.3

a very low false positive score. If the writer had wanted to make that point *immediately* clear, it would have been better to sort the scores in ascending order as in table ☛**8**.

But if the writer had indeed wanted the reader to compare methods where the one step reappears as a second step, then the following table (☛**9**) would have been more to the point.

Table ☛9
The table's layout is better to help the reader compare one and two step methods.

Methods	True positive %	False positive %
BN (1-step)	21.8	1.2
BN&BN	22	1.3
COR&BN	23	1.9
PSY&BN	27.1	2.6
MO (1-step)	24.8	1.9
BN&MO	24.9	1.9
COR&MO	25.8	2.5
PSY&MO	27	2.7
MSV (1-step)	35.9	0.2
COR&MSV	38.1	0.2
BN&MSV	39.2	0.2
PSY&MSV	**66.9**	**0.3**

What if the writer had wanted to make a different point altogether. For example, imagine the visual had to show the lack of significant difference for any single step method between using a combination of two steps and using one step only, except when it came to the MSV method. The accent is now on *the difference* between one step and two steps. Look at table ☛**10**.

From table ☛**10**, the reader immediately sees that the PSY method is systematically the best method as a second step, and that

Table ☛10
The table's message is clear and even more instructive even though less data is displayed. Comparison of all combinations of one — two step methods

One step method	% True positive rate*	2nd step method (BN, COR, or PSY) Best second step	% Increase in true positive rate (PSY as 2nd step)
BN	21.8	PSY	24.31
MO	24.8	PSY	8.87
MSV	35.9	PSY	86.35

* Maximum false positive rate for all one step / two steps combinations is less than 2.7 %

MSV is the best one step method which achieves the best increase in overall score as the first step in the two step combination with PSY. Compared to the original table ☛7, much raw data has disappeared, and in particular the false positive rates. The writer has not left them out. Their maximum value is given in a footnote.

Comparison of one step and two step methods
(For each method,True positive in 1st column, and false positive in adjacent column)
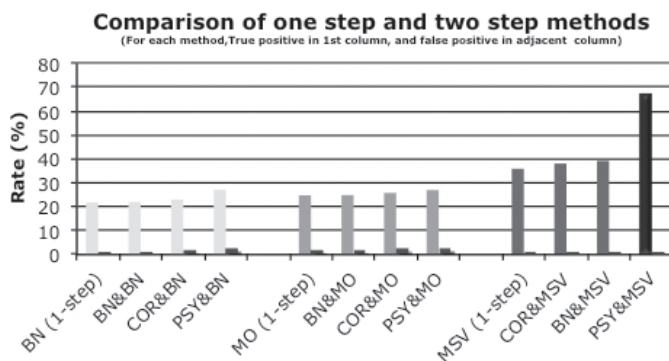
Figure ☛11
The comparison of one and two-step methods reveals that the one-step MSV method (36,5% true positives) is superior to the BN and MO methods. MSV also acts in synergy with the PSY method to nearly double the score with 67.5% true positives while keeping false positive rates low.

Comparison of 1 step vs. two step methods
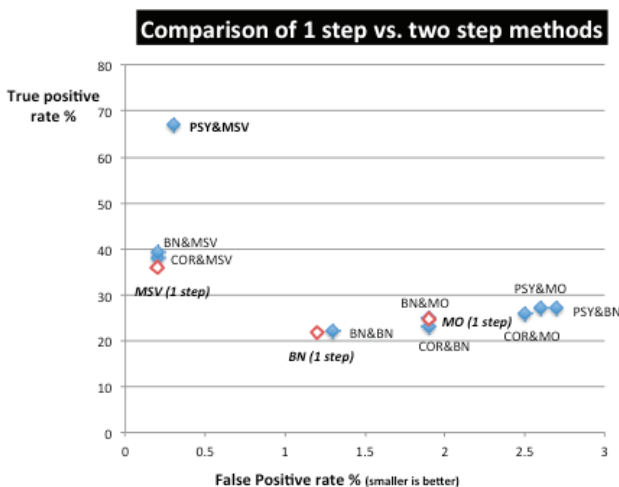
Figure ☛12
The comparison of one and two-step methods reveals that the one-step MSV method (36,5% true positives) is superior to the BN and MO methods. MSV also acts in synergy with the PSY method to nearly double the score with 67.5% true positives while keeping false positive rates low. (One-step method is empty diamond, two-step method is full diamond)

223

How about other types of representations? What would be the best visual to show what table ☞10 reveals, but with all the data (true and false positives)? Two visuals would rapidly make the same point (☞11, and ☞12). Note how the caption has been rewritten to help the reader.

In summary, for each point you want to make, find the most appropriate visual. A specific visual makes a specific point; a different point would be made by a different visual. Choose your data based on their added value towards your contribution and based on their conciseness (do they make the same point with less elements). Once chosen, arrange your data until their organization clearly makes your point. This usually takes many drafts. Granted, improvement takes time, but it is time well spent: a visual is so much more convincing than a paragraph.

Principle 6: A visual is concise if its clarity declines when a new element is added or removed

Each visual has an optimum conciseness. Visuals ☞11 and ☞12 include the false-positive values. Are these critical to the point the author wants to make? Is it possible to come to the same conclusion without them as in ☞10?

To add visual elements to a graph is tempting; to merge two graphics to make room for more text is irresistible. The resulting visual is so complex that it is no longer understandable. It becomes the "everything but the kitchen sink" visual. The density of its elements per square inch hinders rather than facilitates understanding.

The draft diagram in ☞13 had attractive rainbow colors, 3D elements, arrows, links and much more. Some may recognize the yeast cell growth cycle at its center.

It required much time to design and was a masterpiece. It established a parallel between two phenomena that shared the same cell cycle. It would have been perfect, but for a small problem: only its authors understood it. This figure was later simplified, and clarity returned when visual elements were taken away.

If adding or subtracting elements in one visual affects the clarity of another visual, then these two visuals are clearly interdependent (☞14). They have to be redesigned to increase their independence.
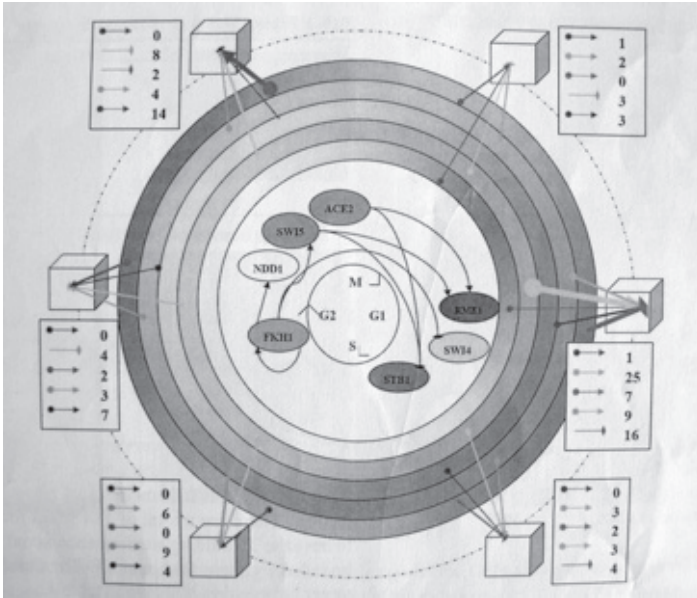
Figure ☞13
Visual gallery of errors. The overly complex visual. This beautiful visual combines two related visuals into one. The resulting increase in complexity greatly reduces clarity and understanding.
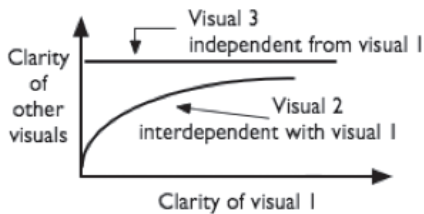
Figure ☞14
If the clarity of a visual is dependent on the existence of another visual, then the two visuals are interdependent, and as a result, the clarity of one visual will affect the clarity of the other.

