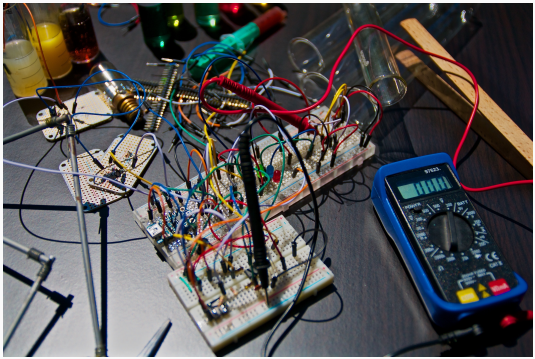
Cardinal sins in scientific publications

M. Bruker

“Paradigms for a more efficient publication workflow” series

February 2022



Picture by Nicolas Thomas on Unsplash

Jefferson Lab

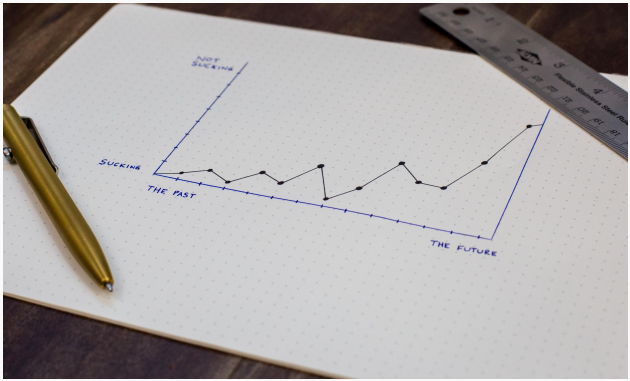
U.S. DEPARTMENT OF ENERGY | Office of Science

JSA

Why are we here today?

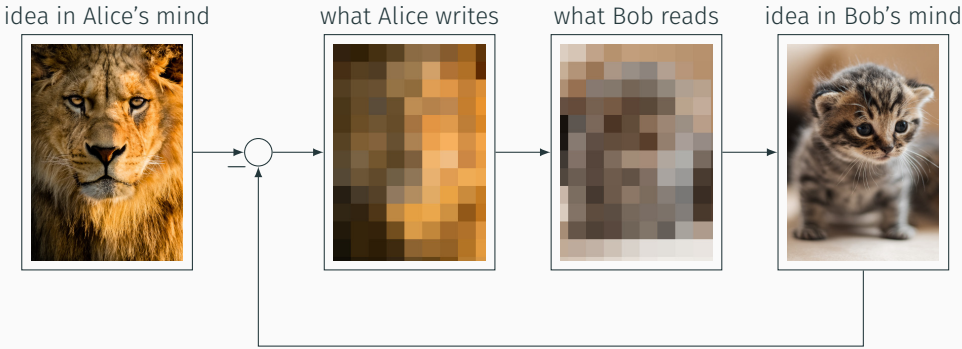
(...and also next week and maybe a bunch of times after that)

- I am not a communications expert
- But I have some thoughts on the matter
- Let's learn together!



