Writing (scientific papers) well

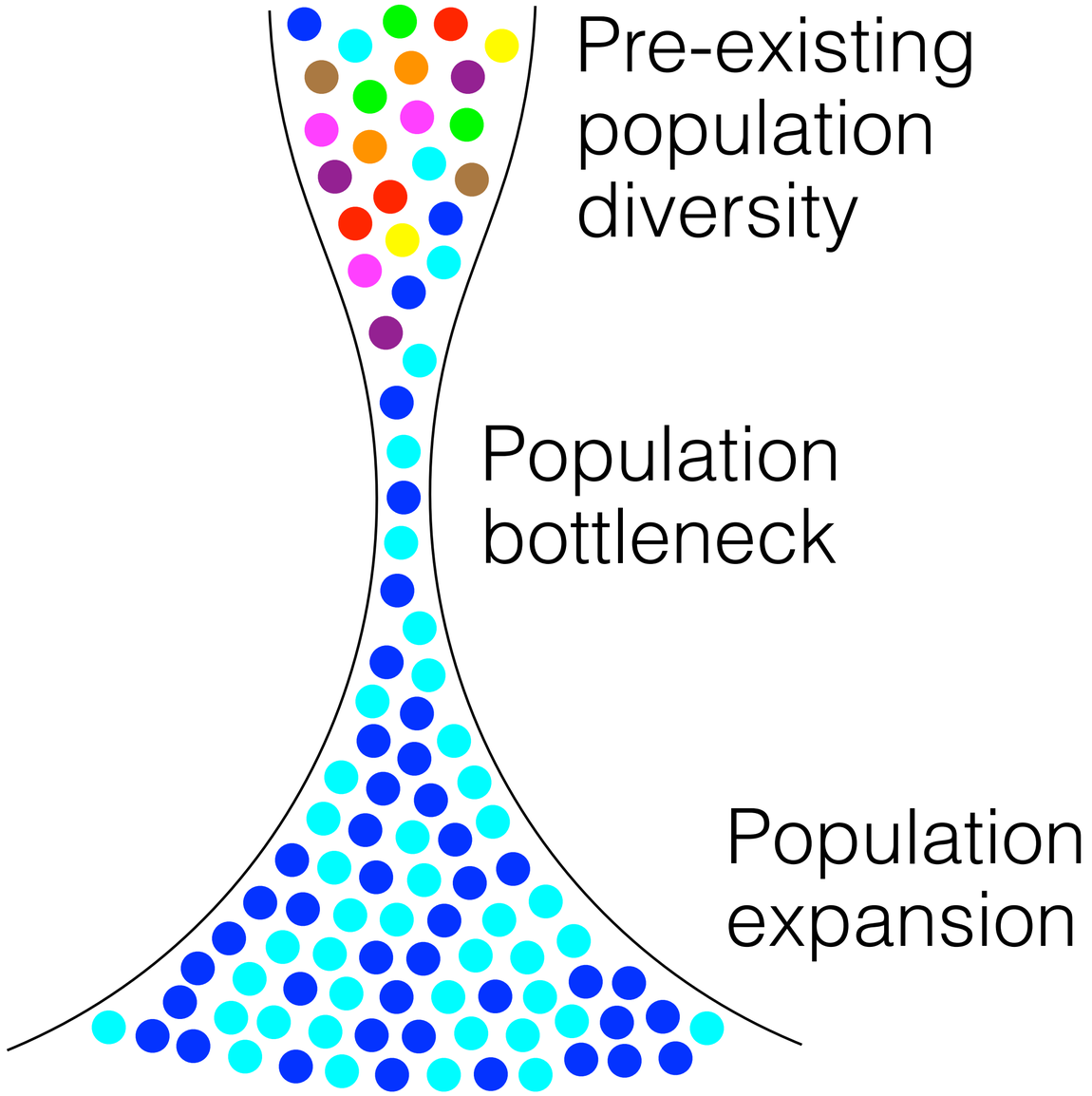
Chris Wymant

Slides from an University of Oxford Pandemic Sciences Institute internal seminar, May 29 2025. Small annotations added to some slides reflect things said but not written.

