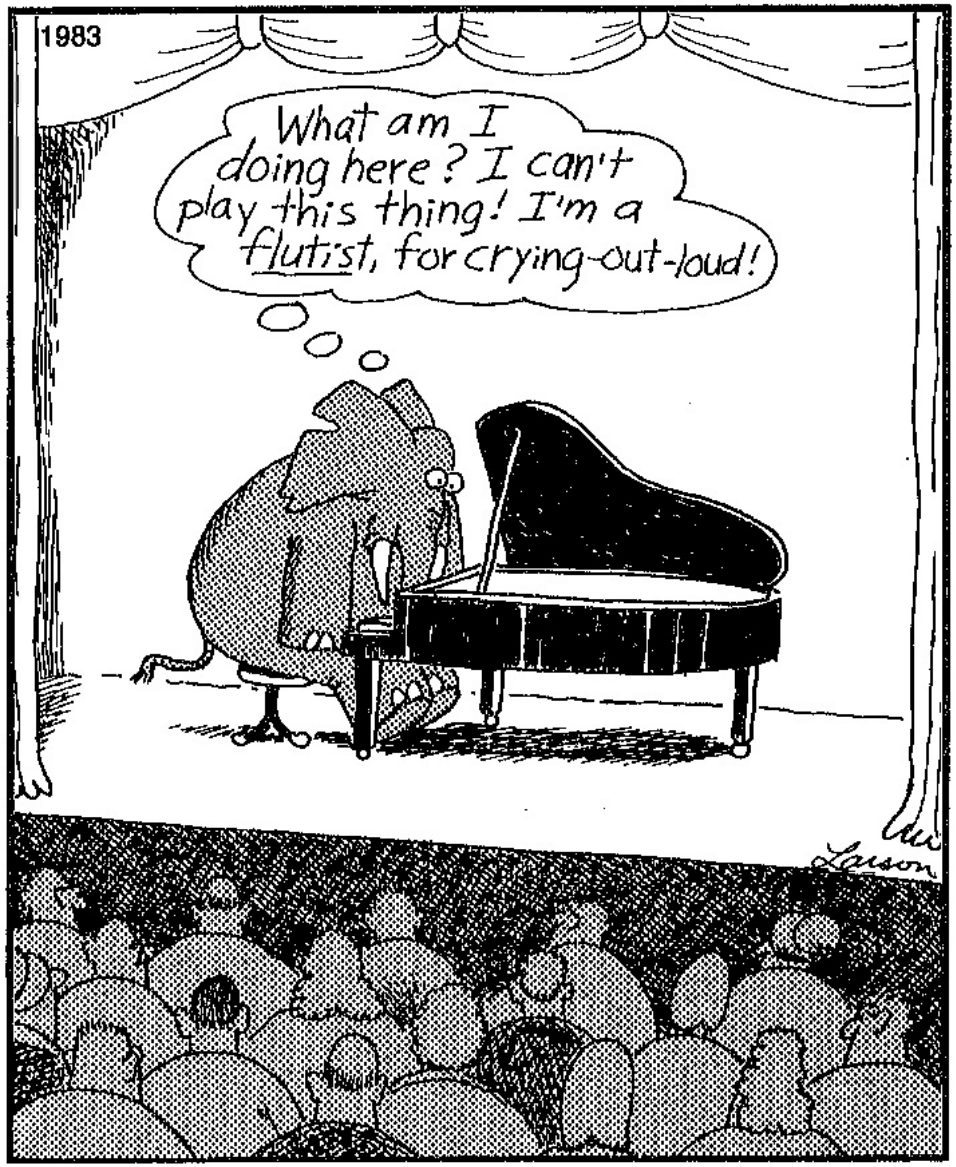
How To Write (in 2×75 mins)

Ken Rice and Don Percival

STAT/BIOST 572

April 25, 2011

Overture . . .



The elephant's nightmare

Outline

- Why writing well matters
- Reminders: grammar, verbs, and using the right word
- Edit edit edit!
- Avoiding nominalizations
- Creating characters
- Stress, and creating 'flow'
- Writing technical material (i.e. statistics and/or math)
- Use of BibTeX
- Planning – and planning your literature review

Time permitting, there will be some in-class exercises.

Note: We'll give *some* essentials, mostly just advice. As with presentation, aim to acquire out a writing style that works for you.

Background

Several of you have been teachers – and thus need little further help getting technical material across in talks

How about writing? How many of you;

- Have had jobs that involve substantial amounts of writing? e.g. journalism, or editing. (Don't count TA positions – unless you wrote much more than students who took the class)
- Write for fun? (e.g. blog entries... or an upcoming NYT bestseller)
- Took science and math classes partly because they *did not* require writing lots of essays? (be honest!)

Your reports/motivation

In 572, you'll write two or three reports;

1. **Literature review:** (2-4 pages) Summarizing the existing literature, and how your paper fits in
2. **Final report:** ($\sim$20 pages) Literature review + explanation/replication of what the paper does, and your critique of the paper's proposed methodology
3. **Prelim:** (not everyone) The same as the Final Report, but with extra revisions/improvements

The longer reports will look very like early chapters in a PhD dissertation.

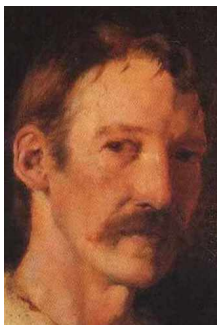
NB Verbatim reproduction of the original paper gets no credit; 'in your own words', please.

Your reports/motivation

It's *not easy* to do this well – as these experts note;

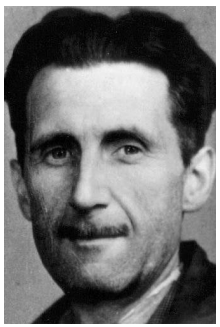
Easy reading is damn hard writing

Nathaniel Hawthorne (1804 – 1864) Novelist



The difficulty is not to write, but to write what you mean

Robert Louis Stevenson (1850 – 1894) Novelist & Travel Writer



Writing is a horrible, exhausting struggle, like a long bout of some painful illness

George Orwell (1903 – 1950) Author and Journalist

Your reports/motivation

Let's rebound a bit from Orwellian gloom and doom . . .

. . . writing is a solitary occupation, and one of its hazards is loneliness. But an advantage of loneliness is privacy, autonomy, freedom.

Joyce Carol Oates (1938 –) Novelist and Professor in Humanities at Princeton University

Note: Prof. Oates gave a public lecture in Seattle on Monday!

Your reports/motivation

But what's the *point*? Why do all that writing?

Writing and learning and thinking are the same process

William Zinsser (1922 –) Author, 'On Writing Well'

Learn as much by writing as by reading

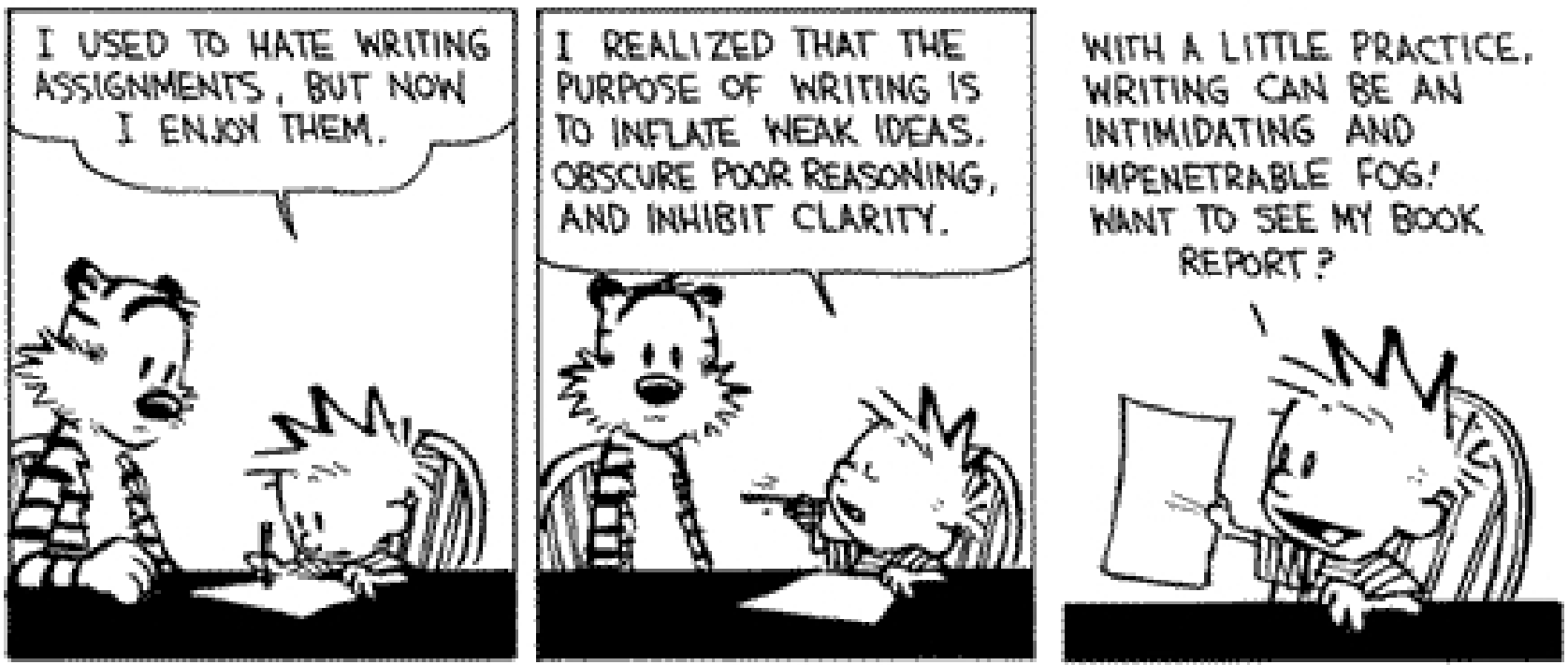
Baron John Acton (1834–1902) Historian and Moralist

Writing is the supreme solace

W Somerset Maughn (1874–1985) Playwright & Novelist

Your reports/motivation

So, writing well helps *you* think well, i.e. clearly. According to one expert*, what does the audience get from it?



* ... *yet another* philosophically-minded non-alive white guy

Your reports/motivation

Just as with talks, you need to think about how your *audience* will react – and this doesn't mean valuing style over substance;

- Readers *will not bother* reading what they can't follow
  $\Rightarrow$ bad writing wastes everyone's time
- Readers *will not believe* what they can't follow
  $\Rightarrow$ bad writing suggests you are (statistically) braindead*
- Readers who persevere with bad writing (e.g. referees, examiners) *will not like* whatever point it makes
  $\Rightarrow$ bad writing costs exams, papers, jobs, grants, etc

Reading really *good* writing can be a pleasure; a reasonable goal is writing that is not distracting – or at least not painful.

* No, this is not fair. But it still happens

Basics

The course text (Williams and Colomb) is an excellent guide to writing with clarity and grace.