To summarize this point, complexity is born out of (1) a lack of discrimination in the choice of the elements included in a visual, (2) a lack of clear relationship between the various elements in a visual, and (3) a lack of independence between visuals.

Simplification decreases visual conciseness whereas consolidation increases it. But conciseness is always the servant of clarity, not its master; that is why a visual can be considered concise if its clarity declines when a new element is added or removed.

225

Examine each visual. What makes it difficult to understand?

Is there a better way to make the same point with less elements? Visuals come in many types: chart, diagram, table, photo, list, etc. Would replacing one type with another make your visual clearer? Would dividing one visual into two make things clearer? Would combining two visuals make things clearer? Would reorganizing the information in your visual make the relationships between its elements more obvious (using arrows, colors, words, or sorting the data in a different order)?

Principle 7: Besides caption and title, a visual requires no external text support to be understood

The strange oasis

An old Bedouin likes to tell the tale of a strange oasis he once came across in the Sahara desert where a sand storm had stranded his caravan. The tallest man, who was perched on top of the tallest camel in the caravan, saw it first. "Oasis straight ahead!" he shouted. They pressed ahead. A short distance away from heat and thirst relief, the travelers noticed that clusters of full coconuts were sitting on the sand dunes, away from the coconut trees in the oasis. Their skin was soft, but they were hot to the touch, so the people took them inside the oasis to drink later. The oasis was small, it had no well, and all its coconut trees were barren, so the only refreshment would have to come from the coconuts found on the sand dunes. Unfortunately, these coconuts were not ordinary. Their soft skin hardened like steel as soon as they were inside the oasis in the shade. The sharpest dagger could not cut them open. So people had to go back out into the desert to open them, and return to the oasis to drink them, a process they found most unpleasant.

The oasis is still there, the Bedouin claims. It is now an attraction for tourists who go and visit it by helicopter (camel rides are just too slow nowadays).

In our story, the oasis is the refreshing visual and its caption, the desert is paragraph text, and the coconuts are what should have been in the caption, but was turned into paragraph text instead. Nowadays, readers are pressed for time. So they parachute themselves directly into your paper with a marked preference for the pleasant visuals, which are far more refreshing that paragraph text. However, they are frustrated because, to understand them, 1) they need to refer back to the text and search for the *'see Figure X,';*

2) they then need to locate the beginning of the sentence; and finally 3) they need to go back and forth between the explanatory text, the visual, and its caption, until their understanding is complete. This time-consuming and iterative process is most unpleasant. In the immortal words of the Bedouin, "Coconuts are meant to be inside the oasis". We no longer live in the days of silent movies. A visual must "tell all" by itself, without the need for text outside of its caption. To accommodate the non-linear reading behavior of scientists, each visual should be self-contained, which means self-explanatory. Some will object that this creates redundancy because the text in the paragraph is then repeated in the caption. They assume that the text in the body of the paper remains the same once the caption of the visual makes it self-explanatory. This is not the case. The paragraph in the main section, now shorter, only states the key contribution of the visual without detail, repeating only what is necessary to move the story along. Doing so has two advantages: (1) the visual can be understood without the need to read the whole article and (2) the body of the paper is shorter (and thus faster to read) because it keeps to the essential.

In the original visual in ☞**15**, the CALB and MCF-C18 acronyms are undefined.

(a)   (b)
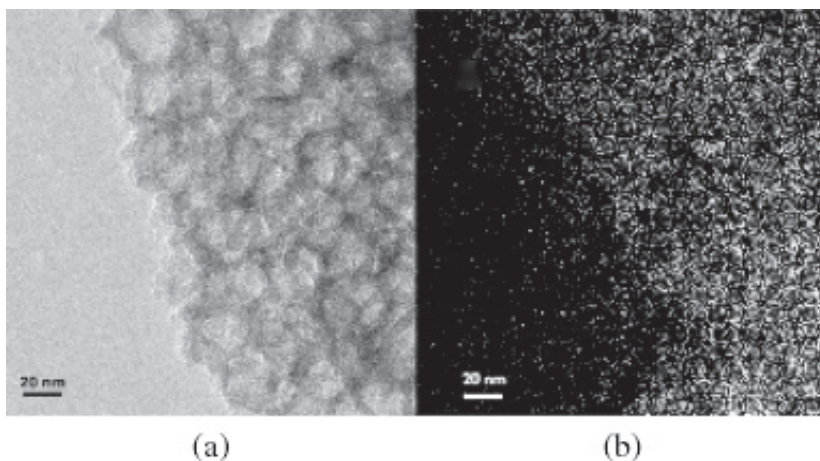
Figure ☞15
A visual that does not stand alone and is not self-contained.

(Caption) Figure 5. (a) TEM micrograph of CALB/MCF-C$_{18}$ from pressure-driven method, and (b) the corresponding ELS elemental mapping of N.

Readers have to go back and forth several times between text, visual, and caption before they understand and "get the full picture."

(Text in body of paper)

> *"Figure 5 illustrates the uniform nitrogen mapping over the CALB/MCF-C18 sample, indicating the homogeneous distribution of the nitrogen-containing enzymes within the mesoporous silica matrix. CALB/MCF-C18 also showed PAFTIR peaks at 1650 cm-1 and 3300 cm-1 (Figure 3c), which were associated with the amide groups of the enzymes, confirming the enzyme incorporation."*

Compare the visual with the modified one in ☛**16.** The new visual is now autonomous, its caption is longer, but the description in the body of the paper is cut down to the bare essentials.

> *(Modified caption)* **Figure 5.** *(a) Transmission electron micrograph of the Candida Antarctica Lipase B enzyme (CALB) immobilized by pressure in the porous matrix of hydrophobic mesocellular siliceous foam (MCF-C$_{18}$), and (b) the corresponding electron energy loss spectroscopy map of nitrogen. Nitrogen is abundant in enzymes. Its detection is used as evidence of their presence. Here, Nitrogen, and therefore the CALB enzyme, is seen as uniformly incorporated and distributed in the porous matrix.*

(a)   (b)
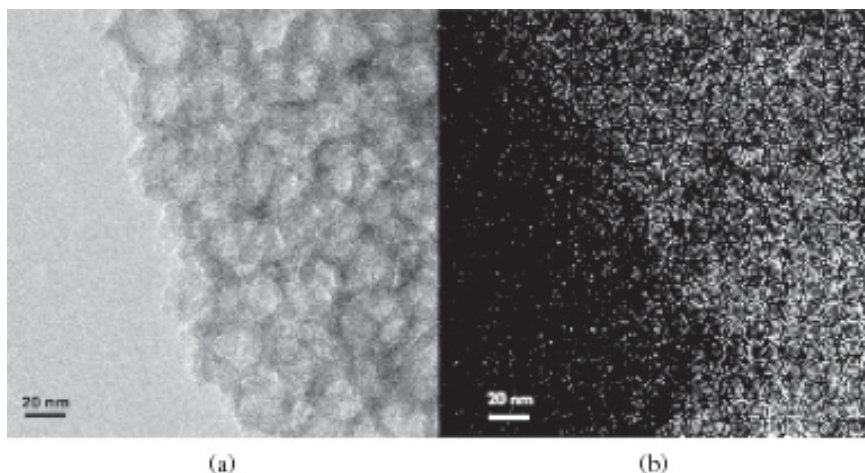
Figure ☛16
A visual that is now stand alone and self-contained.

(Modified text in body of paper)

"Both electron energy loss spectroscopy (Figure 5b) and FT-IR spectrum of CALB/MCF-C$_{18}$ (Figure 3c) confirm the incorporation of the enzyme in the Siliceous foam."