Overlapping material (on writing papers but not writing generally) in prose at

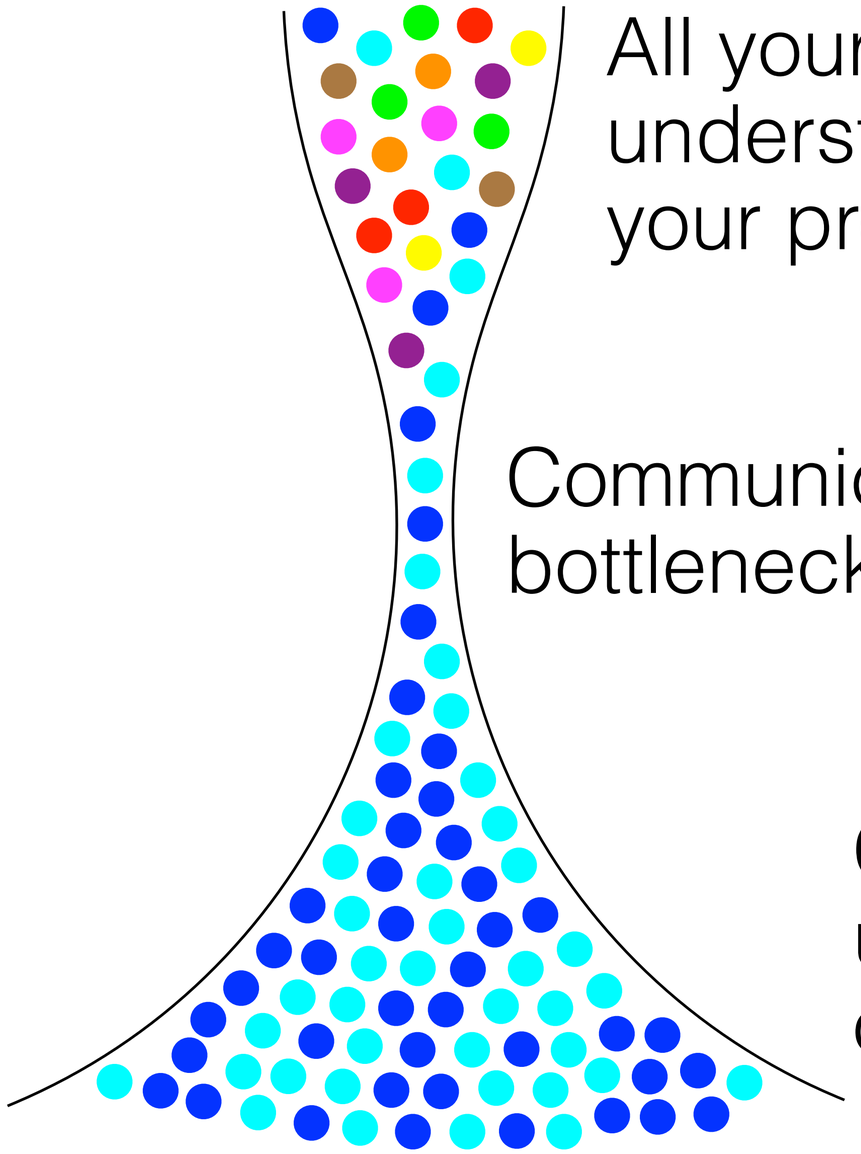
https://github.com/ChrisHIV/teaching/blob/main/other_topics/advice_for_writing_a_scientific_paper.MD

Motivational analogy



Motivational analogy

What you choose to share, and how well you do it, can have a large impact.



All your thoughts & understanding re: your project

Communication bottleneck

Community's understanding of your project

The top three recommendations from this talk:

Paper level: frame your paper up front as a clear, motivated question that you set out to answer.

Paragraph level: plan, organise and analyse your writing at the paragraph level.

Sentence level: identify the main actors and actions, and align these with the simple subjects and verbs.