Before we get there, we'll review a few more fundamental points;

- Word choice
- Some grammar
- Punctuation
- Not-good 'rules' of good writing
- Simple careless mistakes

Basics: Word choice

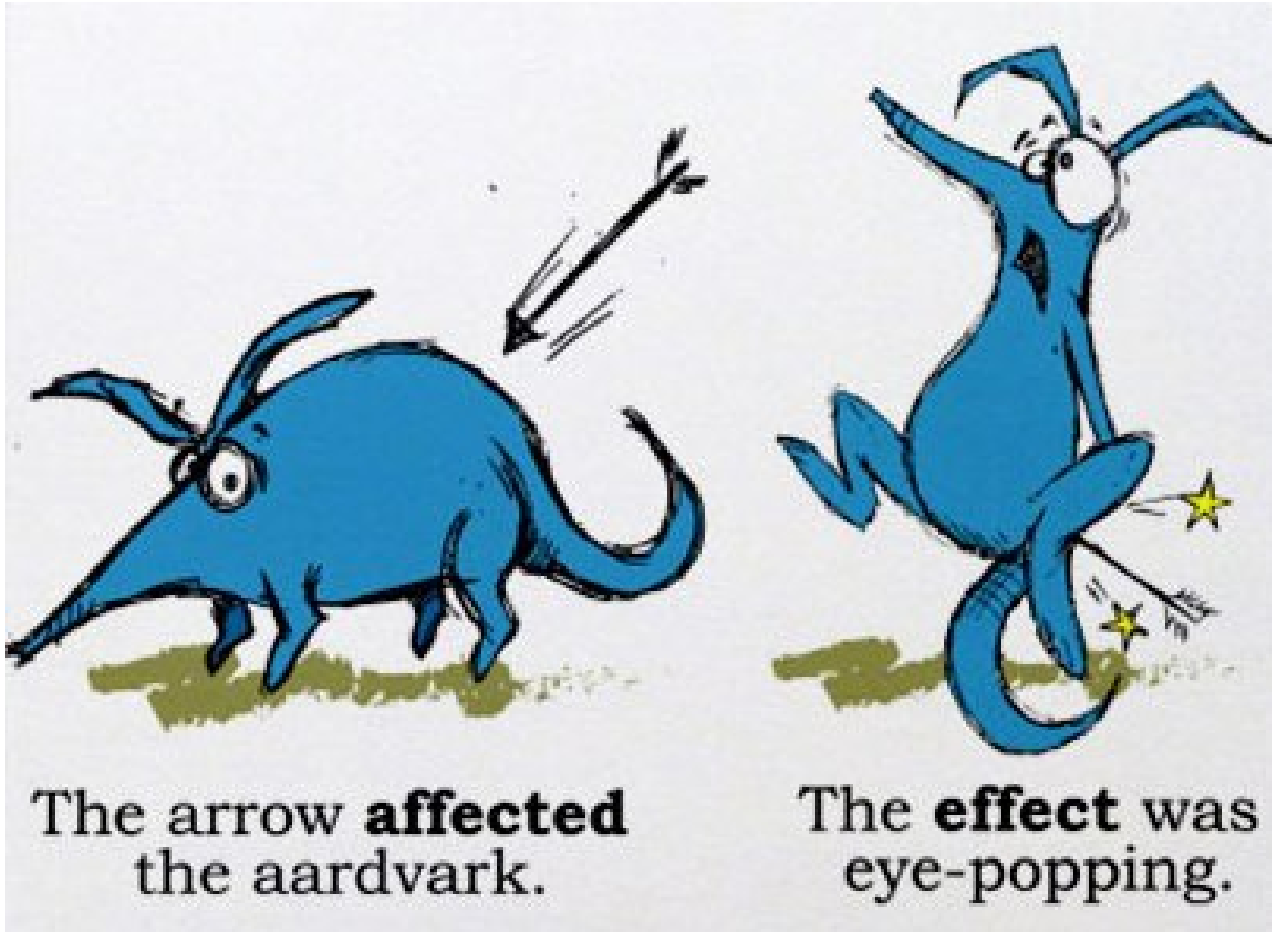
Use words you understand – and use a dictionary

Basics: Word choice

Use words you understand – and use a dictionary

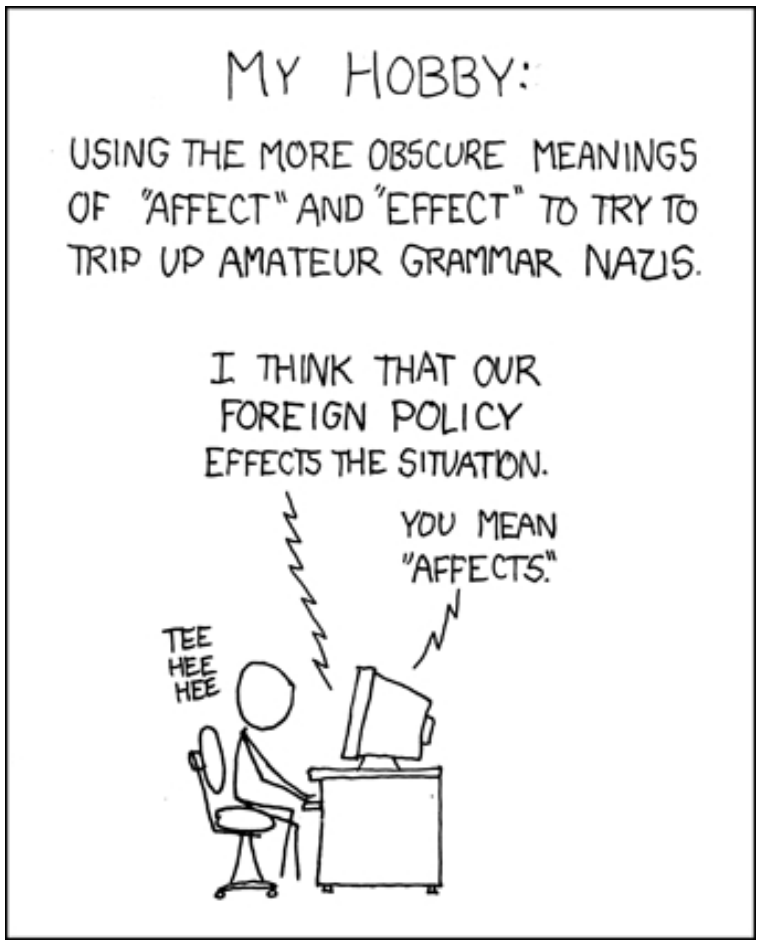
Basics: Word choice

You are expect to know how to use statistical terminology, e.g. the main **e**ffect of X. And causal arrows...



Basics: Word choice

Technical terms may have more than one ‘lay’ meaning;



Basics: Word choice

Avoid redundancy;

Basics: Word choice

Avoid redundancy;

Basics: Word choice

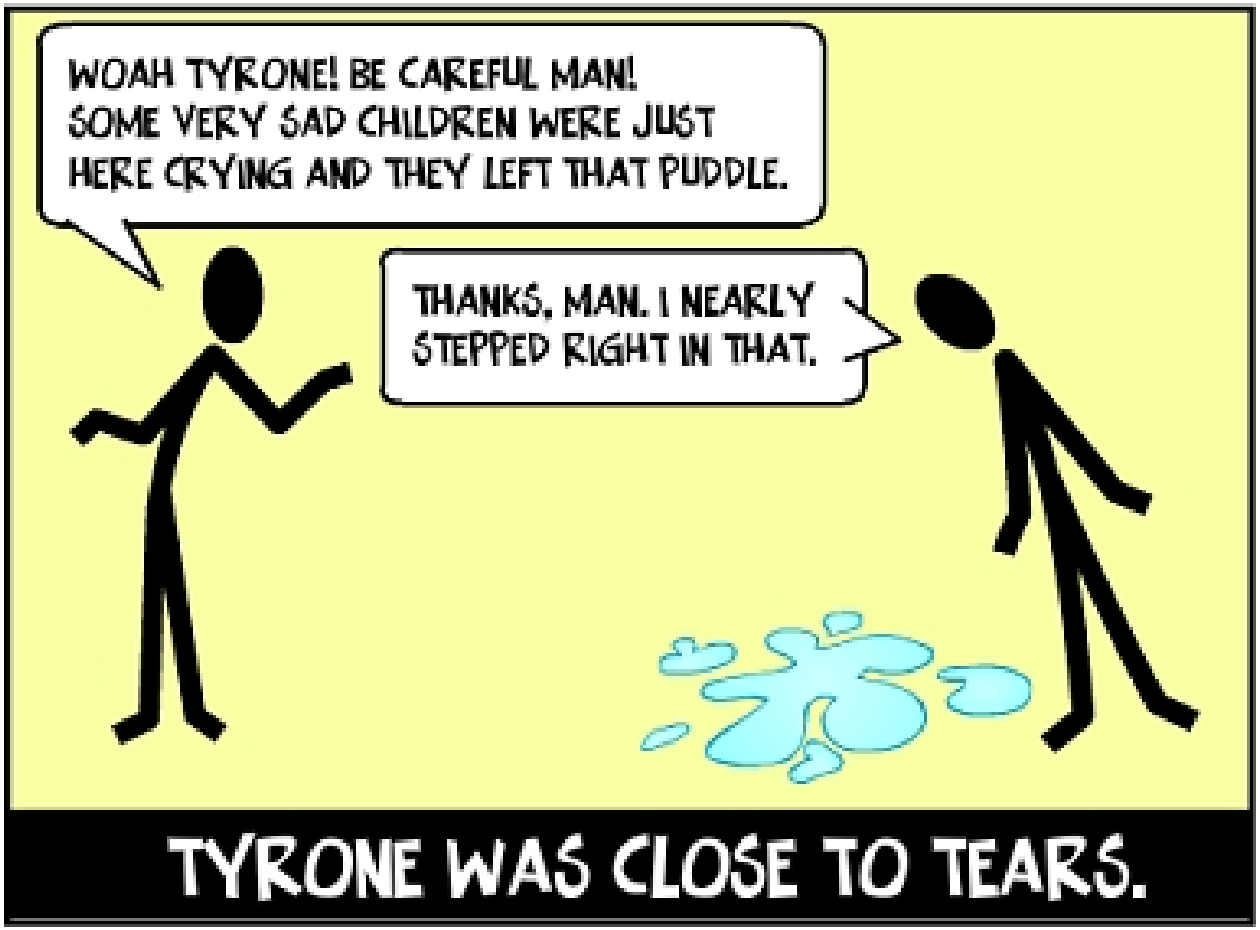
Avoid redundancy *when possible*;

Basics: Word choice

Aim to use simple, unambiguous words;

Basics: Word choice

... and avoid clichés (like the plague!)



Basics: Word choice

Less critical: what's the fewer of two evils?

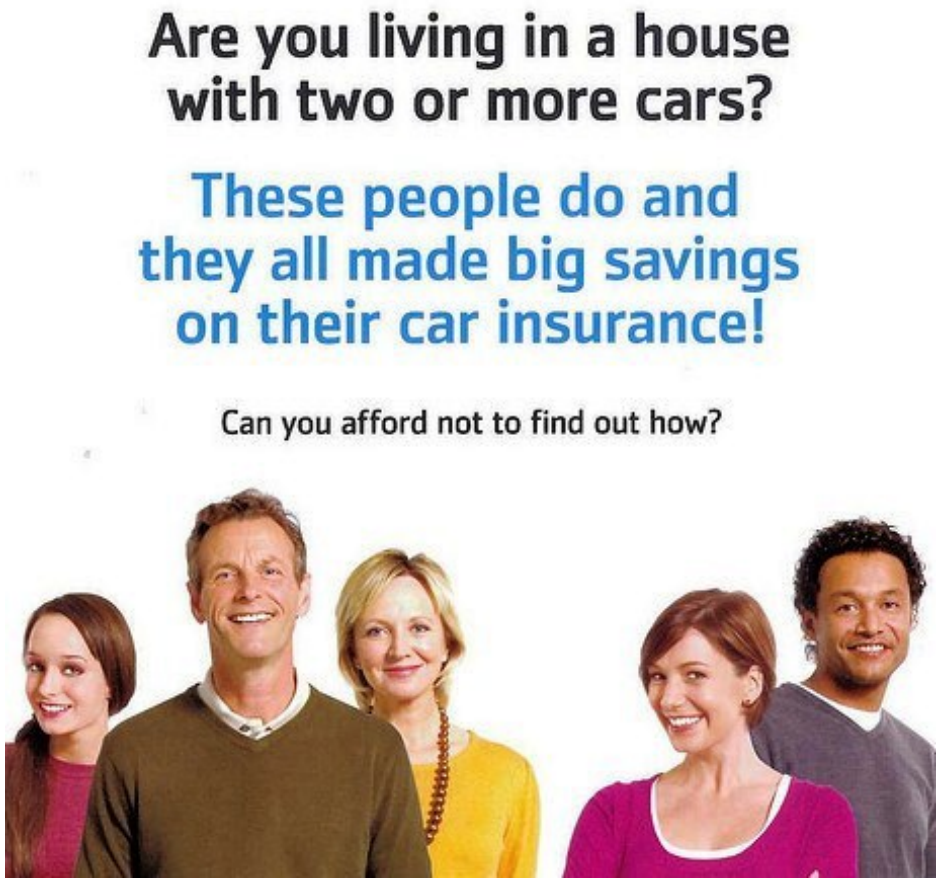
Basics: Word choice

Statistics is hard enough, without 'sentencejunk';

- Say 'use' not 'utilize'
- Delete 'very's. (Replace them with 'damned' first, if this helps – à la Mark Twain)
- 'We investigate' not 'We conduct an investigation of'
- We used X in order to reduce bias
  $\rightarrow$ We used X to reduce bias
- X bears some resemblance to Y
  $\rightarrow$ X is similar to Y
- If you need a specific technical term (e.g. 'estimate') **just use it** – again and again. Resist the temptation to substitute 'guess', 'reckoning', 'assessment' etc.