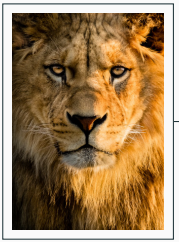
Picture by Isaac Smith on Unsplash

The purpose of scientific communication

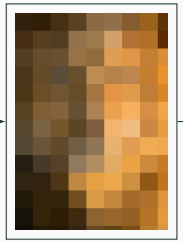


The purpose of scientific communication

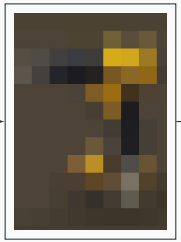
idea in Alice's mind



what Alice writes



what Bob reads



idea in Bob's mind

The true purpose of scientific communication

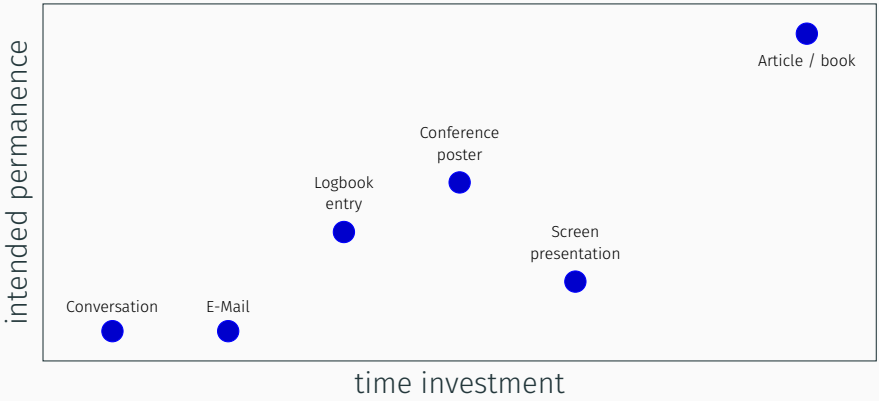
We communicate to...

- share our knowledge
- solve problems together
- make the world a better place

This is all fine, but at the end of the day...

- *We want to look good.*

Types of communication



Different media are different but similar

	Presentation	Stand-alone slideshow	Print article
Text	barely any	minor	verbose
Figures	chunky	chunky	detailed
Data readability	excellent	excellent	excellent
Train of thought	clear	clear	clear
Compromises	none	none	none

Errors vs. bad style

- The worse something is, the easier it usually is to avoid once you understand.
- What matters is how difficult/tedious it is to correct in editing.

Starting with actual errors: pet peeve #1: Math typeface

- Math typeface has a semantic meaning; there are accepted conventions that should be considered binding, e.g., ISO 31
- Quantity symbols are italic; nothing else is, especially not:
  - units (see next slide)
  - particles/elements: e^-, p, ν, μ, γ, π^0, He
  - operators: $d\sigma/d\omega$, δt, $\Delta\phi$
  - indexes that are not quantities: R_{BCS}, L_{eff}, k_B
  - functions: $\sin(x)$, $\exp(x)$, $\ln(x)$
  - special case: symbols of special mathematical constants: $\pi \approx 3.14$, $e = \exp(1)$ and $i^2 = -1$
- Failure to comply with these conventions is common and very dangerous

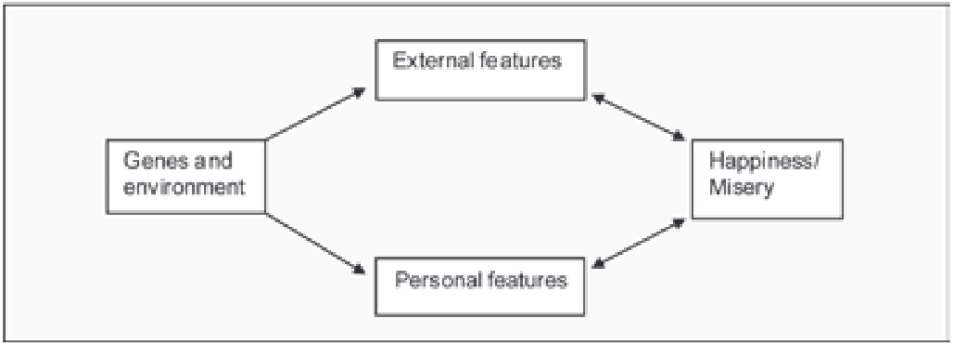
Picture by Nubelson Fernandes on Unsplash

Pet peeve #1.5: SI units

- Unit typesetting is strictly regulated in the SI Manual
- Units and prefixes are always typeset upright
- Space = mathematical product (number–unit or unit–unit, but not prefix–unit)
- No compound prefixes
- No made-up units
- In LaTeX, we use SIUNITX and are fine

wrong	correct
1m	1 m
1 *m*	1 m
1 mmmrad	1 mm mrad
1 mm-mrad	1 mm mrad
10^{34} cm^{-2}s^{-1}	10^{34} cm^{-2} s^{-1}
2e5	2×10^5
1 μs	1 µs
1 us	1 µs
1 Km	1 km
1 psec	1 ps
10 amps	10 A
1 pmm	1 fm
1 Vpp or 1 $V_{p\text{-}p}$	$V_{p\text{-}p} = 1$ V
1 MkW h yr^{-1}	🤯