Talk structure

Writing papers: main points

Writing generally: main points

Writing papers: minor points

Writing generally: minor points

Talk structure

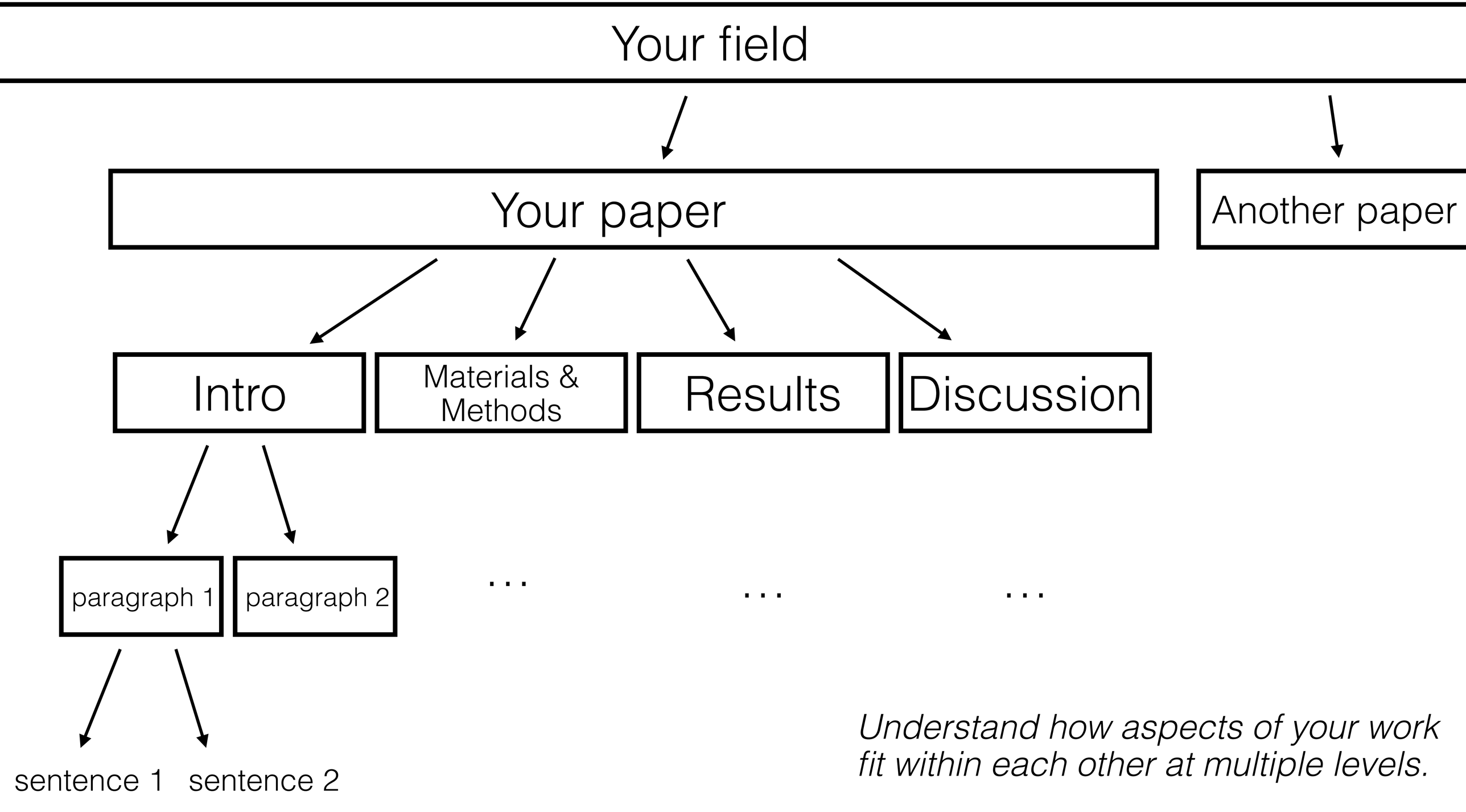
Writing papers: main points

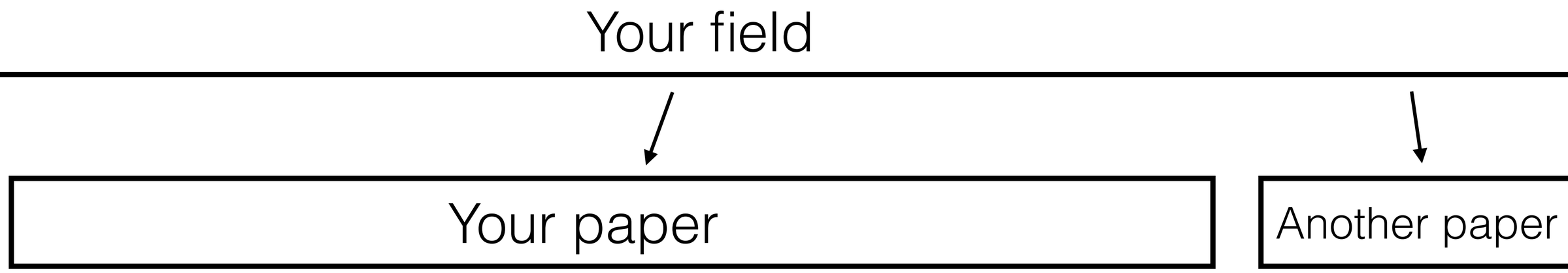
- **Multiple levels; the field level**
- The paper level
- The section level
- The paragraph level
- Other points

Writing generally: main points

Writing papers: minor points

Writing generally: minor points


Your field
Your paper
Another paper
Intro
Materials & Methods
Results
Discussion
paragraph 1
paragraph 2
sentence 1
sentence 2
Understand how aspects of your work fit within each other at multiple levels.



If you can explain how your paper fits into the field it belongs in, that will help the reader see the even bigger picture, and why your paper matters within that picture.
Your paper might be part of multiple fields at once, in which case you can comment on how it fits in each of them.
For example you might write an HIV molecular epidemiology paper, which is part of the HIV epidemiology field, and of the molecular epidemiology field.
And such a paper might provide a result that fills a gap in HIV epidemiology, and a method that's new to molecular epidemiology.

Talk structure

Writing papers: main points

- Multiple levels; the field level
- **The paper level**
- The section level
- The paragraph level
- Other points

Writing generally: main points

Writing papers: minor points

Writing generally: minor points

Your paper: a natural view as an author

You've done some work

You've found some results

You want to tell people about it

You need to publish (or perish)

The problem:

You're forgetting your readers

Imagine one responding *So what?*

Testing a hypothesis

If you

- do various things to your data,
- see what result comes out,
- speculate which hypothesis that result favours,

you're vulnerable to fitting the wrong story to the evidence.

Some common advice, for both conducting and reporting your work, is to start with a hypothesis then test it.