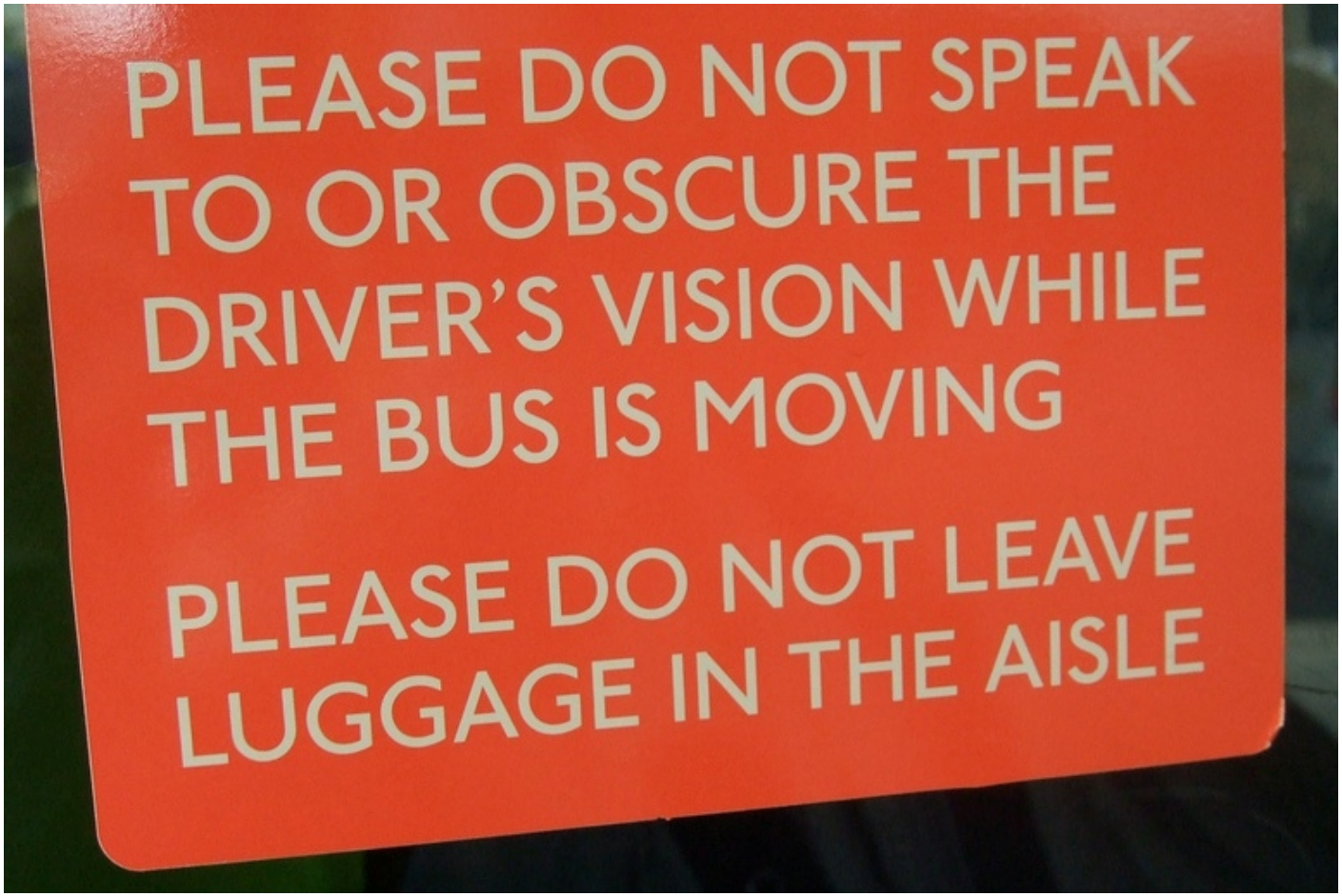
Basics: Correct grammar

An example: verbs should match their subjects



Basics: Correct grammar

Another one: *multiple* verbs should match their object



Basics: Correct grammar

And another one: what's wrong here?

Basics: Correct grammar

What's wrong with these?

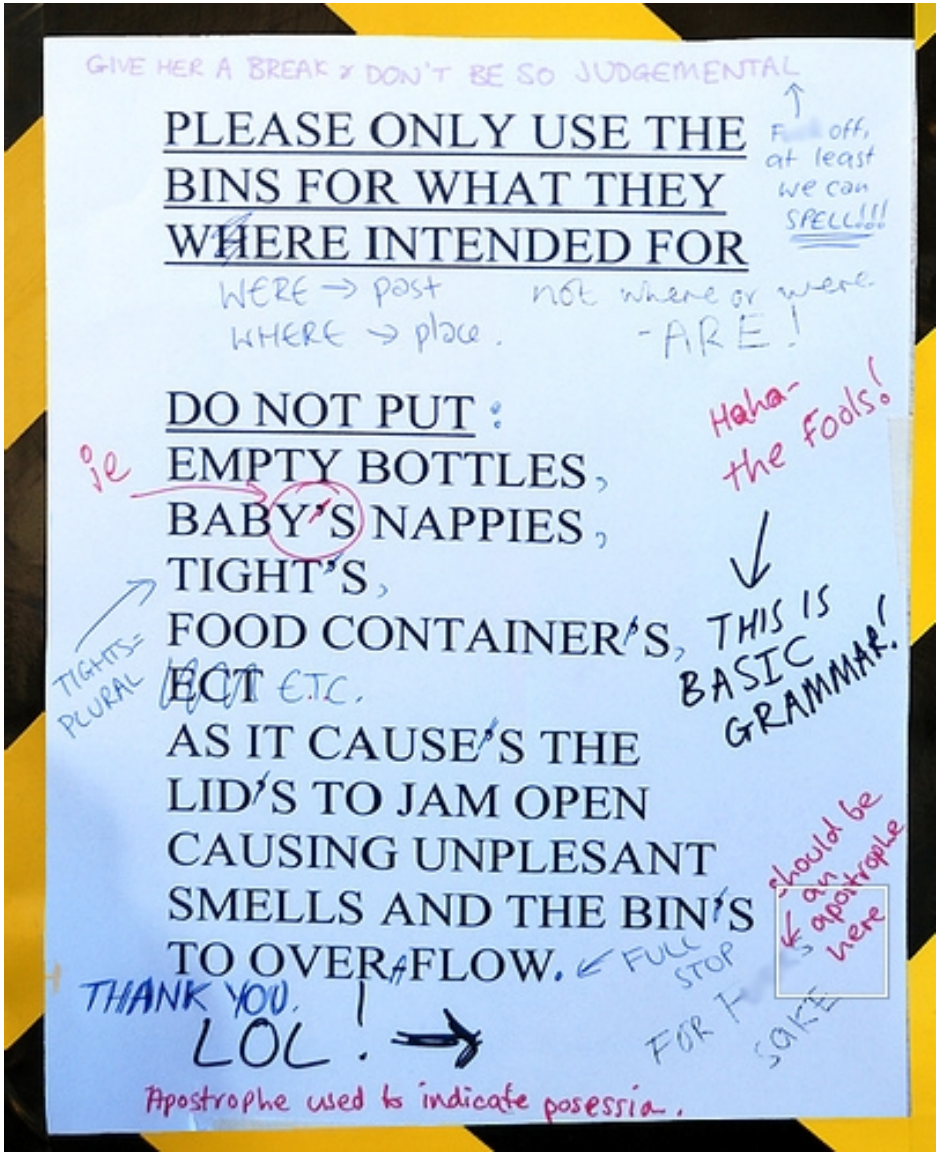
- The formulae in (3) and (17) is therefore equivalent
- Ken and his fellow speed-freak Don commutes every day by bike
- SIMEX is one of the methods that was considered by Carroll and Stefanski
- Each of the students ask questions in class
  Few of the students asks questions in class

Microsoft's grammar check (green underlining) catches slips like these quite well; it can also spot mis-matching tenses.

Basics: Punctuation

Rightly or not, some readers get **very** upset – and distracted – by e.g. incorrect use of apostrophes, and similar mistakes

Hint: *it's* should never appear in formal writing



Basics: Punctuation

Dust to dust, ashes to ash's; what *could* this mean?

Basics: Punctuation

Sloppy punctuation *can* be misleading, as well as distracting. For example, if you're not quoting something, "scare quotes" often indicate sarcasm;

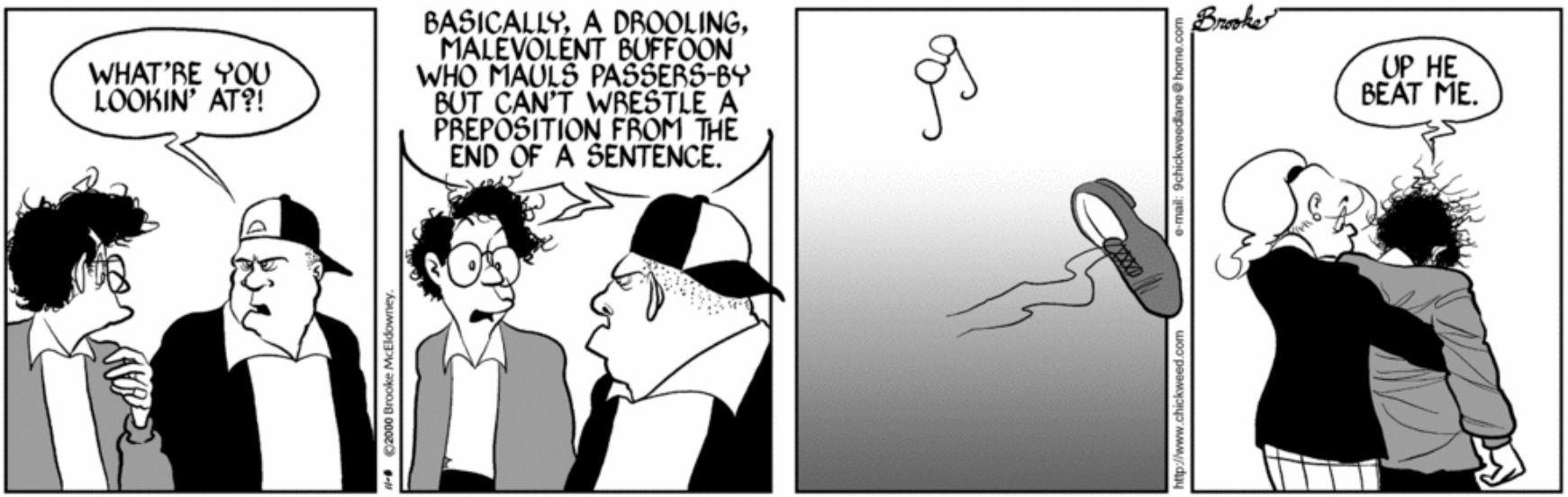
Basics: Punctuation

Punctuation is *helpful*. So use it.

Basics: Troll-avoidance

What silly rule should you never end a sentence with?

The sort of arrant pedantry...



What did you want to bring that book I didn't want to be read from out of up to?