The captions of Figure 5 (a) and (b) are now self-contained, while the text in the body of the paper is cut to focus only on the point the writer wishes to make (the enzyme is trapped in the porous matrix). Even if the reader goes straight to Figure 5 and bypasses the text, the same message is given.

Examine each visual in your paper.
Rewrite the caption to make your visual self-contained.
Revise the description of the visual in the body of your paper to shorten it by stating only the purpose of the visual, or its key contributive point.

Purpose and Qualities of Visuals

Purpose of the visual for the reader

- It allows self-discovery of the paper.
- It helps the reader verify the claims of the writer.
- It saves reading time by allowing faster understanding of complex information and faster understanding of problem and solution.
- It provides a direct (shortcut) and pleasurable (memorable) access to the writer's contribution.

Purpose of the visual for the writer

- It makes the paper more concise by replacing many words, particularly in the introduction where it provides fast context, and helps bridge knowledge gaps.
- It motivates readers to read more, yet allows them not to have to read all.
- It provides compelling evidence, in particular evidence of contribution.
- It enables the writer to represent complex relationships succinctly.
- It (re) captures the reader's attention and improves memory recall.

Qualities of a visual

SELF-CONTAINED.
Besides title and caption, no other element is necessary to understand it. It answers all reader questions.
CLEAR and CONVINCING. It has structure, it is readable, and it includes visual cues to help readers focus on key points.
CONCISE. It contains no superfluous detail. It cannot be combined with other visuals without loss of essential information or clarity, nor can it be simplified.
RELEVANT. It has purpose and is essential to the contribution. It does not distract the reader.
QUICK TO GRASP — It should be understood in less than 20 seconds.

Examine each visual in your paper. Is it concise? Can you hide details in the appendices or footnotes? Is the visual essential? Is it rapidly understandable to a reader who is not an expert in your field? Is it autonomous and understandable without any support from your main text? Should it appear earlier or later in the paper?

Visuals Q&A

Q: I would like to recycle a visual I first published in the extended abstract of a conference proceeding to reuse it for a journal publication. The two papers are very close. Can I just reuse it or do I need to redo it?

A: Unless you have kept the copyright to the original paper, you are obligated to ask the owner of the conference proceedings for the permission to re-use the visual. If the two papers are very close, you also have to give the journal a copy of the original conference paper so that the editor can assert whether there are enough differences to merit publication. In short, redo the visual. There are many ways to present data, and I am sure that the original visual could benefit from another draft. After all, any paragraph published, or any visual published is only a draft. There are so many ways to improve on text, likewise for visuals.

Q: Can I change the contrast of my visual to make it more readable?

A: Any image post-treatment is always looked upon with suspicion, and is often forbidden by the journal, so check the journal's instructions to its authors regarding artwork, and abide by them. There is some tolerance, for example to improve contrast, but again, 1) if improving contrast removes information from your visual, or conceals it, you should not do it (increasing the contrast may remove a gray level that was significant); and 2) you should inform the journal that you have manipulated the image, and provide the original image for comparison.

May I suggest that you read the following guideline from the large academic publisher Elsevier. It will help you understand the issues.[1]

Q: Can a reference to a figure (say figure 2) be made in the caption of another figure (say figure 3)?

A: In that case, figure 3 is not standalone because understanding it requires the prior understanding of figure 2. If you cannot make figure 3 stand-alone, do mention figure 2 in its caption. It informs the reader that figure 2 and 3 are interrelated, and that figure 2 must be examined before figure 3. Not mentioning figure 2 in the caption of figure 3 would leave the reader in the dark.

Q: Can I write the interpretation of the visual in its caption, or do I leave that to the discussion?

A: If the visual is in the result section of your paper, you are not expected to interpret it there. But remember, the way you have laid out your data already guides the reader towards your interpretation. Highlight the salient data or points that will be used in the discussion section when you offer your interpretation. If the visual is in the discussion section, nothing should stop you from providing an interpretation in the caption. This will definitely help you achieve the worthwhile goal of making any visual stand-alone.

Q: Do I need to bother about page layout issues for my visuals since the journal is going to redraw them anyway?

A: You should if you want to increase the chances for your table or graph to be kept close to its reference in the text. This is particularly

---

[1] http://www.elsevier.com/framework_authors/Artwork/Artwork_2010.pdf

relevant when the journal you are targeting has multiple columns on one page. When that is the case, you may want to redesign a horizontal table to make it fit vertically on one column of a two column page. For example, Visual ☞ **11** in this chapter could easily be made vertical by inverting the x-y-axis with no loss of meaning! But keep vertical what people expect to see vertical, like temperature bars; and keep horizontal what which people expect to see horizontal, like distances. And while on the subject of visuals, never hesitate to insert additional space or lines to add structure and to highlight what you consider important. Use (larger) sans serif type fonts if you expect that the visual will have to be shrunk to fit the size of a page or column. Better still; rework your visual to make it fit on a page or column naturally by increasing conciseness and clarity.

Q: There are so many ways to visualize my data, so which way is the best?

A: Before you even ask yourself that question, ask yourself these, IN THE ORDER GIVEN:

- Which point do I want to make in my visual? Asking this question helps you avoid the pointless data dump.
- What data do I absolutely need to make my point, and which data can I leave out?
- Do I have the data I need to make the point I want to make? If not, what point can I make with the data I have?
- Which level of data aggregation or transformation (frequency, percentage, mean, cumulative, log, etc) will best make the point?
- Now that I have the right data, and I know what point to make with it, what is its dimensionality (2D, 3D, nD vector)? What is its nature (qualitative, discrete, continuous, temporal, pictorial, numerical, symbolic as a chemical notation or the elements in a diagram)? What is its accuracy (error bar, range, probability, resolution)? What is its range and scale (finite, infinite, zero to one, set of known attributes, set of given names, etc)?
- Given my type of data, which ordering scheme would support my point: the data intrinsic order (chronological,

numerical, spatial, logical, hierarchical), or a new order such as functional proximity, functional classes, inclusion — exclusion, with — without, before — after, generic — specific, simple — complex, most probable — least probable, low priority — high priority, most favorable — less favorable, most relevant — less relevant, similar — different, parents — children, etc)?

- Which visual representation is the one the reader expects to see in order to be convinced by my ordered data? (more than one may apply — use the one which makes the point most immediately and most clearly) Table, list, line chart, stacked line chart, flowchart, bar graph, photo, Venn diagram, block diagram, tree structure, schematic diagram, 2D bar chart, 3D line chart, Pie chart, doughnut chart, 2D area chart, 3D floating bars, point chart, bubble chart, grid surface chart, mesh surface chart, radar chart, polar chart, high-low chart, error bar chart, funnel chart, box chart, XYY chart, Western blot, Northern blot, Pareto chart, scatter plot, and so many other representations are available.

Visuals Metrics (calculate your score for each visual)

✓(+) The visual is self-contained.

✓(+) The visual has no acronyms (in caption, title, or inside visual)

✓(+) The visual is essential to support the contribution and add to its value.

✓(+) The point made by the visual is seen by the reader within 20 seconds.

✓(+) The visual easily fits on one column.

✓(+) The visual's title, table headings, axis legends are clear and informative.

✓(+) The type of visual used corresponds to what the reader expects, or is better than what the reader expects. It fully supports the point made in the text or in the caption.

✓(+) The visual does not raise more questions than the writer is willing to answer.

✓(+) The caption provides context to help understand the visual.

✓(−) The visual's understanding depends on external support other than caption/title.

✓(−) The visual has acronyms.

✓(−) The visual is not essential and makes a secondary point.

✓(−) The point made by the visual is missed, or unseen, or slow to see.

✓(−) The visual requires much space and could be far from its reference in the text.

✓(−) The visual's title, table headings, axis legends are cryptic or too abbreviated.

✓(−) The type of visual used does not correspond to what is required to make the point made in the caption.

✓(−) The visual raises more questions than the writer is willing to answer.