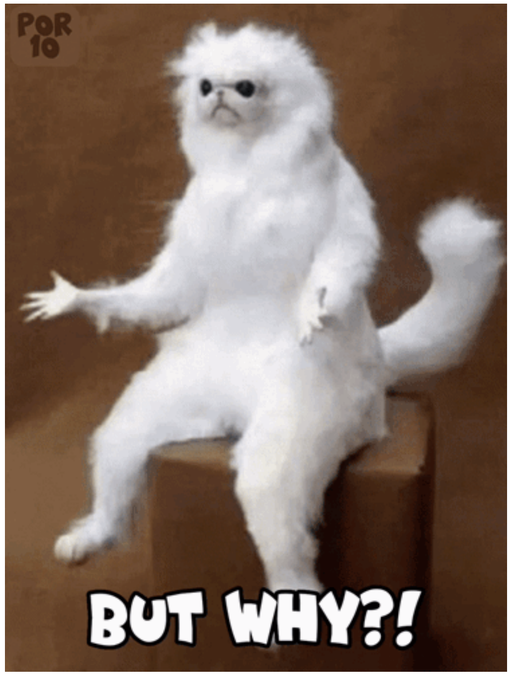
What's wrong with that?

e.g. Hernán & Greenland, JAMA 2024

Testing a hypothesis

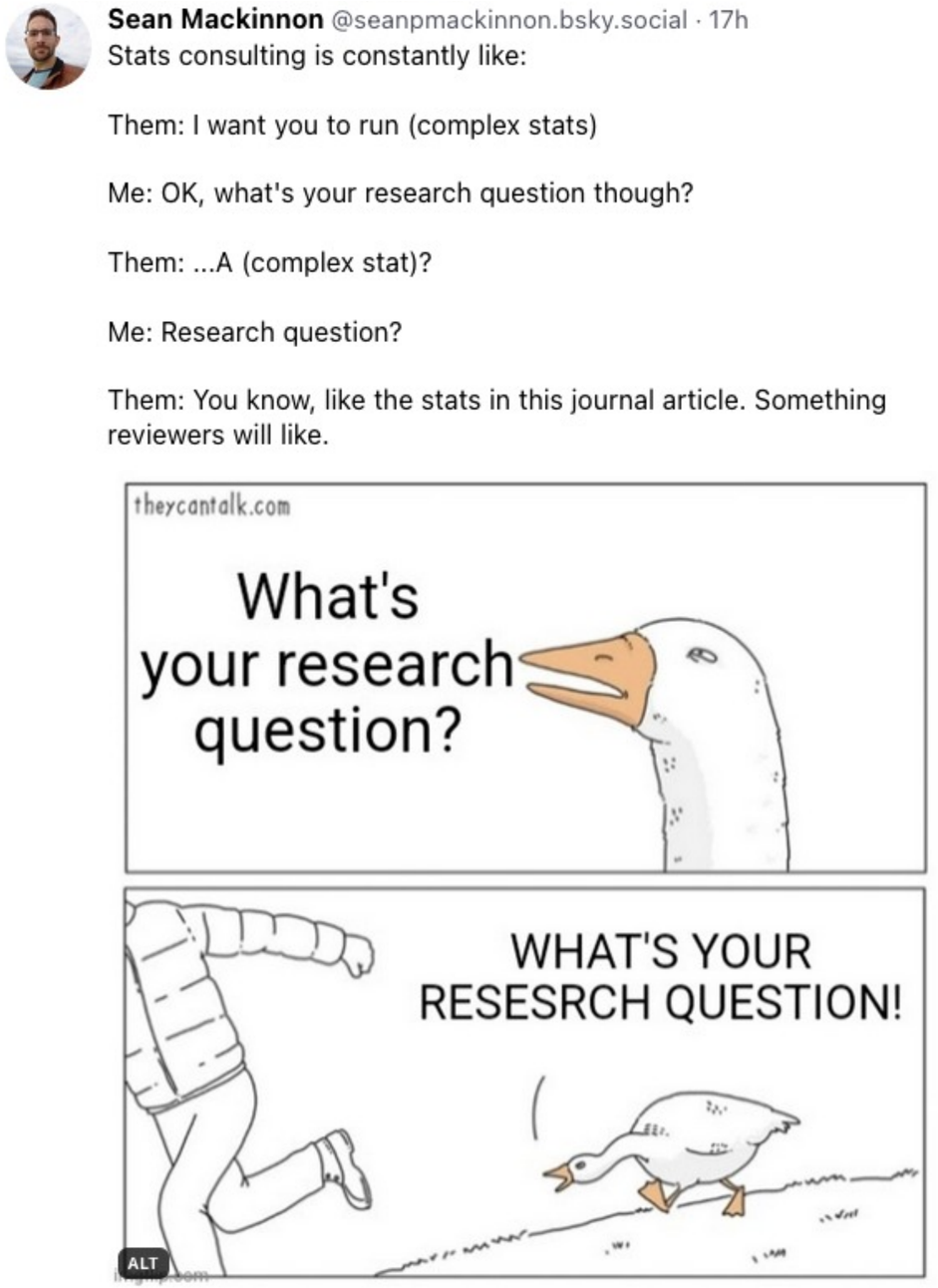
"Start with a hypothesis then test it"

- What about continuous hypothesis spaces?
- Need to choose our favourite answer to the question *before* considering new evidence



~~Start with a hypothesis then test it~~

Start with a question then answer it



Why pose questions?