Pet peeve #2: Rasterized figures (not strictly an error, but very close)



Columbia Climate School—The Earth Institute: World Happiness Report

`https://www.earth.columbia.edu/articles/view/2960`

Pet peeve #2 (continued)

How long would it take to typeset this table natively?
Would it be worth it?

Table 2: Weights on the 33 Indicators

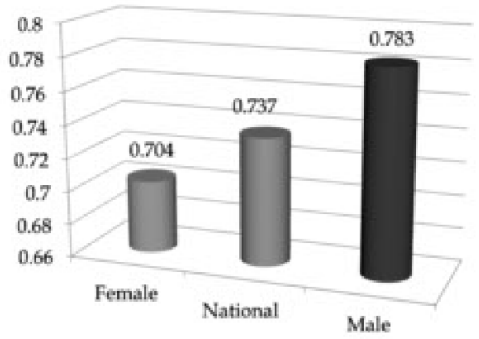
Domain	Indicators	Weight	Domain	Indicators	Weight
Psychological wellbeing	Life satisfaction	33%	Good Governance	Political participation	40%
	Positive emotions	17%		Services	40%
	Negative emotions	17%		Governance performance	10%
	Spirituality	33%		Fundamental rights	10%
Health	Self reported health	10%	Community vitality	Donation (time & money)	30%
	Healthy days	30%		Safety	30%
	Disability	30%		Community relationship	20%
	Mental health	30%		Family	20%
Time use	Work	50%	Ecological diversity	Wildlife damage	40%
	Sleep	50%	& resilience	Urban issues	40%
Education	Literacy	30%		Responsibility towards er	10%
	Schooling	30%		Ecological issues	10%
	Knowledge	20%	Living Standard	Per capita income	33%
	Value	20%		Assets	33%
Cultural diversity	Zorig chusum skills (	30%		Housing	33%
& resilience	Cultural participatio	30%			
	Speak native langua	20%			
	Driglam Namzha (Eti	20%			

Columbia Climate School—The Earth Institute: World Happiness Report

https://www.earth.columbia.edu/articles/view/2960

It gets worse

Figure 12: GNH index by gender



When we decompose the GNH Index by gender we see that men are happier than women. 49% of men are happy, while only one-third of women are happy, a result that is both striking and statistically significant.

Columbia Climate School—The Earth Institute: World Happiness Report

https://www.earth.columbia.edu/articles/view/2960

Pet peeve #3 (specific to screen presentations)

I'm going to tell you everything about this awesome thing I discovered:

- Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.
- Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.
- Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Slide "used as a drunk uses a lamp post—for support rather than illumination"

(source unclear: `https://quoteinvestigator.com/2014/01/15/stats-drunk/`)

What is “good enough”?

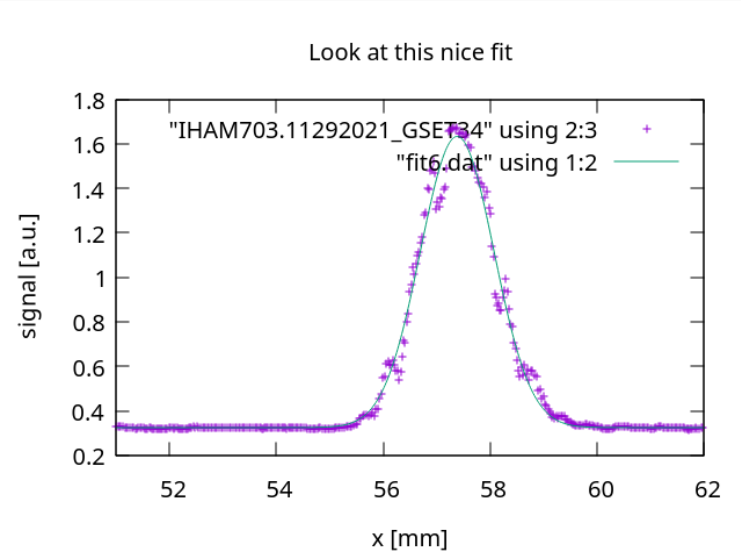
- Getting better results requires time:
  - learning better methods
  - implementing the solutions
- For time spent implementing, there are diminishing returns
- What is the FOM to optimize for? Depends...
- *Beware the lollipop of mediocrity; lick it once and you’ll suck forever.*
  —Brian Wilson
- IMHO: Time spent learning is well spent; however, limit time spent implementing so result is error-free and justifiable but any more polishing would feel excessively cumbersome
- Time spent making something not good enough is spent at least twice.