✓(−) The caption is silent on the context.

AND NOW FOR THE BONUS POINTS:

✓(+++) Reader and writer, or two independent readers agree on which single visual represents the core of the contribution.

Chapter 16

Conclusions: The Smile of Your Paper

After ruling out many choices, I decided that no part of the body could represent the conclusions better than the smile. Why a smile? I thought again of the many conclusions that had disappointed me, and deflated my enthusiasm with self-deprecatory endings such as '*In order to test real performance improvement….,*' or '*This could be greatly improved by…*'. I had read these articles with great interest until, in the conclusions section, I had found suggestions that nothing significant had been accomplished. I felt like the person about to buy a car described as safe, only to be told at the last minute that the car had no air bags and no antilock braking system. Previously undisclosed limitations disguised as future work frequently surface in the conclusions to disappoint the reader who genuinely assumed the author had dealt with them already. Imagine a lawyer who managed to demonstrate the innocence of his client throughout the court proceedings, but who, on the very last day in front of the jury, apologizes because not enough evidence has been produced to justify the plea of innocence. How unbelievable!

The way a defense lawyer ends his plea in front of a jury should also be adopted to end a scientific paper: with assurance, firmly, and smiling, trusting the jury will find the client not guilty of scientific insignificance. Lawyers know that the day the jury gathers for the final plea is an important day. The day the writer writes the conclusion of his paper is also an important day. The writer cannot write it at nighttime, close to exhaustion. His writing will be lifeless. The writer cannot write the conclusion too long after the end of the research. His sense of past achievements may be gone.

So, before writing the conclusion, the writer has to re-energize his pen by reading again introduction and discussion to identify the research milestones. As with the coin collecting hero in the nintendo game, intermediary scientific merit points accumulate to form his final score: the paper's global contribution. Re-energize yourself, smile as you consider your score, stay positively charged… because the negative ions are there: your fatigue, the time that passed since the end of your research, and the limitations that need to be addressed in the future. However, do not waste that positive energy on yourself, to polish your halo, or bask in the sunshine of your glorious past. The conclusion is not an opportunity for an ego trip. It is an opportunity to polish, not your halo, but your diamond of a contribution because you need to sell it to cash in on citations, to encourage others to use your work.

You may have noticed that, in some journals, articles have no conclusions heading; the discussion ends the paper. The need to conclude is still there however, even if the heading is absent. Some journals — Nature is one of them — recommend to finish an article without conclusions. They would rather have the author write the last paragraph "about the implications of what the reader has read",[1] and not summarize what has been accomplished. Professor Railsback with great common sense writes: *"Conclusions are just that — the inferences that can be drawn from your data, not a reprise of the entire paper."* Combining expert advice, *'inferences'* and *'implications,'* it becomes clear that conclusions are not just another abstract.

Abstract Versus Conclusions

Surely readers wouldn't notice that your abstract is similar to the conclusions, would they? They would!

Readers read in a non-linear fashion. They tend to skip large sections of a paper, jumping from abstract to conclusions for example, like the hurried reporter only attending the first and last day in court. From a writer's perspective, that behavior is not ideal, but the writer can use this knowledge to his advantage. First, the writer

---

[1] Nature Physics, "Elements of Style", editorial, Vol. 3 No. 9 September 2007.

now understands how dangerous it is to copy and paste sentences between abstract and conclusions since the reader immediately notices them. Second, the writer should differentiate the conclusions from the abstract to avoid boring the reader. How do the two differ?

- Often the journal recommends the use of the past tense in the abstract. Unfortunately, the main tense used in the conclusions is also the past tense because you are referring to what you did. It is all in the past behind you. Only the facts that have been demonstrated without a doubt, the unquestionable scientific facts, are stated in the present tense. The lawyer says *'my client is innocent,'* not *'my client was innocent'*. The present tense in the conclusions reinforces your contribution. If the journal does not impose the use of the past tense in the abstract, it becomes advantageous to write the whole abstract in the present tense because doing so differentiates conclusions from abstract.

- Whereas the abstract briefly mentions the impact of the contribution, the conclusions dwell on this aspect to energize the reader. In his book *A Ph.D. Is Not Enough*, Professor Feibelman gives his writer's viewpoint.

  > *"The goal of the conclusions section is to leave your reader thinking about how your work affects his own research plans. Good science opens new doors."*[2]

- Conclusions are more comprehensive than the selective abstract. Conclusions bring closure, not on what the self-contained abstract announced, but on what the introduction and discussions opened. They close the door on the past before they open *'new doors'* onto the future.

- The abstract adopts a factual, neutral tone. The conclusions keep the reader in a positive state of mind. Remember that a reader needs strong motivations to read the whole paper, not just your conclusions. The motivating role is traditionally taken by the introduction, but if the reader skips it and jumps directly to the conclusions, then they must also

---

[2] Copyright 1993 by Peter J. Feibelman, "A PH.D is not enough: a guide to survival in Science" published by Basic books.

motivate the reader to reach inside your paper. Therefore, keep your energy level high and think positively about your contribution.

- Everything in an abstract is new to the reader. In the conclusions, nothing is. The conclusions do not surprise the reader who has read the rest of your paper. Following the analogy with the defense lawyer's final plea in front of a jury, it stands to reason that any attempt to convince the jury at the last minute with evidence that has not been cross-examined is not receivable and is objectionable. Such last-minute-theatrical surprises are the realm of Hollywood movies only. Even the section about future work is expected. In the discussion section, you venture explanations that require future validation, or you suggest that different methods might be helpful to bypass constraining limitations. The reader who has read your discussion anticipates that, in your future work, you will explore these new hypotheses or use these different methods.

Examples and Counterexamples

Examples

In the following example, the author repeats a main aspect of his contribution already announced in the discussion section. It is an encouragement for others to use his method.

> *Our method has been used to determine the best terminal group for one specific metal-molecule junction. In addition, we have demonstrated that, in principle, it is applicable to other metal-molecule couplings.*

It is not always necessary to have conclusive results to conclude. Sometimes, the hypothesis presented in the introduction can only be partly validated. The choice of words to say so is yours, but you must admit that the phrasing is quite critical here. Which of these sentences is better?

> *In conclusion, our modified gradient vector flow failed to demonstrate that…*
> *In conclusion, our modified gradient vector flow has not been able to demonstrate that…*

Or

In conclusion, our modified gradient vector flow has not yet provided definitive evidence for or against …

The last sentence is much better, isn't it? The word "*yet*" suggests that this situation may not last. Far for being despondent, the scientist is hopeful. "*Yet*" creates the expectation of the good news that comes later in the paragraph. To convince the reader, the author shares his conviction through the use of the present tense (in bold in the example).

*In conclusion, our modified gradient vector flow model has not yet provided definitive evidence for or against the use of active contour models in 3D brain image segmentation. However, it **confirms** that polar coordinates, as suggested by Smith et al [4], **are** better than Cartesian coordinates to represent regions with gaps and thin concave boundaries. In addition, we have now removed the need for a priori information on the region being modeled without affecting model performance.*

The findings are inconclusive, but they reveal that (1) an undesirable constraint has been removed; and (2) for a particularly complex type of contour, another coordinate representation scheme is confirmed to be more efficient. Even partial achievements are important to the scientific community when they validate or invalidate other people's theories and observations, and when they establish the benefit of a method against other methods for a particular type of experiment. Science explores, step-by-step, a labyrinth of many dimensions. Marking a dead end before turning back is necessary, especially when much energy has been spent exploring what turned out to be a dead-end.

If the findings are conclusive enough, why wait until all the possible paths have been explored before submitting a paper. Mention what you intend to do next to discourage potential competitors, or to encourage others to collaborate with you.

The 25 % improvement in re-ranking the top 10 documents by using words adjacent to the query keywords found in the top five documents demonstrates the validity of our assumption. We anticipate that the high frequency but non-query keywords found in the top five documents may also improve the re-ranking and plan to include such keywords in future research.