What you did and what you found = taking the reader on journey.
It makes much more sense if they know where you are trying to get to!

You should motivate your question: address that '*so what?*' up front.
Some readers will not care what the answer to this question is. That's fine.
Those who do care become invested in the journey.

Why pose questions?

A problem well-put is half solved.
—John Dewey

Looking back, I think it was more difficult to see what the problems were than to solve them.
—Darwin

The formulation of a problem is often more essential than its solution, which may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old questions from a new angle, requires creative imagination and marks real advance in science.
—Einstein

This is slightly tangential to the writing process, but is important to our work generally, and relates to something I've talked about before: the fundamental importance of asking the right questions.
Here are three people talking about how important it is to have a good problem to work on - a problem that's well put, or well formulated.
Solving a problem is similar to answering a question.
Arriving at a good question is hard work and important for making progress.
So it makes sense that communicating to our audience clearly what question we're trying to answer helps them follow what we've done.

2024 Fraser-group examples

Bienfait et al:
How can we best predict susceptibility to neutralising antibodies using HIV sequences?

Rob & Jasmina et al:
For mpox in the UK in 2022, what were the *infection* dynamics (distinct from the *case* dynamics due to dynamic ascertainment delays)?

Michelle & Luca et al:
What additional epidemiological insight can app-based contact tracing provide in near-real time?

Luca & Chris et al:
How did the probability of SARS-CoV-2 transmission vary with app-recorded measurements, e.g. exposure duration?

Matthew et al:
For HIV in Zambia, how does HIV transmission vary with age, sex and drug resistance mutations?

Rewriting history

The question you use to frame your study need not always be what you were asking at the time. The unexpected occurs, and it may be useful to someone.

An exception to question-driven framing

Discovery of something previously not known to exist, whose existence is interesting.

e.g. new species. No natural Q&A:
“Is there an undiscovered species with this genotype/phenotype?” “Why yes!”

Shoehorning into this exception: “I discovered a correlation exists between variables X and Y in my data.” Not actually interesting. The broader question may be.

Talk structure

Writing papers: main points

- Multiple levels; the field level
- The paper level
- **The section level**
- The paragraph level
- Other points

Writing generally: main points

Writing papers: minor points

Writing generally: minor points

Introduction

A funnel shape: channel your readers' attention from the general to the specific.

General background

Specific background

Specific knowledge gap

"Here, we investigated this gap by…[brief summary of your strategy but not results]"

para 1: covid in Eng & Wales

para 2: contact tracing

para 3: app-based contact tracing: how it works

para 4: previous studies of app-based contact tracing. Math modelling, not data.

para 5: "Here we evaluated app-based contact tracing in Eng & Wales using data…"

Materials & Methods

Say how your inputs were generated & what you did with them.

Describe the process following the flow of information on a journey from its original setting to how it was obtained and processed from raw to simplified form. e.g.

1. the affected population (i.e. describe the epidemic)
2. the cohort (inclusion criteria)
3. how we obtained biological samples from each participant
4. what experiment we ran on each sample
5. how we simplified each experimental result from complex data into simple data
6. what inclusion criteria or imputation we applied to the simple data
7. what steps we ran on simple data for the main analysis

Results

What did you find: what were the results of what you did.

Save your main commenting on your results for Discussion.
Short comments on your results within Results may help
readers to keep following your story (thanks Katrina).

(If the journal puts Materials & Methods at the end, start Results with
the briefest summary of M&M possible for the results to make sense.)

Discussion

Possible topics:

- Summaries of what you did & found
- Your interpretation of the findings
  - How robustly can we generalise from this dataset or thought experiment to the world at large, and so what have we learned, what do the results mean?
- Limitations
  - Don't just state platitudes. *All* models are wrong. Which of your assumptions are most suspect?
- Implications / outlook: what next?
  - Should policy change in light of your results?
  - Should other researchers use your method?
  - Are more/bigger/different studies needed for this question?
  - Or fewer studies, if you have a convincing negative result?
  - Does your study suggest new questions?

Previous related work

Previous related work can fit in both Intro and Discussion.

Intro:
Say to what extent previous studies have answered this question.
If that's "reasonably well":

- summarise their answers
- explain why you asked the question again,
  e.g. to test using new kind of data, to resolve disagreement.

Else:

- explain your reasonable claim on novelty here

Discussion:
If previous studies did "reasonably well":

- compare results,
- speculate on the cause of differences,
- explain again the added value of your paper.

Else:

- state again your novelty.