Basics: Troll-avoidance

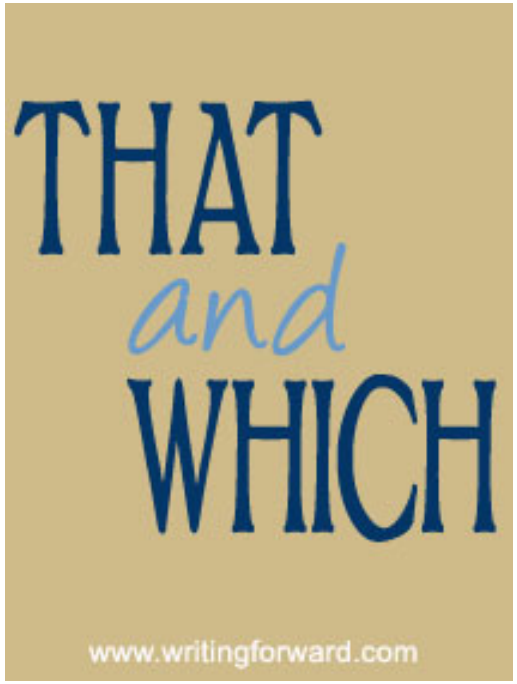
An example to boldly show the ‘weakness’ of split infinitives;

The effect is mild, but real. Halmos called rules like this “reactionary grammarian balderdash” – but Halmos won’t be reading *your* paper, and a grumpy troll might.

Basics: Troll-avoidance

For a full review of 'wicked whiches', see this site (right)

Or...



If it sounds alright to replace a 'which' by a 'that', then Strunk & White say replace it

Leslie Lamport (LaTeX guru)

This is advice **that** I usually follow, but be aware that Strunk & White *is* troll-oid in parts – and is also rather old.

Basics: Do not accidentally write garbage

Engage brain before hitting 'submit'... and watch out for hanging danglers

- I am out of the office with no email until April 20th
- δ is the minimally clinically important difference between groups
- After standing in boiling water for an hour, examine the flask
- When pickled, I like herring
- Ken is on trial. If convicted, his course notes will still be available on the class site
- Walking through the quad, the cherry trees seemed prettier than ever before

Basics: Do not accidentally write garbage

Editor's wife rented to 2 suspects, FBI says

Chicago Tribune 10/15/01

McDonald's Fries the Holy Grail for Potato Farmers

Wednesday, September 23, 2009
Associated Press

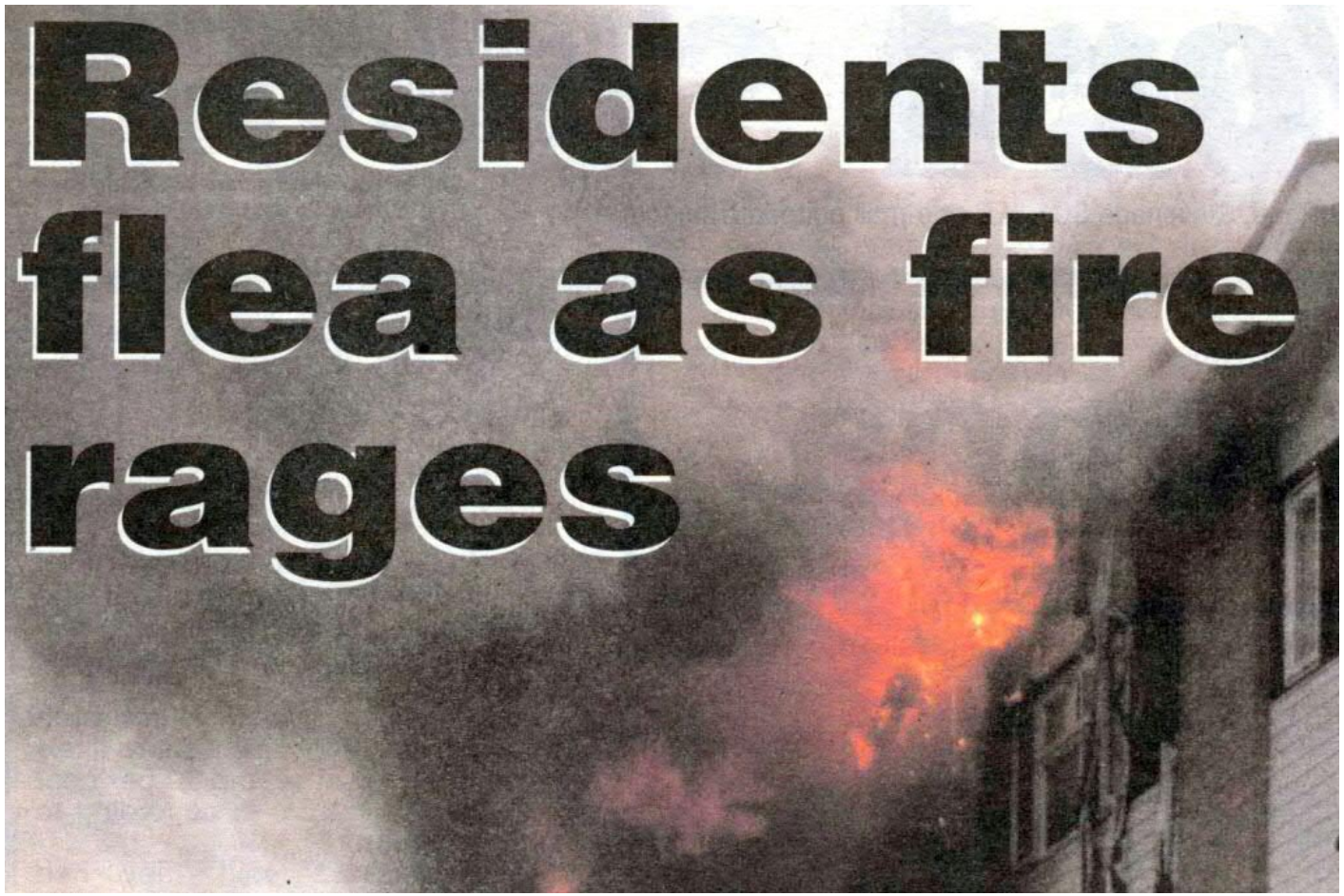
Print

KIMBERLY, Idaho — From the fields of Idaho to tasting rooms in suburban Chicago, potato farmers, researchers and industry representatives are in the midst of an elusive hunt: finding a new spud for McDonald's french fries.

Newspaper staff call these 'crashblossoms' – after the Japan Times' classic 'Violinist Linked to JAL Crash Blossoms'

Basics: Do not accidentally write garbage

Catching brainfarts *is* hard; (Letchworth & Baldock Comet)



Basics: Do not accidentally write garbage

... but saves confusion.

Basics: Do not accidentally write garbage

We can all forget an occasional ‘not’ – but when we do, the audience is *forced* to mind-read;

Basics: ... are your responsibility

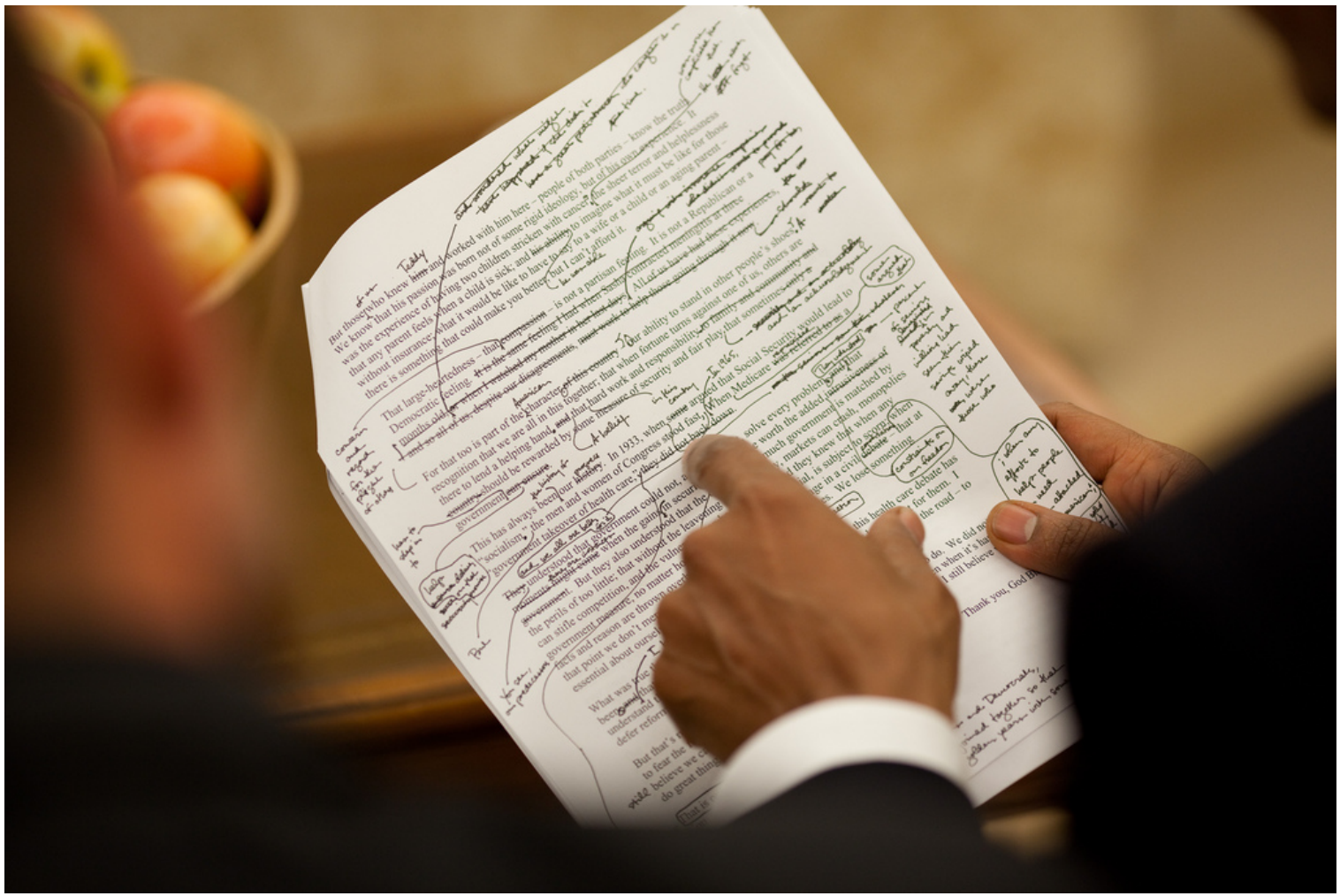
Your supervisor/professor is not there to teach you basic grammar and spelling

The more time and emotional energy she or he spends on correcting basic English usage, the less remains for issues of content or fine-tuning. You are responsible for mastering the basics of the language; save your supervisor's time for more substantive issues. A few glitches and non-parallel tenses will slip through your own careful editing, but there is no excuse for frequent ungrammatical sentences. Similarly, with word processors and spellcheckers having become standard writing tools, typos or other spelling errors should be very rare. Use a spellchecker before submitting anything to anyone.