In the previous example, the writer stated precisely his future research plan to establish the anteriority of his idea and protect future research.

Presenting in conclusions any limitation that has put a lower ceiling on your high hopes is a perilous exercise. But the ceiling is not permanent — at least, that is what you want to convey to the reader. You know that tackling these limitations is definitely worthy of future research. Relaxing one of your strong assumptions or finding a way to bypass a limitation may enable others to solve their problems. Taking the time to state assumptions and limitations is not only good scientific practice, it is also a way to promote science and your name in science. But how does one constructively present these opportunities in a conclusion? The next example adapted from an IEEE paper starts with a sentence that sends a chill down one's spine.

> *Finally, we summarize the limitations of our optimizing algorithm and offer our future research plan.*

There are so many limitations that the authors find it necessary to summarize them. Two of the authors of the paper were senior fellows, and I suspected they knew how to remain positive in the face of adversity. Indeed, they knew. Here is the first item on their list:

> • *Parameter tweaking. As discussed in section 4.2, the value of alpha is obtained without difficulty, but a satisfactory gamma value is obtained only after experimenting on the data set. We have given the reader pointers to speed up the determination of gamma in this paper. We plan to investigate a heuristic method that allows direct determination of all parameters. In this respect, we believe that Boltzmann simulated annealing will be an effective method.*
> • *…*

The parameter tweaking limitation is minimized in two ways: (1) by emphasizing that a method has been given to speed up the labor-intensive part of the algorithm, and (2) by showing confidence that a solution is at hand to bypass this limitation.

Counterexamples

When it comes to conclusions, be conservative and exercise restraint. Do not destroy your good work with sentences like these:

> *In the future, we would like to validate the clustering results not only from the promoter binding site analysis, but also incorporate more information such as the protein-protein interactions, pathway integration, etc, in order to have more convincing and accurate results.*

As a reader, how did you view the achievements? Did you feel that the author was pleased with his contribution?

Here is a familiar sentence written in a humble, self-effacing way, too low key to encourage the reader.

> *Our method has been used to determine the best terminal group for one metal-molecule junction only, although in principle, it can be applied to other couplings.*

Would you trust the conclusions of a paper that ends with the following sentence?

> *In the future, we intend to experiment our approach using larger data sets.*

Does it mean that the current method relies on data sets which the authors think too small?

The next sentence seems fine… if only the writer had not used *'we believe.'*

> *"Although these protocols will continue to change, we believe they represent a reliable starting point for those beginning biochip experimentation."*

The positive contribution is placed in the main clause at the end of the sentence, but some readers perceived this sentence as slightly negative. Read this sentence again and skip *'we believe.'* You may find the protocols more appealing. The facts appear to speak for themselves, without the need for beliefs to influence the decision of the reader.

In the next sentence, both main and subordinate clauses contain positive facts. Since the main clause contains information about the future, the future should appear appealing. But this is not quite the case:

> *Although the model is capable of handling important contagious diseases, new rules for more complex vectors of contagion are under construction.*

Both subordinate and main clause establish positive facts, yet the overall perception is not always positive. Why? The readers are confused. Ordinarily, if the *although* clause contains a positive argument, readers expect the main clause to negate or neutralize the value of that argument. In this case, the main clause also contains a positive argument. As a result, the overall impression is mixed.

Purpose and Qualities of Conclusions

Purpose of the conclusions for the reader

- They bring better closure to what the introduction announces, by contrasting pre-contribution to post-contribution. What was unproven, unverified, unexplained, unknown, partial, or limited is now proven, verified, explained, known, complete, or general.
- They allow readers to understand the contribution better and in greater detail than in the abstract to evaluate its usefulness.
- They tell the readers that follow-on papers may be expected from the same author.

Purpose of the conclusions for the writer

- They place a particular emphasis on what the contribution allows others to do, immediately or potentially.
- They propose new research directions to prevent duplication of effort, to encourage collaboration, to point to unexplored fields, or to claim anteriority for novel ideas.

Qualities of conclusions

POSITIVELY
CHARGED. They maintain the
excitement created in the introduction.
PREDICTABLE content. There are no surprises. Everything
has been stated or hinted to in other parts of the paper.
CONCISE. Focus on benefits to readers.
Close the door. Open new doors.
COHERENT with title, discussion
and introduction. They are a part of one
same story.

Examine your conclusions. How positively charged are they? How consistent are they with the claims you made in the abstract and introduction? Do they "open new doors"?

Conclusions Q&A

Q: If the research stopped, should we still mention something in the future work part of the conclusion?

A: The worst possible thing one could do is to invent future work to have something to write about. You may, however, mention open research issues. The reader will understand that your work has stopped and that you do not intend to pursue these issues, thus leaving them "up for grab" as they say.

Q: To me, listing the many benefits of my research to the reader feels like a list of sentences starting with 'in addition,' 'moreover,' 'furthermore'… Is there a better way?

A: When you mention these benefits, you may want to give the reader a time frame: the ripe hanging fruit ready for immediate picking without much else to do, the green fruit that will require extra research to come to maturity (your next paper maybe), the flowers pregnant with the promise of a fruit — (a longer term research effort by your group), and the flowers awaiting pollination from other bees (not your research group).

Conclusion Metrics (if you have a conclusion)

✓(+) The conclusion is positively charged.

✓(+) The conclusion is significantly different from the abstract.

✓(+) The conclusion is longer than the abstract.

✓(+) The conclusion does not claim new benefits or findings not already presented before.

✓(+) The conclusion encourages the reader to benefit from the contribution or to further the work.

✓(–) The conclusion presents limitations as a severe drawback instead of an opportunity to improve.

✓(–) The conclusion does not differ much from the abstract.

✓(–) The conclusion simply restates the results, and if there is an impact statement, it is a plain restatement from the abstract, without elaboration.

AND NOW FOR THE BONUS POINTS:

✓(+++) The reader is able to reconstruct the title from the conclusions.

Chapter 17
Additional Resources for the Avid Learner

First, I would like to congratulate you. If you are reading this chapter it is because you have decided to further understand the science of scientific writing by venturing online. I have spent many years exploring the rich humus of the Internet garden and collected over 50 sites worthy of your attention. Their URLs have been tested at the time of publication. They may change over time but you will always find an up-to-date list on the book site: www.scientific-writing.com. And once you have been published, and invited to present your paper at a conference, you will need a new set of skills. My scientific presentation blog will be of great use to you, and so will the book with DVD *"When the scientist presents,"* which I wrote as a companion for this book. It is also published by World Scientific Publishing.

On Visuals

Graphics and visuals are essential in scientific papers.

Only a few websites deal with the topic. I like the site from Bates University with the title page *How to write a paper in scientific journal style and format*, and in particular the page entitled *Everything You Wanted to Know About Making Tables and Figures*.

http://abacus.bates.edu/%7Eganderso/biology/resources/writing/HTWtablefigs.html

Naturally, the Grand Master of visualization is Edward Tufte. On the page *Cancer survival rates: tables, graphics*, I enjoyed the example demonstrating that a table is not the poor man's tool, and that it can

overtake the corresponding graphic in terms of clarity and interest, when the two are blended.

www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id = 0000Jr&topic_id = 1

On Grammar

Your first priority as you finish reading this book, is to download a pdf book, print it, and use it for reference. You will not find this book for sale simply because it is free. This book is on grammar, but not any kind of grammar. This 108 page grammar is written by scientists for scientists; Where else would you find a grammar that recommends to write the abstract of your paper in the present tense! It comes to you courtesy of the scientists working at the NASA Langley Research Center In Hampton Virginia, and in particular Mary K. McCaskill. The title is *Grammar, Punctuation, and Capitalization; a handbook for technical writers and editors*.

www.sti.nasa.gov/publish/sp7084.pdf
www.sti.nasa.gov/sp7084/contents.html

Here is another site you cannot bypass (the French say *'incontournable'*). Better still, download the PDF file when you find it (or buy the book), print it, and read it once a year. Professor William Strunk, Jr. wrote this little book to encourage his students to write English properly (Grammar, English composition). He intended it to be concise and to the point, which is why it is still read by generations of writers even today. Its title is *The Elements of Style*.

www.bartleby.com/141/ (html)
www.cs.vu.nl/ ~ jms/doc/elos.pdf (pdf)

Another article I strongly recommend is the one written in the American Scientist by George Gopen and Judith Swan entitled: The science of scientific writing. It touches on aspects of writing rarely covered but essential for fluid reading: text progression.

www.americanscientist.org/issues/feature/the-science-of-scientific-writing

Large universities help their students improve their writing skills by making resources available online. I recommend you visit three sites.