Talk structure

Writing papers: main points

- Multiple levels; the field level
- The paper level
- The section level
- **The paragraph level**
- Other points

Writing generally: main points

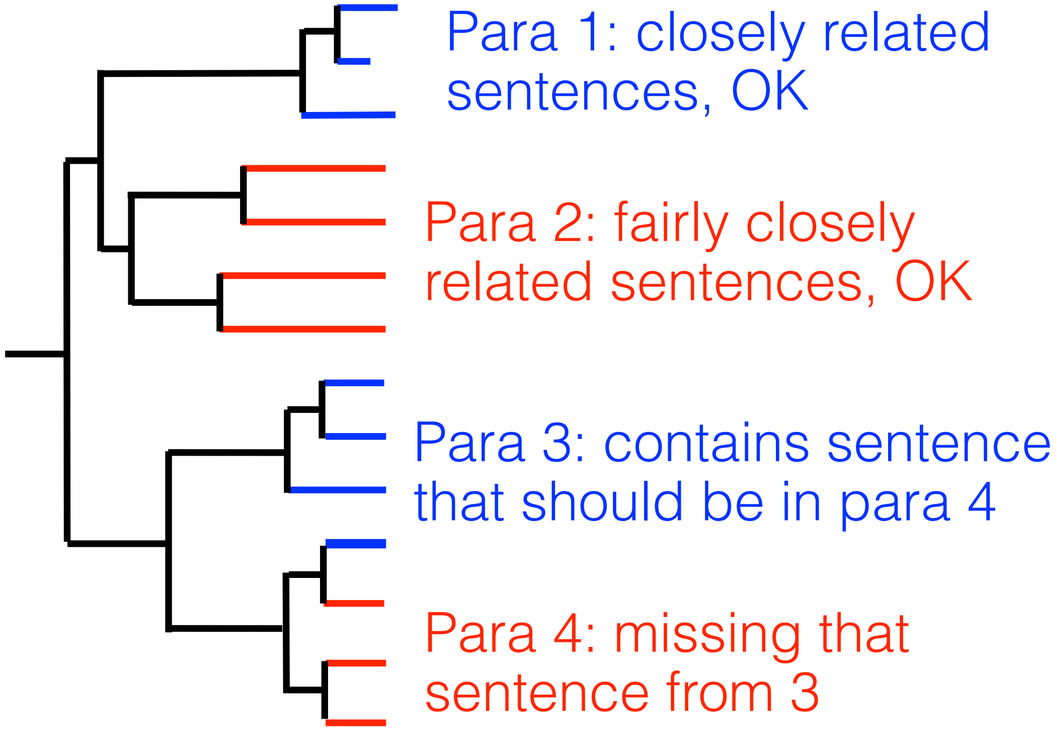
Writing papers: minor points

Writing generally: minor points

Paragraphs

Each paragraph should have a single topic, different from the topic of other paragraphs.

Grouping sentences into paragraphs = a clustering algorithm.
A sentence should be more closely related to sentences in the same paragraph than it is to sentences in other paragraphs.



It's very helpful to know what makes a mammal a mammal and a bird a bird if you want to put a species into one of those two groups. In the same way, if you want to put a sentence into the right paragraph, it's helpful to be able to see a paragraph as a thing in its own right, i.e. taking a higher level view than the sentence level, seeing a paragraph not just as: one sentence that says this then a second sentence that says that. Instead, seeing the single topic that all these sentences are part of. That might be: here is my paragraph saying what our results were, here is my paragraph about the implications of those results. Or here is my paragraph about our first result and what it implies; here is my paragraph about our second result and what it implies. There's different ways of organising your thoughts; the point is that they should always be organised.

Paragraphs

Many readers expect the first sentence in a paragraph to define the paragraph topic. Following this expectation helps.

Our dataset included data from a Ugandan cohort.
[Summary of Ugandan cohort.]
Our dataset also included data from a Botswanan cohort.
[Summary of Botswanan cohort.]
👎

Our dataset comprised data from cohorts in Uganda and Botswana.
[Summary of Ugandan cohort.]
[Summary of Botswanan cohort.]
👍

Planning paragraphs

A paragraph plan: a few words / short sentence saying what is the single topic of each paragraph.

Before you start writing: write a paragraph plan, and get feedback from coauthors at that stage: easier & quicker for you & them. Update the plan as you write.

If it's too late for that: write the paragraph plan describing what you've already written, to check structure and flow.

e.g. you might start with this and get coauthor feedback towards this

summary: what we did & found, interpretation

previous studies

outlook: more studies

limitation: sampling bias limits generalisability

limitation: showed correlation not causation

summary: what we did & found

why this matters: novelty & interpretation

previous studies

limitation: not strong evidence of causation

outlook: more studies including interventions

Talk structure

Writing papers: main points

- Multiple levels; the field level
- The paper level
- The section level
- The paragraph level
- **Other points**

Writing generally: main points

Writing papers: minor points

Writing generally: minor points

Writing to inform and persuade

Different kinds of writing has different aims.

Background, what we've done, what we've found: writing to *inform*

What we think our findings mean, and what that implies: writing to *persuade*.

Every point should be information or persuasion (or something else) *but not both*. Clearly indicate which, through style. Don't 'spin' facts.

	Objective writing style	**Subjective writing style**
Statement is truly objective	We analysed 2218 genomes ✅	We leveraged a dataset of more than 2000 genomes ❌
Statement is truly subjective	We measured the effect of X on Y ❌	We believe that the correlation we estimated between X and Y reflects a causal relationship ✅

Impress with content, not style

If impressing people is your aim, clearly communicate *some impressive science*. Choose the clearest word, not a more impressive or dramatic sounding alternative.