`http://course1.winona.edu/mdelong/EcoLab/21%20Suggestions.html`

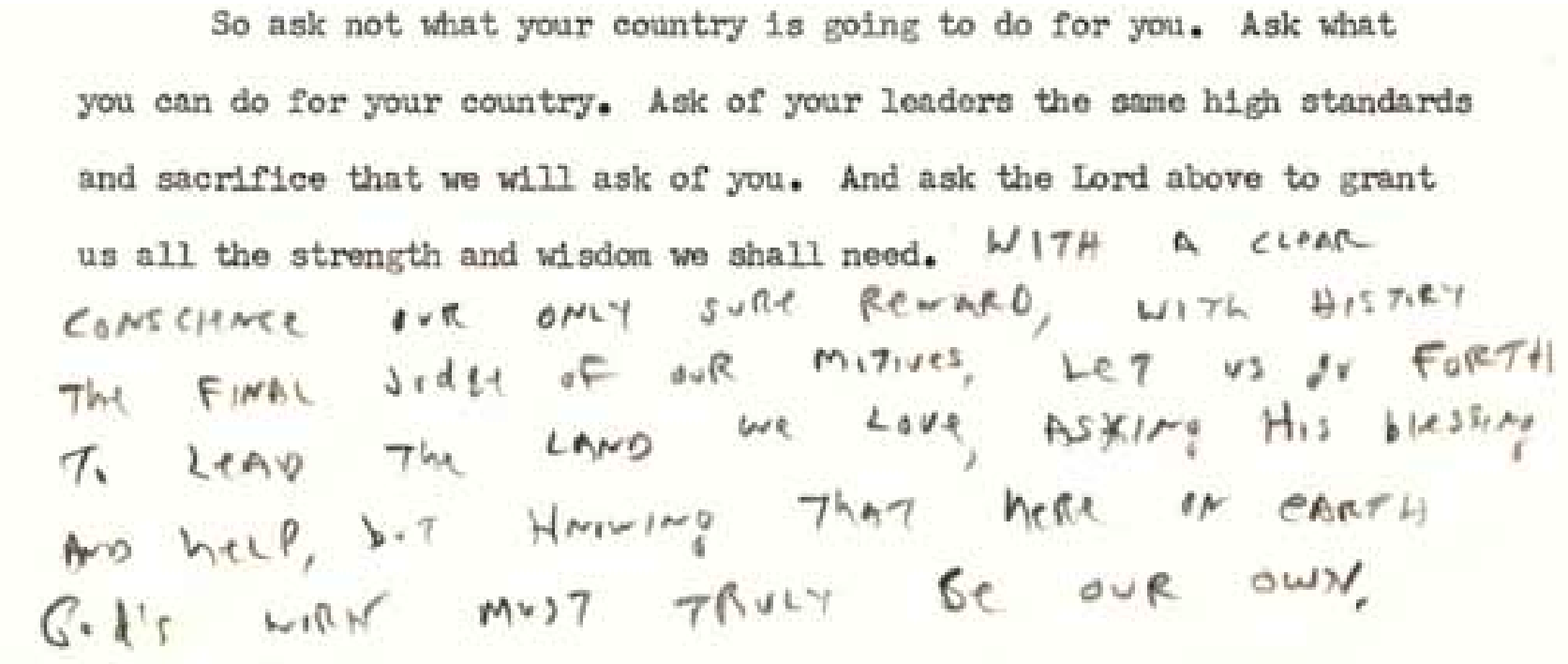
Edit! Edit! Edit!

From a speech on healthcare reform;



Edit! Edit! Edit!

Editing is not new – here's a draft of JFK's inaugural;



So ask not what your country is going to do for you. Ask what you can do for your country. Ask of your leaders the same high standards and sacrifice that we will ask of you. And ask the Lord above to grant us all the strength and wisdom we shall need. WITH A CLEAR CONSCIENCE OUR ONLY SURE REWARD, WITH HISTORY THE FINAL JUDGE OF OUR MOTIVES, LET US GO FORTH TO LEAD THE LAND WE LOVE, ASKING HIS BLESSING AND HELP, BUT KNOWING THAT HERE ON EARTH GOD'S WORK MUST TRULY BE OUR OWN.

Edit! Edit! Edit!

... and the version you may know better.

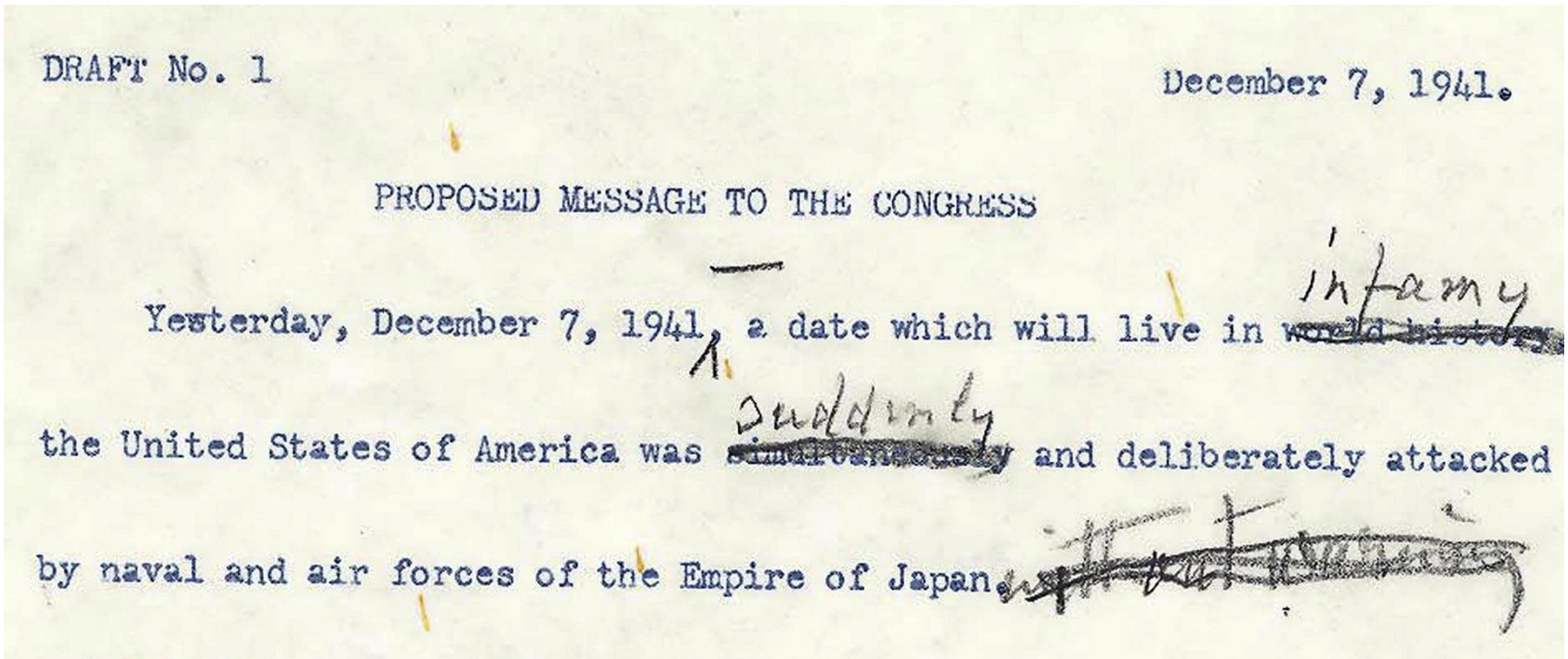
And so, my fellow Americans: ask not what your country will do for you -- ask what you can do for your country.

My fellow citizens of the world: ask not what America will do for you, but what together we can do for the freedom of man.

Finally, whether you are citizens of America or of the world, ask of us the same high standards of strength and sacrifice that we shall ask of you.

Edit! Edit! Edit!

Before him, FDR did the same thing, right until showtime;



DRAFT No. 1

December 7, 1941.

PROPOSED MESSAGE TO THE CONGRESS

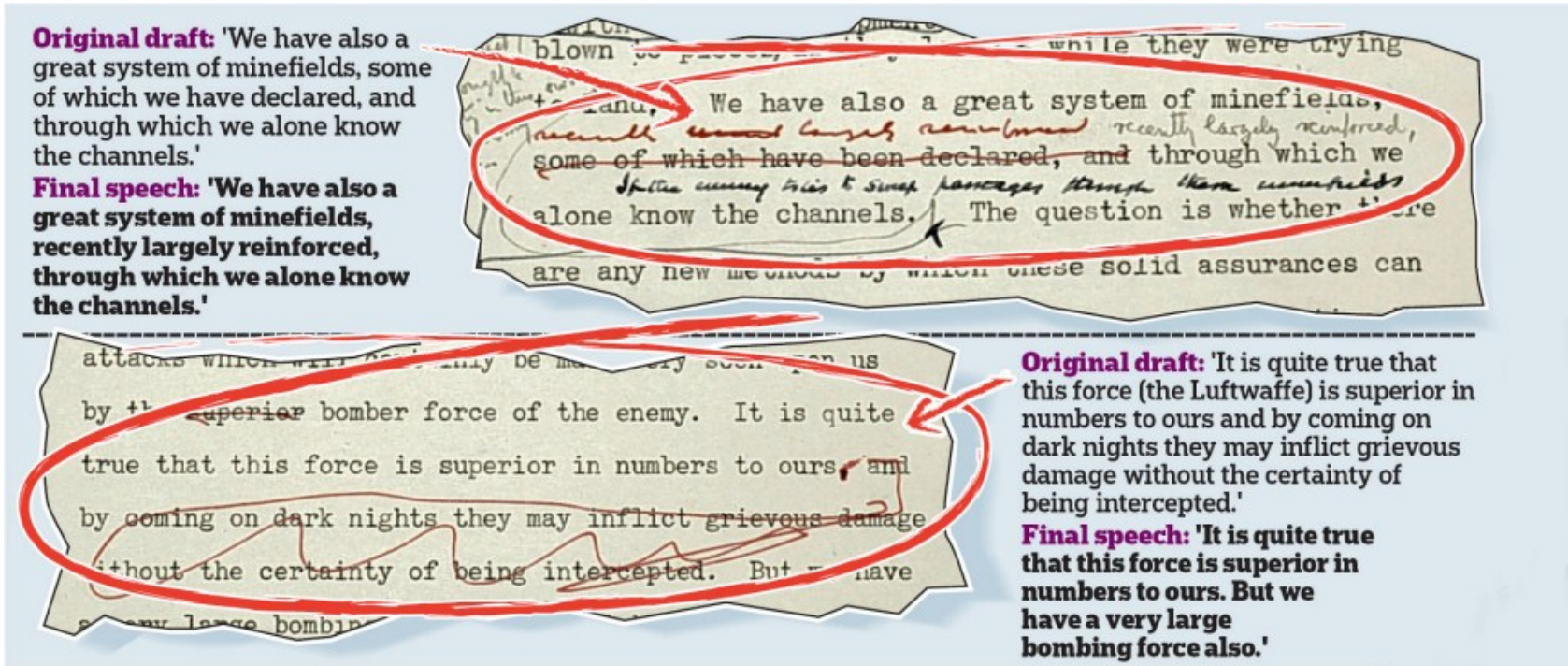
Yesterday, December 7, 1941, a date which will live in infamy

the United States of America was suddenly and deliberately attacked

by naval and air forces of the Empire of Japan

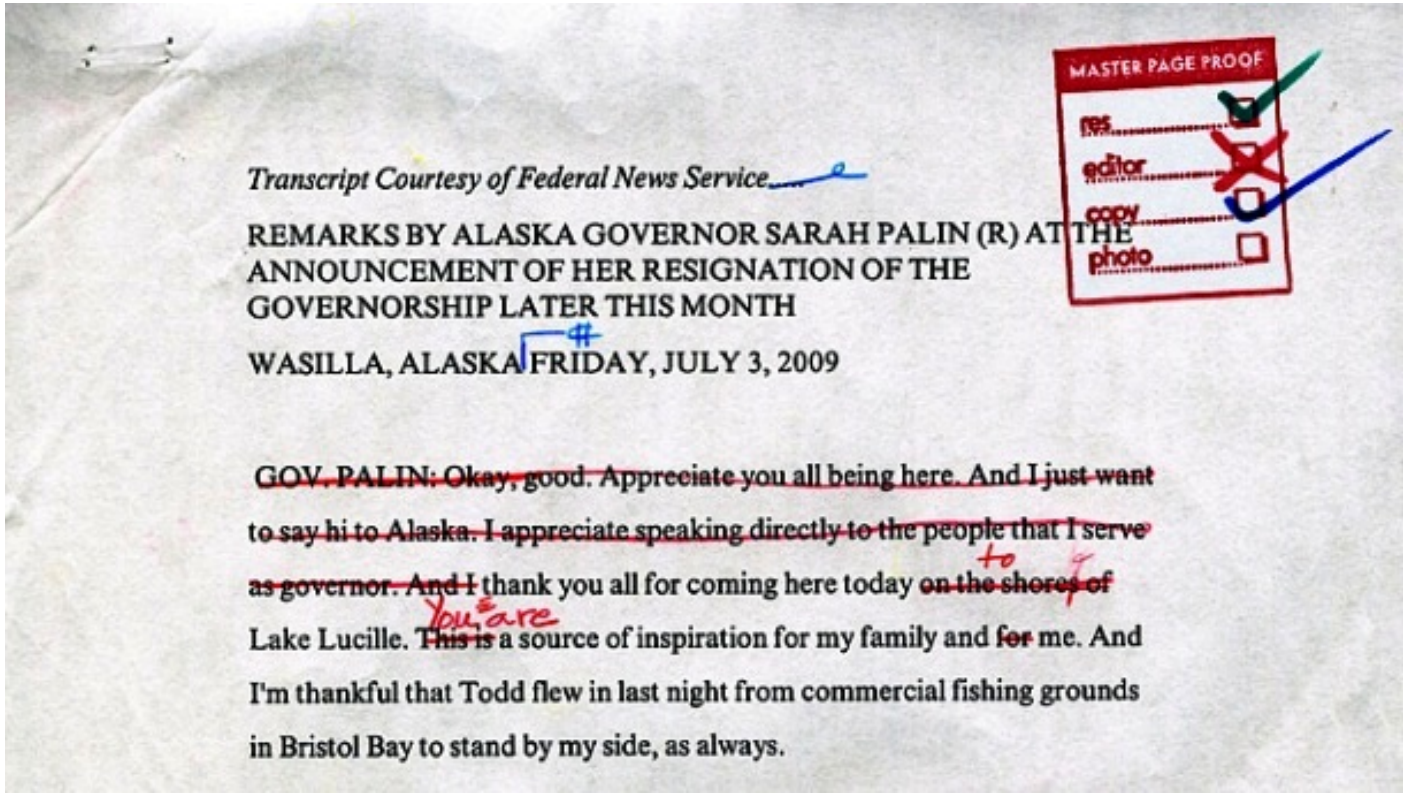
Edit! Edit! Edit!