Platform-adaptable content

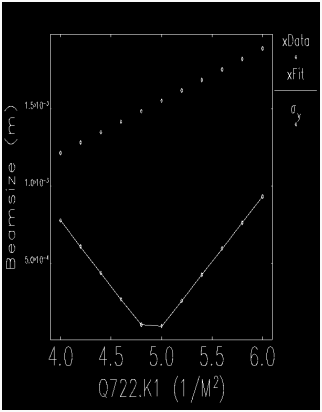
- Figure dimensions, fonts, etc. are fixed upon rendering, regardless of vector/raster property
- We still lack a good media-independent, seamlessly integrated format for data visualization (think BOKEH)
- Wait for paradigm changes & tech evolution… Until then, we *must* re-render all content from scratch for each document
- Use TikZ

Picture by Mick Haupt on Unsplash

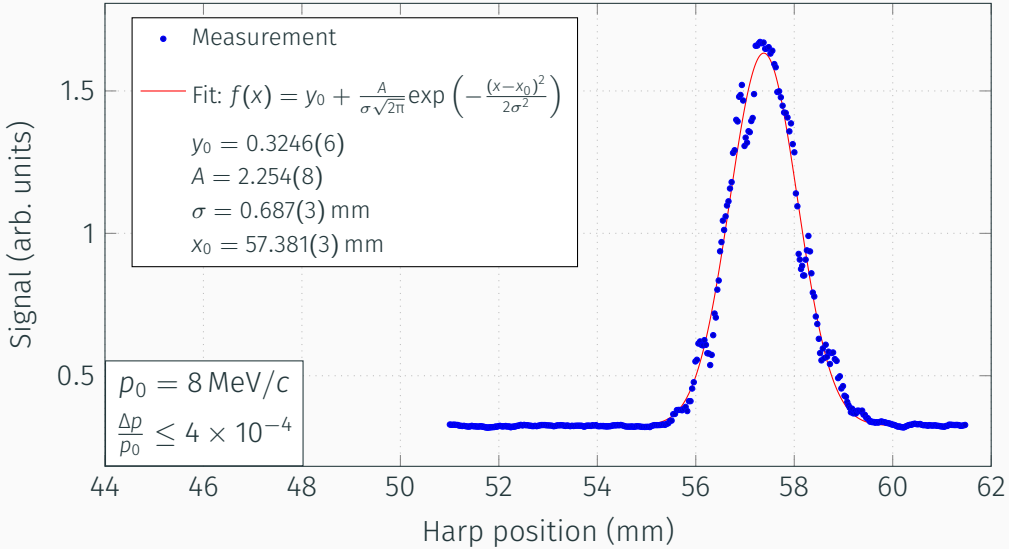
How not to make a plot



Speaking of aspect ratio distortion



Example of a decent plot



All WYSIWYG software is useless

- Stop playing with toys like WORD and POWERPOINT.
  They have been obsolete since 1983 / 1987.
- LaTeX is still the only typesetting system for us:
  - Creates presentations, posters, articles, and books from the same code
  - Separate content from style; abstract representation of information
  - Bibliography (use biblatex and biber)
  - Reusable, abstract figures with TikZ and pgfplots
  - Trivial handling of cross-references etc.
  - Built-in programming functionalities (TeX and now Lua)
- “Converting” between formats is a misconception