Use methods, analyse data, find findings 👍

Leverage (👎)

Harness (the power of) 👎

Unlock insight 👎

But don’t be timid either. One robustness qualifier is enough.

*It’s **possible** that these results **may suggest**…* 👎

*These results **suggest**…* 👍

Transferability to talks

Basically all of this so far applies to giving talks too. e.g.

- Use a funnel-shaped introduction from general to specific.
- End your introduction with a clear question: frame your work as answering this question.
- Provide motivation for that question.
- A paragraph in writing $\approx$ a slide in a talk.

Reading

Reading about writing is helpful: recommendations with explanations, plus many good-bad examples. See next section.

I think reading essays is helpful: writing to inform & persuade. I like(d)

Changing My Mind, by Zadie Smith

Feel Free, by Zadie Smith

Selected Essays, by George Orwell

Selected Literary Essays, by C. S. Lewis

The Oxford Book of Essays, OUP

Talk structure

Writing papers: main points

Writing generally: main points

Writing papers: minor points

Writing generally: minor points

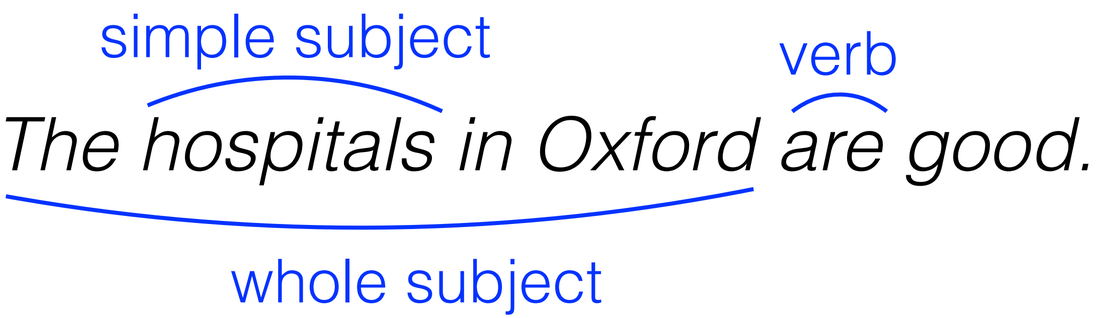
Background

The verb: the action (grammatically)

The subject: who (or what) is doing that action (grammatically)

- The whole subject: what it sounds like
- The simple subject: the smallest unit of the whole subject that determines whether the verb is conjugated in singular or plural form.

The object: who (or what) the action is happening to.



Storytelling

Which do you prefer? Why?

Once upon a time, Little Red Riding Hood was walking through the woods, when the Wolf jumped out from behind a tree and frightened her.

Once upon a time, as a walk through the woods was taking place on the part of Little Red Riding Hood, the Wolf's jump out from behind a tree occurred, causing her fright.

Style: Lessons in Clarity and Grace, by Williams & Bizup

Key point from this book:

Identify the main actors and actions, and align these with the simple subjects and verbs.

Style: Lessons in Clarity and Grace

Once upon a time, Little Red Riding Hood was walking through the woods, when the Wolf jumped out from behind a tree and frightened her.

Once upon a time, as a walk through the woods was taking place on the part of Little Red Riding Hood, the Wolf's jump out from behind a tree occurred, causing her fright.

Style: Lessons in Clarity and Grace

subject

Once upon a time, Little Red Riding Hood

verb was walking through the woods, when the Wolf *subject*

verb jumped out from behind a tree and frightened her.

verb

subject

Once upon a time, as a walk through the woods

verb was taking place on the part of Little Red Riding Hood, the Wolf's jump out from behind a tree *subject*

verb occurred, causing her fright.

verb

Style: Lessons in Clarity and Grace

Grammar level:	*Subject*	*Verb*	*Object*	
	We	discussed	the problem.	*Aligned* 👍
Story level:	*Main character*	*Main action*		

Grammar level:	*Subject*	*Verb*	*Object*	
	The problem	was	the topic of our discussion.	
Story level:			*Main character*	*Main action*

Unaligned 👎

Writing that does not align the grammar and story levels suffers more from big composite nouns linked with non-specific verbs.

subject

The main advantage our proposed study has for discovering new science, stimulating innovation and translating into public health benefits is…

verb

subject *verb + verb = verb* *verb* *verb*

We aim to discover new science, stimulate innovation and translate into public health benefits. Our proposal's main advantage is…

subject *verb*

Style: Lessons in Clarity and Grace

Put early in the sentence: things that are

- simple
- familiar
- expected

Put towards the end: things that are

- complex
- unfamiliar
- unexpected

HIV is a retrovirus. The error-prone reverse transcription of RNA into DNA causes retroviruses to have high mutation rates.
👎

HIV is a retrovirus. Retroviruses have high mutation rates because the reverse transcription of RNA into DNA is error-prone.
👍

Several short sentences about writing, Verlyn Klinkenborg

Write shorter sentences: they're harder to get wrong and easier to fix. Readers can get lost in long meandering sentences.