Winston Churchill also re-drafted every last nuance;



Edit! Edit! Edit!

It seems Sarah Palin does not Edit[3]. So Vanity Fair did it for her... research, language and copy-editing;



Transcript Courtesy of Federal News Service

REMARKS BY ALASKA GOVERNOR SARAH PALIN (R) AT THE ANNOUNCEMENT OF HER RESIGNATION OF THE GOVERNORSHIP LATER THIS MONTH

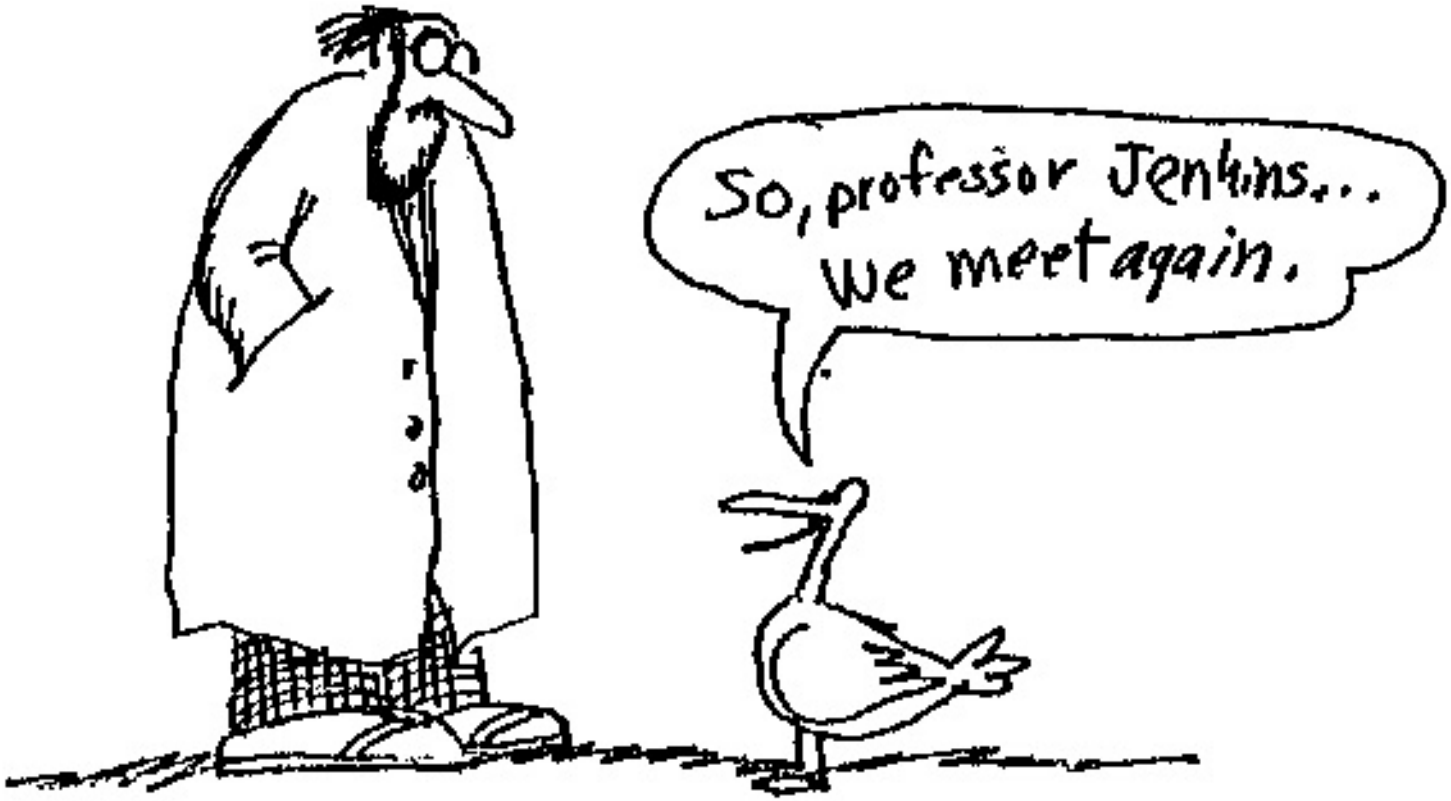
WASILLA, ALASKA FRIDAY, JULY 3, 2009

~~GOV. PALIN: Okay, good. Appreciate you all being here. And I just want to say hi to Alaska. I appreciate speaking directly to the people that I serve as governor. And I~~ thank you all for coming here today ~~on the shores of~~ Lake Lucille. ~~This is~~ a source of inspiration for my family and ~~for~~ me. And I'm thankful that Todd flew in last night from commercial fishing grounds in Bristol Bay to stand by my side, as always.

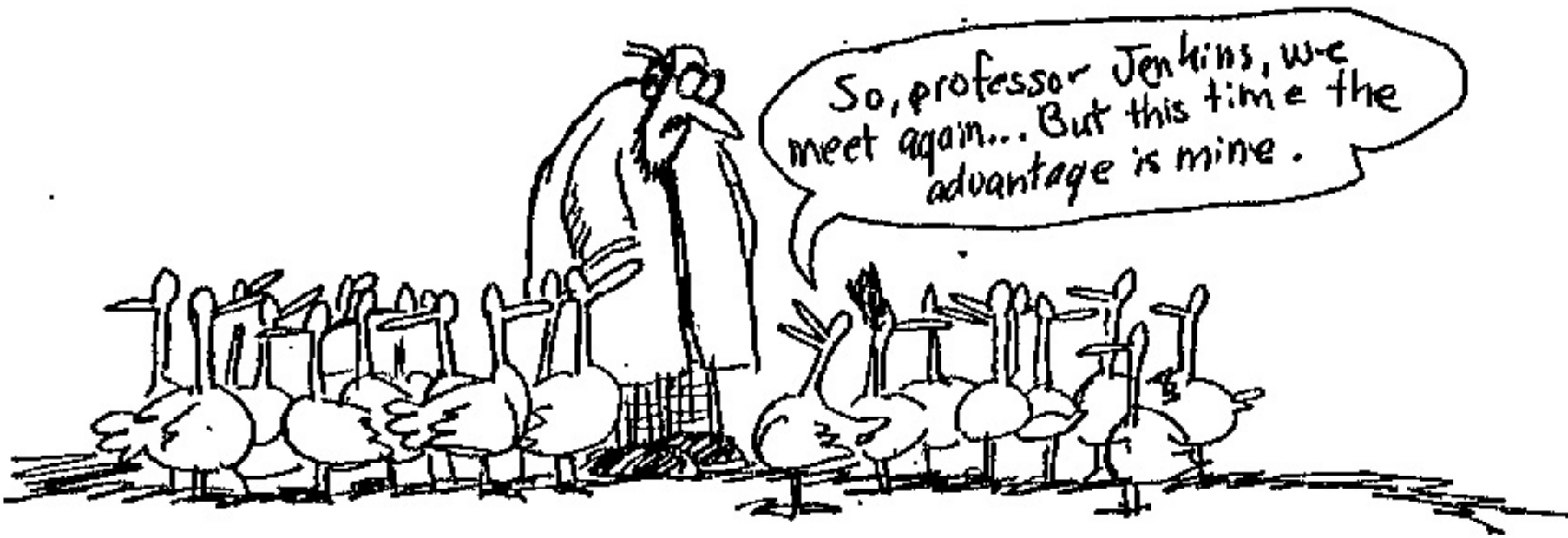
MASTER PAGE PROOF
res
editor
copy
photo

,

Edit! Edit! Edit!

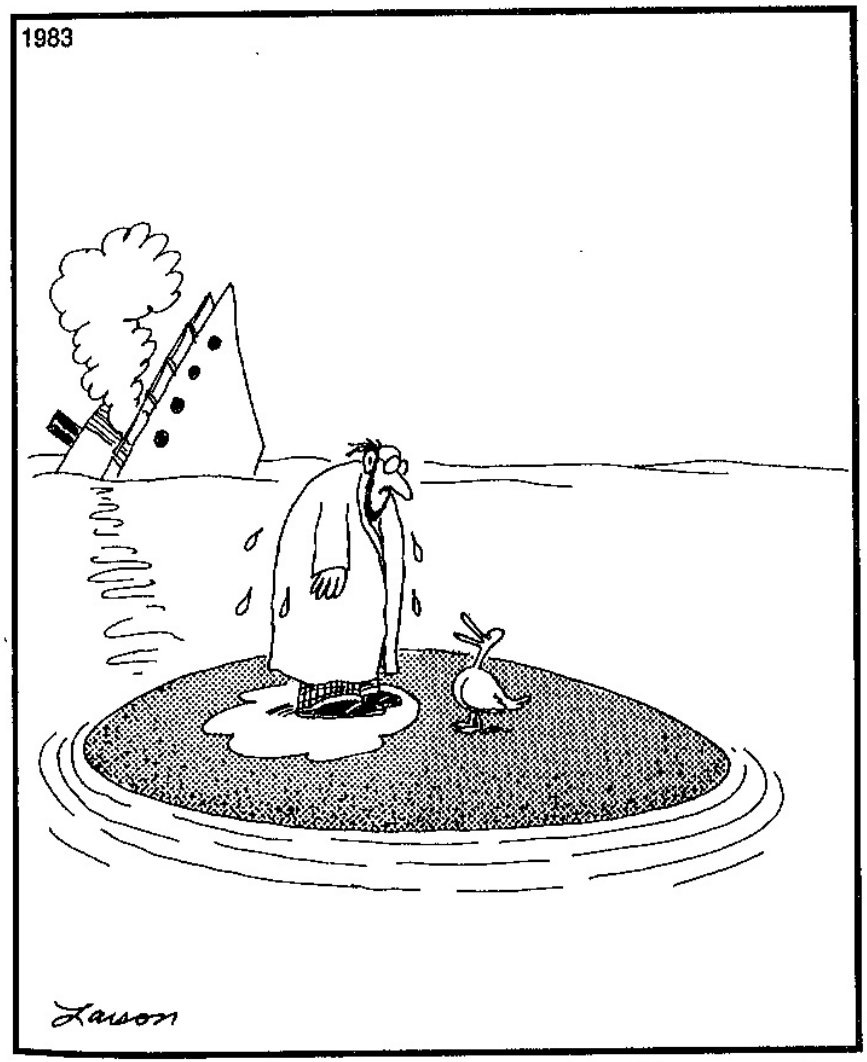


Edit! Edit! Edit!



Edit! Edit! Edit!

Edit! Edit! Edit!



"So, Professor Jenkins!...My old nemesis!...We meet again, but this time the advantage is mine! Ha! Ha! Ha!"

Edit! Edit! Edit!