The Purdue university Online Writing Lab (OWL) is extensive.

http://owl.english.purdue.edu/owl/

The university of Wisconsin has a great handbook containing an extensive grammar section.

www.wisc.edu/writing/Handbook/GramPunct.html

Penn State University proposes many online exercises, and they refer to Michael Alley's work. The header on top of the web page says *Writing Guidelines for Engineering and Science Students*. '*Writing,'* here, covers an extensive range: from the journal article to proposals, lab reports, thesis and dissertations, and even résumés and correspondence. Click on the link *'Writing Exercises'* under *'Student Resources'* at the top left of the page, and have fun learning.

www.writing.engr.psu.edu

If English is your second language, you probably struggle with the use of the articles *'a,' 'an,'* and *'the.'* Indeed, the incorrect use of the article will immediately betray the non-English background of a writer. Learn from the experts. Three documents will help you: the Writing Center at Rensselaer Polytechnic Institute has a good page entitled *Article Usage* written by John R. Kohl. It is exhaustive.

www.rpi.edu/web/writingcenter/esl.html

Be sure to also read section 1.5.1 in the NASA grammar for a different perspective, and the page entitled *The Use and Non-Use of Articles* in the grammar of Purdue University writing lab.

http://owl.english.purdue.edu/owl/resource/540/01/

Finally, punctuation makes a significant difference in the clarity of writing. Do make sure to read on the use of punctuation. Each of the websites mentioned above has a section on punctuation. The NASA grammar dedicates 32 pages to this important topic, from section 3.1 to 3.16. I also enjoyed reading two pages from the Purdue University Online Writing Lab (OWL) entitled: *Commas with Nonessential Elements* and *Proofreading for Commas*.

http://owl.english.purdue.edu/handouts/grammar/g_commaproof.html
http://owl.english.purdue.edu/handouts/grammar/g_commaess.html

On the Evolution of Scientific Writing

No, Darwin did not write anything on this topic. But Alan Gross and Joseph Harmon have written interesting articles. They have one article freely available online. It is entitled *The Scientific Article: From Galileo's New Science to the Human Genome*.

http://fathom.lib.uchicago.edu/2/21701730/

If you subscribe to "The Scientist" online, do read their article entitled *What's right about Scientific Writing* (The Scientist 6 December 1999, 134(24):20). You may need to be an online subscriber of the journal to be able to read this article.

www.the-scientist.com/article/display/18803/

and then read Michael Seringhaus in a later issue of The Scientist (21 November 2006) arguing against the scientific paper in its current form. His article is entitled *The Death of The Scientific Paper*.

http://papers.gersteinlab.org/e-print/paperdeath/preprint.pdf

If you want to discover the views of Roald Hoffmann, 1981 Nobel Prize Winner in Chemistry, on scientific writing, and how he wishes it to evolve, read the fascinating article he co-wrote with Pierre Laszlo entitled *The Say of Things* published in the Social Research journal, Vol 65, No 3 (Fall 1998).

www.roaldhoffmann.com/pn/modules.php op = modload&name = Downloads&file = index&req = viewdownload&cid = 1&min = 360&order by = titleA&show = 20

Professor Henrietta Tichy, a specialist in technical and scientific writing, has a few choice words aimed at the old guard of science writers. Her article entitled *Advice To Scientist-Writers: Beware Old 'Fallacies'* (published in The Scientist 2[20]:17, Oct. 31, 1988) starts with the following sentence: "Bits of advice from fallacy land have a strong influence on writing". You can read the rest – as long as you are an online subscriber to the journal!

www.the-scientist.com/article/display/8862/

Another scientist, cancer researcher Chris McCabe, is on *a mission to sex up scientese* (title of an article published on the Guardian Unlimited Newspaper website on February 4, 2004, in its life section).

www.guardian.co.uk/life/feature/story/0,13026,1140718,00.html

On Persuasion

You will have to convince the journal that your paper deserves its attention. There are two documents I have found that help in that respect. The first is an extensive page written by editor Kwan Choi. The page contains pdf files on topics such as *"General publication strategies," "Writing strategies," "Preparation and submission," "Rejection and revision,"* and even a topic on *"being a good referee."*

www.roie.org/how.htm

The second document is a paper entitled *How to write a frequently cited article* by Henk van der Vorst. If you ever want to know how citations work, this article is for you. Do not get caught trying to use your influence to increase the number of citations. On the positive side, there are things the author can do to attract attention in a good way. This paper reviews them.

www.austms.org.au/Publ/Gazette/2004/May04/vorst.pdf

If your paper is rejected, it is because somewhere, somehow, you did not convince the peer reviewer that it is worth publishing. A good article published in a medical journal reviews the reasons for rejection. Trisha Greenhalgh wrote an article for reviewers entitled *To read a paper: getting your bearings (deciding what the paper is about)*. One of its subtitles is *"The science of 'trashing' papers"*. The paper is about clinical research studies.

www.bmj.com/cgi/content/full/315/7102/243

Another interesting paper also in clinical studies is written by Arthur Croft. It is entitled *Scientific Writing: Discriminating Good from Bad*. Part 1 covers statistical errors, part 2 surveys, part 3 assignment, assessment, and analysis in experimental studies, part

4 interpretation and extrapolation of experimental studies, and part 5 sensitivity and predictive value.

www.chiro.org/LINKS/croft.shtml

Logical fallacies abound in scientific writing. Scientists unknowingly fall prey to logical fallacies. It matters to be aware of what they are. An excellent article by Stephen Richardson entitled *Logical Fallacies in Scientific Writing* disappeared from its original website, possibly because it applies logic to the heated arguments between creationists and evolutionists, but reappeared here.

www.cs.auckland.ac.nz/ ~ cristian/i2rcs/i2rcs_docs/logic.htm

On Scientific Papers

If writing a review paper interests you, here is a journal paper that covers the subject.

http://dbem.ws/Writing % 20Review.pdf

For an original metaphor on scientific paper writing, the *Kipling's guide to writing a scientific paper*, is worth a read even though the writer's views on the introduction section differs from the one presented here. The same medical journal also published the *uniform requirements for manuscripts submitted to Biomedical journals: writing and editing for Biomedical publication.*

www.cmj.hr/2002/43/3/12035130.pdf
www.cmj.hr/2003/44/6/20ICMJE.pdf

I enjoyed the Bates college page entitled The Structure, Format, Content, and Style of a Journal-Style Scientific Paper. It systematically goes through all parts of a paper.

http://abacus.bates.edu/ ~ ganderso/biology/resources/writing/
HTWsections.html
http://abacus.bates.edu/ ~ ganderso/biology/resources/writing/
HTWtoc.html

I also enjoyed the 20-page guide written in January 2006 by Sarah Deel from Carleton College. She recommends to write the abstract in the past tense — I prefer to follow the guidelines of NASA on this one because the abstract needs to attract the reader into the paper, and

the present tense does that better than the past tense, yet I like her style of questions and answers. The title of the guide is *Lab Report Guide: How to Write in the Format of a Scientific Paper.*

http://serc.carleton.edu/files/sp/carl_ltc/quantitative_writing/examples/lab_report_guid.pdf