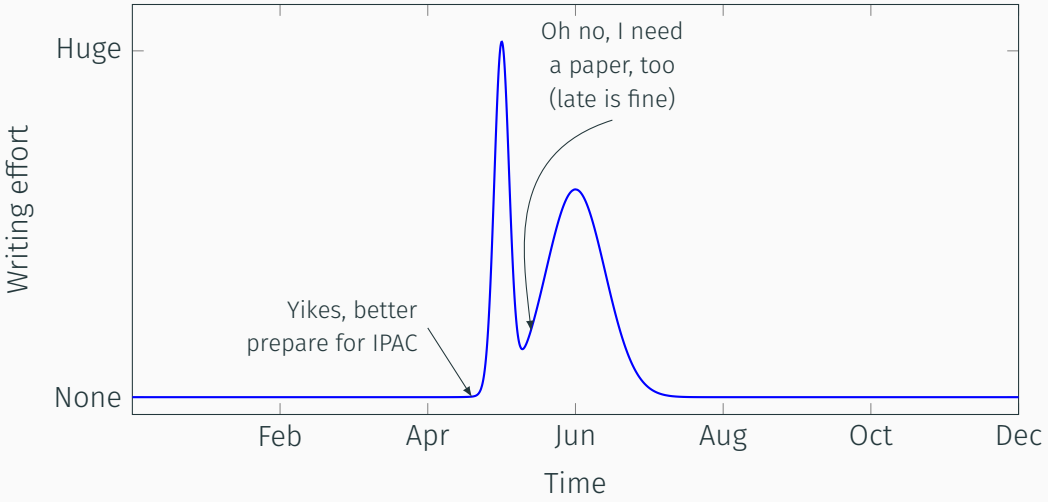
Most graphics/plotting software that isn't LaTeX is of limited use

- Visualizing data
  - GNUPLOT: popular for quick plotting, very difficult to get good results from
  - OCTAVE/MATLAB: superseded by PYTHON
  - MATPLOTLIB: see GNUPLOT
  - BOKEH: similar to MATPLOTLIB in terms of usage, better for web deployment
  - Internal graphics engines in simulation codes (SDDSPLOT et al.): usually garbage

  Export the numbers and use PGFPLOTS.
- Visualizing concepts
  - Vector software like INKSCAPE: can be fine, but difficult to use; issues with font inconsistency hard to avoid
  - GIMP, PHOTOSHOP: useless; results are of low quality and not easily adaptable/reusable

  Use TikZ.

Distribution of work load



Reviews and the publication signature queue

- Reviewing a finished piece of work under time pressure is not helpful
- Can we start talking about figures etc. early on?
- I don't need everyone to be a Ti*k*Z expert (I'm not one myself)...
- There should be a panel of enthusiasts to help with these things; encourage continuous improvement

Picture by jeshoots.com on Unsplash

Workflow implications

- Archive all raw data that accrue in an easily accessible location
- Make all plots and figures in Ti*k*Z the first time you need them
- Store the source code alongside the data files
- Copy code & data to any document they go into
- If you ever make temporary figures during analysis (in PYTHON etc.), keep the code

Time for a field trip :)

FAQ

- The program I'm using makes PNG files and can't export raw numbers.
- My supervisor is asking me for a plot to include in their PowerPoint presentation or Word document.
- Why is (*insert idea here*) so difficult to do in LaTeX?
- How does collaborative editing work if all you have is text files?
- You mentioned logbook entries and e-mail.
  How to properly visualize my ideas/data there?

The Jefferson Lab beamer template

- BEAMER: popular LaTeX document class to make slide shows
- Template made in 2019 based on METROPOLIS;
  resembles official JLab templates
- Code not very flexible/customizable, mostly for my own needs so far, but reasonable starting point
- Will hopefully be available on the website soon
- Take a look at the source of this document?