Editing, revising, re-editing and re-writing are *essential*;

> *Every single word that I publish, I write at least six times*
>
> Paul Halmos
> Author, 'How To Write Mathematics'

- Be ruthless with yourself. If you think an explanation *sort-of* works, your reader will be totally lost
- Catching errors in your own writing is hard. Learn to read vigilantly, when required. Tick off sentences/paragraphs while re-drafting – reading sections aloud can also help
- Practise version control. Save files as `v1`, `v2`, or similar
- Get someone else to read through it – or part of it. Also, leave a segment for a while, then come back to it

Edit! Edit! Edit!

Editing and revising takes a long time;

- Very few students (and faculty) over-edit. The returns diminish eventually, but you'll probably hit a deadline before that happens
- Insufficient editing is obvious – and unimpressive
- Getting *something* (anything!) down and *then* editing is fine – many people write this way. But do plan time for editing, i.e. don't start the night before
- Writing/Editing often takes much *less* time than coding, or derivation, or just getting hold of data. Viewed this way, a few extra hours work may not seem so bad – and compared to the effort of re-submitting, it's nothing

Avoiding Nominalizations (. . . Well, Usually)

A lesson from Williams and Colomb (Chapter 3 – 'Actions');

Which of these two sentences do you prefer?

- The promise of our new approach is reduction of bias.
- Our new approach promises to reduce bias.

Avoiding Nominalizations (. . . Well, Usually)

Most of you (hopefully!) preferred the second, but why?

According to Williams and Colomb (W&C), a sentence is more appealing to readers when

- its main character is a subject of a verb and
- the verb expresses a specific action.

Let's boldface the main characters and italicize the verbs:

- The promise of our new **approach** *is* reduction of bias.
- Our new **approach** *promises* to reduce bias.

Avoiding Nominalizations (. . . Well, Usually)

Verb in first sentence ('is') is vague; verb in second sentence ('promises') is more substantial (has a punch to it).

W&C state two principles of clarity:

1. Make main characters subjects.
2. Make important actions verbs.

Nominalization is the technical term for a noun that is derived from a verb or adjective.

Nominalization is bad, because it strips verbs of their power and masks main characters.

Avoiding Nominalizations (. . . Well, Usually)

Return to our example:

- The **promise** of our new approach is reduction of bias.
- Our new approach **promises** to reduce bias.

In the first sentence we nominalized a verb ('promise') into a noun ('promise' also) and allowed it to mask the main character ('approach'). The second sentence avoids a nominalization.

Note: here the verb and its nominalization are the same, but this is not often the case ('discuss', 'discussion'; 'analyze', 'analysis'; 'justify', 'justification').

Avoiding Nominalizations (. . . Well, Usually)

Quote from W&C:

> *No element of style more characterizes turgid writing, writing that feels abstract, indirect, and difficult, than lots of nominalizations, especially as the subjects of verbs.*

Lesson 3 of W&C talks about

1. locating nominalizations in a sentence you have written;
2. picking out main characters and the actions they perform;
3. rewriting your sentences to obey the two principles of clarity.

Avoiding Nominalizations (. . . Well, Usually)

For next Monday (April 25th), do even-numbered parts of W&C's Exercise 3.6 (pp. 40–1), which will give you practice in revising dicey sentences so that they sparkle.

Also, read over pp. 42–4, which discuss reasons for *sticking* with certain nominalizations (hence the '. . . Well, Usually' qualifier on the slide title).

Note: arguably there is a small goof-up in the organization of W&C's material on pp. 42–4. Can you spot it? And should this prompt an eleventh edition?

Creating Characters

Another lesson from Williams and Colomb (Lesson 4);

Make the subjects of most of your verbs
the main characters in your story

Subjects? A reminder*...

	Subject	Verb	Object
	The cat	sat on	the mat
	The mat	sat under	the cat
The entity that	Sat on the mat	was	the cat

For our purposes, subjects are 'what the sentence is about'; the subject performs the verb's action.

* Most languages start sentences with subjects, but not all, e.g. Gaelic and Hawaiian are VSO. English's SVO is less common than SOV, seen in e.g. Japanese.

Creating Characters

W&C's principle in a book that's widely-read (from)

Mr. Tickle and the dragon stood beside Little Miss Splendid's swimming pool.

"It is too cold today to swim in Little Miss Splendid's pool," said Mr. Tickle. "Do you think you could do anything about that?"

The dragon thought for a moment.

Then he took a deep breath and breathed out through his nostrils. Flames licked across the surface of the swimming pool. In no time at all the pool was steaming.

Obviously, the goal here is to make the story easy to follow – but *your* story should be easy to follow, too.

Creating Characters

These would confuse the audience – badly;*

- The swimming pool was what Mr Tickle and the dragon stood beside
- A moment was the duration of the dragon's thought
- After a deep breath was taken by the dragon, it was through his nostrils that he breathed out.
- The surface of the swimming pool was licked by flames

As well as being complicated, they feel flat, indirect and impersonal

* Happily, reported speech is rarely a concern when writing about methods

Creating Characters

A 'proper' example; imagine we estimated coverage of three methods, using 1000 simulations. We could write this;

The estimated coverage of methods A and B is similar, but method C's estimated coverage is smaller. Monte Carlo error was small.

But the plausible 'characters' here are the methods – not the coverage values. So something like this is better;

Methods A and B gave nominal coverage, up to Monte Carlo error, but Method C never gave coverage above 90%

With methods as characters, it's straightforward to compare them, and giving their absolute properties is also easy. What you compare them *to* also matters – 2nd version seems better.

Creating Characters

Main characters needn't *always* be subjects.

The subject may be implicit, or irrelevant

- The bootstrap was developed just as desktop computers became affordable
- Brad Efron developed the bootstrap just as desktop computers became affordable

It's fine to write that 'the likelihood was maximized' *if* you want to focus on the likelihood. But *is* that the character in *your* story? Consider whether the character could be the coverage, estimate, parameter – or something else

Creating Characters

Main characters needn't *always* be subjects.

The subject is a convenient nominalization

Say you're describing how Hodges showed that, at the null value θ_0, his estimator is super-efficient, i.e. has smaller variance than the MLE. What comes next?

- Super-efficiency has also been seen in some Bayesian estimators
- We can also see super-efficiency in some Bayesian estimators
- Some Bayesian estimators also exhibit super-efficiency

In statistics we *often* discuss abstract ideas. These abstractions may not 'do' any verb – so the passive voice can be useful

Creating Characters

Is this advice familiar to you?

Avoid writing in the passive voice.
Use the active voice

- This is *not* terrible advice; it's easy to over-use the passive. Also, active forms often use fewer words
- But, as we've seen, it's *more* important to ensure that subjects are relevant to your story – and passive voice can help achieve this
- *Never* using the passive is just 'reactionary balderdash' – Microsoft's green underlining is only for advice

Creating Characters

Let's recap the advice of W&C;

Make the subjects of most of your verbs
the main characters in your story

Here's a blunter version, with broader consequences;

Get to the point, **quickly**.
Stick to the point

- Begin each paragraph by indicating its topic
- Open sections and chapters by declaring their topic(s)
- If/when the topic changes, make this obvious

Writing with no obvious point (e.g. listing properties of a particular estimator, without context) will confuse and annoy the reader.

Abbreviated noun use criticism interlude

W&C (pg 62) advise not using long strings of (modified) nouns;

Dugong Age Length Discrepancy Observation
Adjusted Maximum Profile Likelihood Estimator
Minimum p-value Permutation Test Approach
Robustified Double Hierarchical Generalized Linear Model

Standard terms should be retained – your reader knows them – but try not to make up new jargon.

Also, while it's impractical to replace 'sea snake diet data' with 'data on the diet of the snake of the sea' (Baer and Hildebrand, Science, 1983) a few hyphens can help; do you find 'sea-snake diet data' easier to read?

Stress in sentences

A lesson from Gopen and Swan;

Place each unit of information where readers expect to see it

'Units of information' can be words, phrases, sentences... all the way up to paragraphs and sections.

- Paragraphs and (particularly) sections should be sorted out in the planning stages
- Getting a sentence down in a *helpful* order takes care and forethought – or lots of re-drafting
- As a principle: Start with the topic; then say anything else you need; finish with what you want to stress

Stress in sentences

Imagine writing a review of some methodology; what material might sensibly **precede** each the following?

- Bootstrapping can be used to estimate standard errors
- Standard errors can be estimated by bootstrapping

At face value, these sentences say the same thing. But we've seen that they are *used* differently.

More spot the difference;

- PhD completion rates are improving in statistics
- PhD completion rates in statistics are improving

– when would you say **after** each one?

Stress in sentences

The final 'stressed' content indicates where you will go next; what's going on in each of these?

- Ken laughed when Don suggested co-teaching
- Ken laughed when co-teaching was suggested by Don
- When co-teaching was suggested by Don, Ken laughed

In each case, what might you say next?

Stress in sentences

The final 'stressed' content indicates where you will go next; what's going on in each of these?

- Ken laughed when Don suggested co-teaching
  ...he'd expected a very different suggestion
- Ken laughed when co-teaching was suggested by Don
  ...Don was the last person he wanted to work with
- When co-teaching was suggested by Don, Ken laughed
  ...and after laughing, he wept tears of bitter regret

In each case, what might you say next?

Stress in sentences

A useful tip from W&C (pg 89)

Read your sentences aloud.
As you speak the last few words, raise your voice and tap the table with your fingers. If you're stressed the wrong words, your voice and table thumping will feel wrong

- ... your voice and table thumping will feel wrong if you've **stressed the WRONG WORDS** ...?
- Topic, then stress – then topic, then stress
- Familiar information, then the new stuff – again and again

Stress in sentences

Like making characters subjects, this 'dovetailing' idea is a guideline, not an absolute rule.

The end of a sentences is *also* a good place to introduce long and complex ideas and phrases (W&C pg 85). Back to our introduction of Hodges' estimator;

- Hodges' estimator, at some point in the parameter space, attains a smaller asymptotic variance than regular efficient estimators
- Hodges' estimator attains a smaller asymptotic variance than regular efficient estimators at some point in the parameter space

Which do you prefer? Why? What helps you choose?

Stress in sentences

More from Gopen and Swan (paraphrased!)

If words or phrases relate to each other, keep them nearby

- Place modifiers near whatever is being modified
- Formally, place subjects and their verbs near each other
- Also; verbs should be indicative to the reader that some action is taking place (as we saw already)

Stress in sentences

An example (can you improve it further?)

- **Propensity score approaches** for bias reduction in treatment comparisons in non-randomized studies **led** to another round of debates about the merits of correlational analyses to make inferences about causality.
- Propensity score approaches can reduce bias when comparing treatments in non-randomized studies. Following their introduction, researchers again debated the merits of using correlational analyses to infer causality.

Unlike e.g. writing a novel, short 'choppy' sentences can be **just fine** in technical writing. Shorter sentences *connected logically* tend to be less ambiguous than long 'twisty' constructions – substance comes before style.

Writing Technical Material: basics

This is from pg 379 (!) of *Principia Mathematica*, by Alfred Whitehead and Bertand Russell

$*54{\cdot}43.\quad \vdash :. \alpha, \beta \in 1 . \supset : \alpha \cap \beta = \Lambda . \equiv . \alpha \cup \beta \in 2$

Dem.

$\vdash . *54{\cdot}26 . \supset \vdash :. \alpha = \iota`x . \beta = \iota`y . \supset : \alpha \cup \beta \in 2 . \equiv . x \neq y .$

$[*51{\cdot}231] \qquad \equiv . \iota`x \cap \iota`y = \Lambda .$

$[*13{\cdot}12] \qquad \equiv . \alpha \cap \beta = \Lambda \qquad (1)$

$\vdash . (1) . *11{\cdot}11{\cdot}35 . \supset$

$\vdash :. (\exists x, y) . \alpha = \iota`x . \beta = \iota`y . \supset : \alpha \cup \beta \in 2 . \equiv . \alpha \cap \beta = \Lambda \qquad (2)$

$\vdash . (2) . *11{\cdot}54 . *52{\cdot}1 . \supset \vdash . \text{Prop}$

From this proposition it will follow, when arithmetical addition has been defined, that $1 + 1 = 2$.