The document entitled *Scientific Writing Booklet* is compiled by Dr. Marc Tischler from the University of Arizona. It is a short 24-page booklet that, besides covering the various parts of a paper, explains the differences between active and passive voice and when to use which (see the *evolution of writing style* sites for more views on the use of the passive voice in scientific writing – and in particular the paper by Rodman).

www.biochem.arizona.edu/marc/Sci-Writing.pdf
www.jacweb.org/Archived_volumes/pdf_files/JAC2_Rodman.pdf

Do read the set of slides prepared by Simon Peyton Jones from Microsoft Research in Cambridge, England. It is entitled *How to write a great research paper*. I do not agree with everything he says: his definition of the introduction is too narrow (See chapter 13 and 14 of this book for more details), and I suspect that leaving all related work for the end of the paper is too late. Related works are most welcome in several sections of the paper: in the introduction, the discussion, and sometimes the methodology.

http://research.microsoft.com/Users/simonpj/papers/giving-a-talk/writing-a-paper-slides.pdf

I appreciated the brevity and common sensical recommendations of two web pages from Professor Railsback from the University of Georgia entitled *Some thoughts on writing a scientific paper or thesis* and *some comments on writing and editing (the latter is as important as the former).*

www.gly.uga.edu/railsback/writing1.html
www.gly.uga.edu/railsback/writing2.html

Finally, There are two documents scientists cannot ignore because each and everyone is responsible for the reputation of Science (let alone our own reputation); The first one, a 76 page PDF document

published in 2006 by the Council of Science Editors (CSE), is entitled *CSE's White Paper on Promoting Integrity in Scientific Journal Publications*. It has two main parts: 1) "Roles and responsibilities in publishing," and 2) "Identifying research misconduct and guidelines for action."

> www.councilscienceeditors.org/
> www.councilscienceeditors.org/files/public/entire_whitepaper.pdf

The second document comes from my favorite website:

> *www.the-scientist.com*

They have an online daily, which I scan every day. All scientists, not just those in life science, will find something of value in it. For example, they recently had an article on a *glossary of retractions* March 2, 2007, listing out the various degrees of retraction, from the simple correction, to the expression of concern, the partial retraction, and the total retraction with or without permission. For each situation, a sample letter is given.

Dos and Don'ts of Technical Writing by Dennis R. Morgan, an article published page 22 to 25 of the August/September 2005 issue of IEEE Potentials.

In this editorial, you find the most common errors authors make when writing formulas. A whole section of the paper covers mathematics. The General Writing section mostly covers punctuation errors, and the Word Usage section briefly reviews the most common word errors peppering the text of engineering journals. The free version of this article is found online through a simple search on its title, but it should normally be bought on the IEEE website.

> http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber = 1502501&am
> p;isnumber = 32235

Citing and referencing guide: Harvard style is a fifteen page pdf file from the Imperial College in London. The article advocates the Name Year referencing scheme where the name of the author is immediately followed by the year of publication of the document. This guide is less extensive than the massive book of the Council of Science Editors, but if the journal you are targeting for publication requires you to follow the Name Year scheme, this article is for you. The citing

guide makes clear how to cite, and the referencing guide details how to reference the usual journal or conference papers (electronic or in print), but also more unusual documents such as podcasts, patents, blogs or online lectures.

www3.imperial.ac.uk/library/subjectsandsupport/publications

On Writing Process and Reviewers

Bates college has a web page entitled *A Strategy for Writing Up Research Results* which covers the whole writing process, including the all essential peer-review.

http://abacus.bates.edu/ ~ ganderso/biology/resources/writing/ HTWstrategy.html

The peer review occurs after your draft is complete. You have made sure your logic is sound and you have tried to make your paper concise without losing clarity. You then show your article to a reader not familiar with your research. If that reader is used to reviewing other people's papers, then you are all set; More often, however, you may have to explain what you expect of him or her. The peer-review form put together by Bates college is great. It helps a reader, even you, check out a paper methodically before sending it to the unknown super-reviewer whose publishing veto power will mean everything to you and your future.

www.bates.edu/ ~ ganderso/biology/resources/peerreview.html

The OWL writing center from Purdue University has a page entitled *General Writing Concerns (Planning/Writing/Revising/ Genres)*. The many links to the extensive contents of its site are divided into four categories, two of which are about the writing process: "Planning/Starting to Write" and "Revising/Editing/ Proofreading".

http://owl.english.purdue.edu/handouts/general/

I recently found a 55 page *Guide to Academic Writing* from the University of the Western Cape, completed in September 2003, written by Nelleke Bak. Although its emphasis is on writing a thesis, many of its topics are applicable to the writing of a scientific paper.

There is an interesting section on literature review and on critical reading (you will need to read many papers to prepare the related work section of your paper). It also covers the topics of plagiarism and citations.

www.uwc.ac.za/usrfiles/users/4196080617/Microsoft_Word_-_ACADEMICWRITINGGUIDEcompletedraft.pdf

You may remember Stephen D. Senturia, MIT professor and Senior editor of the JMEMS journal, who was mentioned in chapter 14. Dr. Senturia, shares his great insights on reviewers in the paper *How to avoid the reviewer's axe: One editor's view.*

www-bsac.eecs.berkeley.edu/ ~ muller/jmems.web/sds_editorial_june_2003.pdf

Do visit the following website to understand what is required of a typical reviewer.

www.rsc.org/Publishing/ReSourCe/RefereeGuidelines/Refereeing
ProcedureAndPolicy/index.asp

Books on Scientific Writing

A PH.D. Is Not Enough

Peter J. Feibelman

The book is worth buying. I bought twenty copies and use them in my seminars on scientific writing. With 109 small pages of large text divided into eight chapters, the book can be read in less than two hours.

I like this book because every time I re-read it, I discover new wisdom. Professor Feibelman is not only an experienced scientist, he is also a good writer. You will find information on his book at the following address:

www.amazon.com/Ph-D-Not-Enough-Survival-Science/dp/
0201626632

You may be curious to know how I use this book in the seminar: I have the participants read chapter 4 entitled "*Writing Papers: Publishing Without Perishing,*" 13 short pages in all. I then put participants into groups and ask them questions on the writing process,

the relative importance of the various parts of a scientific paper, a successful publishing strategy, and the qualities that a successful publishing scientist should possess. This chapter contains the answers to all. Its richness is astonishing. No wonder Carl Djerassi, in a quote prominently featured on top of the cover, writes: "It took me over forty years to learn from experience what can be learned in one hour from this guide".

Scientific Style & Format

Council of Science Editors (CSE)

This massive book (658 pages), in its seventh edition (2006), is an authoritative reference book.

www.councilscienceeditors.org/

The book subtitle is "The CSE Manual for Authors, Editors, and Publishers". It is a resource that you will probably not read in its entirety. It has 32 chapters divided into four sections. I spent time adding the number of subheadings one to four level deep, and found 935 of them! You can imagine the level of detail available here. This book covers the special scientific conventions for writing everything scientific, from the electromagnetic spectrum and astronomical objects and time systems to the chemical formulae and genes, chromosomes, and related molecules.

I use it often to answer questions such as "How do I reference a paper to be published, but not published yet?" or "How do I refer to a paper seen on the Internet?" or "How do I refer to non-text information, like a video, an animation...?"

Chapter 27 is about Journal Style and Format. Of interest will be section 27.7 "Journal Articles and Their Parts". But there is also a chapter on Tables, Figures, and indexes (chapter 30), on proof correction (chapter 32), on copyright (chapter 3), and three chapters on grammar (chapters 5, 6, and 7).

I recommend that you ask your library to buy it if they haven't already got a copy. The latest edition has been updated to take into account the rapid developments in life sciences. So if your research is in life sciences, get the latest edition!