- As an exercise in pure logic, it's awesome – and flawed
- As a helpful explanation, it is a total disaster
- *You* are writing about statistics, not pure math

Writing Technical Material: basics

Even in heavily mathematical material (e.g. proofs), everything should be written in complete, punctuated sentences – because this helps the reader understand them.

For example, when declaring some new notation;

Bad: Consider $\hat{\theta}_n$, $n > p$.
Good: Consider $\hat{\theta}_n$, where $n > p$.
Better: Consider the estimator $\hat{\theta}_n$, where $n > p$.

- Equations and other 'displaymath' are just part of the text
- Read the material aloud. Did you include enough signposting?
- Top Tip: do this in your theory exam!

Writing Technical Material: basics

Write precisely; what's the difference in meaning here?

- $U^T DV$ is the singular value decomposition of X
- $U^T DV$ is a singular value decomposition of X

When declaring notation, it is *particularly* important say whether e.g. R is *a* representation or *the* representation

Don't start sentences with formulae;

Bad: $f(\lambda)$ has p roots, so ...
Good: The polynomial $f(\lambda)$ has p roots, so ...

– and try to connect formulae with words, not $\Rightarrow, \therefore$ etc

Writing Technical Material: personal pronouns

Technical writing should not sound like a report of your grade school vacation;

I maximized the likelihood. I inverted its Hessian.
Finally, I made a lovely confidence interval.

- *We* (i.e. 'you the reader and me the author, together') can now see that Theorem X is true
- A little bit of passive voice is not the end of the world
- Try to avoid 'I' unless persona is relevant
- When that's not possible, *consider* adding fictitious co-authors... or using $\mathcal{I}$ as a parameter

Writing Technical Material: which tense?

If the default tense isn't obvious, write in the present; e.g. 'this algorithm is an efficient way to compute X'.

- Sections on literature review and data collection can be written in the past tense
- If someone proved it, it *is* currently true $\Rightarrow$ present tense
- For your reader, 'Table 1 displays' and 'Figure 2 illustrates' $\Rightarrow$ present tense
- 'In Section 3, we apply this method to Problem Z' – present tense, because readers jump around
- Conclusions can be written in the past tense; 'We showed that X was a bad idea' – but perhaps 'We conclude that X **is** a bad idea'.

Writing Technical Material: UW-ese

Try describing the UW campus to a Hawaiian...

- 'Precision variable' is not a universal term (or concept)
- 'Robust standard errors' also go by other names
- Try to reach a wide audience; citing a textbook may help

Writing Technical Material: lists

Listing items causes these and other headaches?

- Get parallelism right; likelihood methods are used for estimation, tests and to predict?
- Modified objects must all match the verb; Smith *et al* calculate Bayesian estimates, intervals, and philosophy?
- Indicate that you are starting a list; clearly demark entries on that list. (Relax this a bit for just two items)
- Writing about factorial studies can be hard, for anyone
- Few papers use bulleted lists
- Don't interrupt a 'First ... second ...' sequence with other uses of 'first', 'second'

Writing Technical Material: other points

- Principle Components Analysis or Principal Components Analysis? Spellcheckers do not speak statistician
- Use minimal Latin, and use Latin correctly;
  - i.e. meaning 'that is' or 'in other words'
  - e.g. meaning 'for example'
  - etc meaning 'and so on', i.e. extending a list
  - Smith *et al* for Smith and ≥ 3 authors
  - *vice versa*... use it sparingly
  - *a priori*... also for occasional use

  ... but stop using 'ab hinc', *ab hinc*. And *ad nauseum*.
- Use notation that's standard *for your audience*, e.g. i for rows, j for columns, Y for outcomes...
- Use of ξ and ϑ invites your reader to #@&% off

Writing Technical Material: other points

- Be **very** precise with use of θ (parameter) and $\hat{\theta}$ (estimate). Writing that 'regression to the mean lead to inflated θ', you'll be assumed to be braindead*
- Use 'model' to refer to an assumed structure. Linear regression is not itself a model; linear regression may or may not use an assumed mean model
- Some technical terms get used ambiguously, e.g. 'semiparametric' and 'Bayesian'. If using them, ensure *your* interpretation is clear to everyone
- 'This may suggest X is Y' is waffle; 'This suggests X *may be* Y' is weak, but often appropriate

* Yes, this may itself be braindead. But it happens

Writing Technical Material: other points

Technical writing is formal – more formal than talks – hence;

- No contractions; isn't, doesn't, should've etc
- No jokes – and no exclamation points; they're "like laughing at your own jokes"
- No glib comments – you will sound vapid, i.e. dumb
- No slang – 'these authors' not 'these guys'
- Minimal acronyms; MLE and OLS are okay, UMVUE is not. Acronyms are less well-known than acronym-users would like – many TLAs and eTLAs are unhelpful jargon
- No emotive language; the difference between the methods was 'large', not 'gigantic'. Also avoid sounding snotty; don't refer to 'so-called experts' or 'supposedly-objective methods'

Writing Technical Material: other points

Beware: Statistical pioneers stole a lot of otherwise-useful words, to make their methods sound impressive;

Significant, Bias, Confidence, Consistent, Sufficient, Coherent, Identifiable, Estimable, Likely, Normal, Robust ($\times 3$), *Resistant, Profile, Credible, Error, Residual, Oracle, Optimal, Marginal, Reference, Empirical*

- If you use the 'lay' meaning of any of these, it may be misinterpreted by a statistically-aware reader
- Avoid ambiguous use of terms pertinent to your topic
- Using terms from other areas (e.g. 'coherent' in a non-Bayesian discussion) may be okay – but try to consider your reader's background

Writing Technical Material: other points

Sundry other hints and tips;

- To help skim-readers, give all main numerical results in either Tables or Figures *but not both*. Incidental numeric results, e.g. the value of your BGR diagnostic, can appear just in the text.
- Never round a p-value down to zero
- Beware font trouble with l, I and 1 (...`l, I and 1`)
- Criticism must be accurate; you must explain that e.g. 'this method does not allow for X, and when X holds Y may happen'. Don't just say 'this is rubbish because it ignores X'
- Do not rant. Lengthy criticism – that omits any balancing argument – is very unlikely to convince informed readers

Using BibTeX

BibTeX is TEX's citation manager. Given a file containing bibliographic information, it will;

- Insert citations like 'Rice (2010)' or '[1]' in your document, in place of `\cite{rice:2010}` in the source code
- Construct a bibliography at the end of the document, in whatever format you desire (e.g. Vancouver, Harvard)
- Automatically re-label everything if, during revision, you include new references, or re-order references, or correct typos in existing references

Once you have such a file, its contents can be used again in any other document; this is a **huge** time-saver. (I've been using and expanding `KenRefs.bib` since 1998)

Using BibTeX: file structure

Bibliography files (with `.bib` extensions) are text files with entries like the following;

```
@article{Cox:regr:1972, #this is the 'citation key'
    author = {Cox, D. R.},
    title = {Regression Models and Life-tables
(with Discussion)},
    year = {1972},
    journal = {Journal of the Royal Statistical
Society, Series B: Methodological},
    volume = {34},
    pages = {187--220}
}
```

- The citation key should be sane, short, and memorable
- Google Scholar gives these ready for cutting/pasting – set your 'Google Preferences'

Using BibTeX: file structure

Some standard LaTeX rules apply;

- White spaces don't matter
- Closing parentheses and commas do matter
- Typos results in irritating error messages

Some non-standard rules;

- Separate authors by `and`
- Some 'fields' are required, e.g. omitting an article's `title` will lead to an error message.
- Omitting e.g. publisher address should just give a warning
- Capitalize names yourself; protect lower-case with e.g. `{de}Villiers`

Using BibTeX: making citations

In your LaTeX source code, use `\cite{keyname}` wherever you want the citation to appear

- Most times, use ~`\cite{keyname}`, to enforce a space between the preceding character and e.g. '(Rice, 2010)'
- In the preamble, declare `\bibliographystyle{plain}` – or the appropriate style for your target journal
- At the end of the source code, `\bibliography{kenrefs}` to get the bibliography (note no `.bib` extension)
- See also ~`\citet{keyname}` and ~`\citeauthor{keyname}` in the `natbib` package

Using BibTeX: making citations

LaTeX makes a separate file of bibliographic information, which is updated as LaTeX 'compiles' your source file sequentially. Therefore, to inegrate your document and this information;

1. LaTeX your file – to set up required citations
2. Run BibTeX – this constructs a `.bbl` file for your document
3. LaTeX your file again **twice**

Realistically, the last changes you make to drafts are to e.g. typos, not references. If you run BibTeX after you've finished citation-related edits, you will run LaTeX enough times anyway.

A tip: you can edit the `.bbl` file, if needed

Using BibTeX: unusual citations

Not everything you'll cite is a journal article;

- See classes `@book`, `@inproceedings`, `@misc` etc
- Web pages can be cited – with date accessed – but it's standard to cite something on paper
- `R` gives you its BibTeX entry using `citation()`, or e.g. `toBibtex(citation("sandwich"))`
- When citing books for specific points, give the page(s) in the main text, e.g. Casella and Berger (2001) pp 166–168. When citing a chapter or other large chunk of a book, use the `@inbook` class.

Using BibTeX: what to cite?

For methods work, consider whether to cite;

- The paper that give the source of the general idea – e.g. Neyman & Scott (1946) showing by example that not all MLEs are consistent
- A paper, book or textbook that explains the idea, and how it applies to whole areas of research – e.g. Casella and Berger's material on consistency

Neither choice is 100% right. If you cite to indicate 'this statement is justified', the latter is most helpful. For scholarly literature reviews, try to cite something near(er) the source.

When citing, try to have the paper/book open in front of you

References/Recommended Reading

Some example BibTeX output;


D. R. Cox.
Regression models and life-tables (with discussion).
Journal of the Royal Statistical Society, Series B: Methodological, 34:187–220, 1972.


A. S. C. Ehrenberg.
Writing technical papers or reports.
The American Statistician, 36:326–329, 1982.


G. Gopen and J. Swan.
The science of scientific writing.
American Scientist, 78:550–558, 1990.

References/Recommended Reading


P. Halmos.
How to write mathematics.
L'Enseignment Mathematique, 16:123–152, 1970.


N. J. Higham.
Handbook of Writing for the Mathematical Sciences.
SIAM [Society for Industrial and Applied Mathematics], 1993.

The class site has more, with links. In addition to these references, a lot can be learned just from reading stylish writers – Andrew Gelman, Ronald Christensen, David Spiegelhalter

Planning the whole document

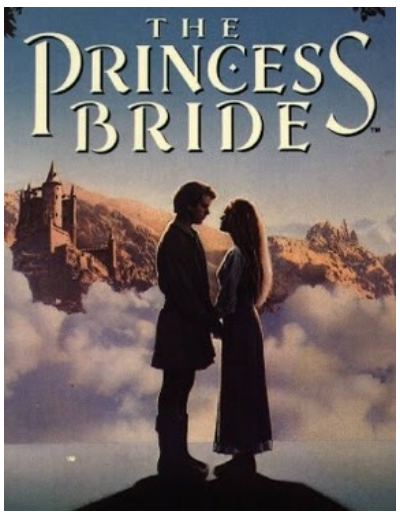
Like oral presentations, written documents should be planned. For example, in a 2-4 page literature review;

1. Introduction; statement of The Terrible Problem, and why it's of interest
2. a Overview of methods I (e.g. Empirical Bayes)
   b Overview of methods II (e.g. Likelihood approaches to measurement error)
   c etc
3. How these (could) go together to solve The Problem
4. Possible strengths and weaknesses of such an approach

Part 3 & 4 would be short, in a literature review

Planning the whole document

Sound familiar?



The over-arching structure helps the reader tolerate, appreciate and understand diversions along the way

Planning the whole document

Find a structure, even if you think the area is a mess;

- List the topics and decide the order in which they will appear
- Figure out how the topics connect – if they *don't* connect, you have a major problem
- Assess how much your audience knows about each topic – this can change dramatically between papers. Write for a specific audience
- Break the task into manageable pieces, and work on one at a time