Successful scientific writing

Janice R Matthews, John M Bowen, and Robert W Matthews

The subtitle of this book reveals its targeted audience:
"A step-by-step guide for the biological and medical sciences".

www.amazon.com/Successful-Scientific-Writing-Step-step/dp/0521789621

A thoughtful and thorough 230-page guide that goes down to the type of practical advice you don't expect, as in page 71 of the book in a subheading about writing momentum, and how to keep it.

"If you must eat [while writing], choose foods that require only one hand, don't stain papers, and don't make you thirsty or (worse yet) sleepy."

This book is thorough and complete. It answers many of the questions the novice writer has when facing the daunting task of writing a first draft — including answers to the dreaded writer's block (pages 72 and 73).

A large 24-page chapter on visuals entitled "Supporting the text with tables and figures" is well worth reading. And naturally, three chapters on grammar and composition complete the book. These chapters are not dry. They are always related to the scientific article, making the grammar points immediately relevant to both reader and writer.

The font size is small (10 point) so there is much material to digest in this book (including many tables). The authors break down the text using numerous subheadings. It can be overwhelming at times, and not all information given is of immediate use. So read selectively, but take the time to do the many exercises given (the correct answers are given at the end of the book). Although the guide is targeted for life science researchers, much of it applies to all scientific domains. This is probably the type of book you can read many times and always learn something new as your writing skills improve.

Expectations: Teaching Writing from the Reader's Perspective

George D.Gopen

George D.Gopen is a very interesting and provocative writer. I reviewed his book on Amazon.com.

www.amazon.com/Expectations-Teaching-Writing-Readers-Perspective/dp/0205296173

"I have been teaching Scientific writing. I helped many scientists but nobody was helping me understand why the sentences I had rewritten were far better than the originals, why paragraphs had regained energy and clarity. This made it difficult for me to transfer my knowledge to others. Then... through a Scientific American article, I discovered George Gopen and this book. I appreciated George Gopen's thoughtfulness. Part two of his book is for us, teachers of writing. Entitled "Pedagogy," it shares George's methods; It also answers the students' common objections to the reader approach. The book's appendix provides an interesting historical perspective on the topic. I recommend this book without reserve."

After reading this book, while home schooling my son, I discovered to my great surprise that French schoolteachers teach 9th graders the same grammatical principles. Who likes grammar! Is it not enough to enjoy reading and not worry about the reason for our enjoyment? Text progression techniques are essential to reading speed and clarity of understanding, yet they are too succinctly presented in books on writing, often under the title "transitions". What a shame!

The Craft of Scientific Writing

Michael Alley

I already mentioned the Penn State University website which provides grammatical exercises. Many of these advise the reader to look for answers in chapters of Michael Alley's 1995 book. You will find more information on the Amazon web page, including some eloquent reviews of the book.

> www.amazon.com/Craft-Scientific-Writing-Michael-Alley/dp/0387947663

When I was asked to conduct scientific writing seminars, at the turn of the last century - it's not that long ago :-), this was the first book I selected to prepare the course. Being French (we don't like rules, but we love principles), and having discovered that participants sent to the class felt as if they had been sent to the gulag, I decided to move away from the dryness of grammar. Michael's book was a refreshing departure from grammar and sentence structure books.

Like the book of Professor Feibelman, the pages in his book are small and the large text readable. The grammatical part is pushed back to the appendix "avoiding the pitfalls of grammar and punctuation". The examples are not too technical and the author frequently mentions the readers, thereby balancing both reader and writer points of view. The book gave me the idea to change the focus of attention from the writer to the reader. All participants in the writing class are expert readers. They can relate better to the solutions proposed to improve their writing after they understand what writer-created problems cause inefficient reading. Later on, the book by George Gopen allowed me to refine my understanding of the reader and I was able to improve on Michael Alley's chapter 9 ("Language: Being Fluid).

The book focuses on structure (chapters 2 and 3), language (chapters 3 to 9), and illustrations (chapter 10 and 11). These chapters help you identify the difference between good and bad writing. The remaining chapters are only brief introductory chapters on presentations, writing genres, and the writing process.

Writing for Computer Science

Author Justin Zobel

Here is my review of the book on its amazon.com page.

www.amazon.com/Writing-Computer-Science-Justin-Zobel/dp/1852338024

"My roots are in computer science where I started my professional life. Therefore I was attracted by this title. I borrowed the book from the library and read it from cover to cover. I found myself wanting to use a yellow highlighter pen over the text of quite a few chapters.

The pertinence of the examples chosen proves that Justin has reviewed his fair share of papers. If you referee papers, chapter twelve on "Refereeing" is for you. If you are in computer science or if you use computers in your research, then chapter 5 ("Mathematics"), 7 ("Algorithms"), and 11 ("Experimentation") are written specifically for you.

Scientists outside of computer science are not left out. Chapters two to four are about style and grammar. His examples will bring

a smile to your face as you recognize your own mistakes. I cannot say I always agree with Justin's examples. For example page 32, he writes: "Beginning a paper by stating that a topic is popular or that a problem is important is flat and uninspiring." I could not agree more, of course. He illustrates such a flat start with the great example "Use of digital libraries is increasingly common". But then the "may well be preferable" counter example that follows has the same problem: "Digital libraries provide fast access to large numbers of documents". It uses two imprecise adjectives and does not enhance the knowledge of even the most junior researcher in computer science. It is easy to critique anyone's work, and I am sure mine deserves its fair share of criticism; Justin's book has the merit of systematically illustrating the principles of writing he recommends to the readers.

Chapter 6 on graphs, figures, and tables gives many examples. Justin believes in making figures "less dependent on the paper's text" (page 112) by making their caption more informative. I do too, and even go one step beyond his recommendation: make your figures, tables, and graphs stand-alone through a complete caption.

Chapter 9 gives general instructions for writing the various parts of a scientific paper, from its title to its conclusions".

This page intentionally left blank

Epilogue: Your Future Work

My work ends here, and yours starts now. Writing a book is not easy. Sometimes, only after rewriting and rereading a chapter for the nth time does its structure finally appear. Sometimes, the structure of a chapter is in place even before its contents, and the hard work consists in finding examples and metaphors to make things clear. But one thing is constant: the more you rewrite, the clearer your paper becomes. The memorable words of Marc H. Raibert, President of Boston Dynamics, ex-head of the Leg lab at CMU and MIT still ring in my ears: 'Good writing is bad writing that was rewritten'. How true!

Writing is hard. To avoid making it harder than it already is, start writing your paper as soon as you can. It will be less painful, and even pleasant at times. At the beginning, write shorter papers (e.g. Extended abstracts or letters to journals). You can write more of them, and chances are, a few will be accepted. On the way, a few good reviewers will encourage you and pinpoint the shortcomings in your writing, while a few good readers will tell you where you lack clarity.

Each chapter in this book contains exercises that involve readers. Value your reader friends. The time they spend reading your paper is their gift to you. Accept their remarks without reservation and with a grateful heart. Thank them for their help and return the favor (being French, I recommend giving them a bottle of red Bordeaux for their services, but feel free to offer other vintages). Do not take negative remarks personally; instead, consider them as golden opportunities to improve your writing. It is pointless to try to justify yourself because, in the end, **the reader is always right**, and in particular, the reviewers. Just take note of their remarks and questions, and work to remove whatever caused them to stumble. Do not argue.

Let your introduction convey research that is exciting, and let that excitement become your readers' excitement as they look forward to

the future your research has opened. Show the world that scientific papers are interesting to read. Create expectations, drive reading forward, sustain the attention, and decrease the demands on you readers' memory. To make reading as smooth as silk, iron out the quirks in your drafts with the steam of your efforts.

And may the fun of writing be with you!

Index

Clear

Concise

Convincing (assuming your text/visuals are already clear)