Write long sentences by linking short sentences with clear conjunctions like *and, if, because, however.*

"Is the subject of the sentence an actor capable of performing the action of the verb?
Can you adjust the sentence so it is?
Or does the subject of the sentence hide the action of entities that are able to act—humans, for instance?"

Politics and the English Language, George Orwell

"modern writing at its worst… consists in gumming together long strips of words which have already been set in order by someone else, and making the results presentable by sheer humbug. The attraction of this way of writing is that it is easy. It is easier - even quicker, once you have the habit - to say *In my opinion it is a not-unjustifiable assumption that* than to say *I think*."

Elements of Style, Strunk & White

Clear explanation of many basic rules of English.
Concise, short book. Recommended.

Active not passive voice

Active voice: we made mistakes.

Passive voice: mistakes were made (by us).

Style: Lessons in Clarity and Grace:
Prefer the active voice

Several short sentences about writing:
Prefer the active voice

Politics and the English Language:
Prefer the active voice

Elements of Style:
Prefer the active voice

A suggested exception

The passive voice is more concise and just as informative where the subject is clearly the authors of cited works.

This has been studied previously [1-4].

is slightly better than

Researchers have studied this previously [1-4].

is much better than

Alice et al., Bob et al., Charlie et al., Daniella et al. have studied this previously [1-4].

Talk structure

Writing papers: main points

Writing generally: main points

Writing papers: minor points

Writing generally: minor points

Naming things

Clearly establish and then stick to a one-to-one mapping:

[things] ↔[names for those things]

Any differences in naming a thing introduce doubt as to whether a different thing is being named.

e.g. tricky in our COVID-19 app papers.
Proportion TPAEN = proportion (of individuals) Testing Positive (and reporting) After an Exposure Notification
≈ secondary attack rate
≈ ‘infection probability’
With/without correction for background risk (not index)?

c.f. equations: don’t define a variable x and then refer to X

Be specific about your paper's value

Freakonomics: adverts for expensive houses tend to use specific words (titanium). Adverts for cheap houses tend to use vague words (lovely).

What suspicions do you have when reading a paper that claims it

- "sheds light on" the phenomenon under study,
- "characterises" the phenomenon under study,
- "has implications for" future studies or interventions,

without elaborating?
(Answer: maybe it does not do the thing claimed.)

Tenses

A suggestion:

- use the past tense for the past
- use the present tense for the present
- use the future tense for the future

This **is** the problem.
Previous studies **did** this and **found** this.
Here, we **did** this and **found** this.
This **suggests** that the answer to our question **is** this.
The outlook is that such interventions **will be** beneficial.

A subtlety for methods:

When describing a method designed as a one-off of these results, say what you **did**.

When describing a method designed for re-use by others, say what the method **does**.
(Readers read such methods to understand what the method **is doing** as they use it.)

Talk structure

Writing papers: main points

Writing generally: main points

Writing papers: minor points

Writing generally: minor points

Elements of Style

Rule 19: Express coordinate ideas in similar form.
When making a point with the same conceptual structure,
use the same sentence structure.

Authors A used method a and found result X [A].
Result Y was obtained [b] leveraging methodology b.
Employing c, C et al [C] demonstrated Z.
👎

A et al used method a and found X [A].
B et al used method b and found Y [B].
C et al used method c and found Z [C].
👍

Repeating the structure clarifies what's changing with lower mental effort.
c.f. rows of data in a rectangular table.

Style: Lessons in Clarity and Grace

When you refer back to something from a previous sentence using words like *this, these, that, those, another, such, secondly,* put this as close to the start of the current sentence as you can. This makes it easier for the reader to cast their mind back.

Avoid ‘throat clearing’ at the start of the sentence: get to the point.

It is important to note that… 👎

What we did is we [verb] 👎 *We [verb]* 👍

If the main topic of each sentence is indicated close to its start, then the different sentences in the same paragraph more clearly connect to each other.

Politics and the English Language

Here is a well-known verse from Ecclesiastes:

I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all.

Here it is in modern English:

Objective consideration of contemporary phenomena compels the conclusion that success or failure in competitive activities exhibits no tendency to be commensurate with innate capacity, but that a considerable element of the unpredictable must invariably be taken into account.

(Our aim is not poetry, but it sure as hell is not to sound like the second version, despite what you will come across in papers.)

Several short sentences about writing:

“The logic of writing, as you learned it in school,
Turns out to mean little more than an obsession with transition
And the scattering of rhetorical tics—overused, nearly
meaningless words and phrases.

In fact.
Indeed.
On the one hand.
On the other hand.
Therefore.
Moreover.
However.
In one respect.
Of course.
Whereas.
Thus.

These are logical indicators. Emphasizers. Intensifiers.
They insist upon logic whether it exists or not.
…
These words take the reader's head between their hands and force
her to look where they want her to.
…
If a piece is truly assured in its order… It needs no logical indicators.
It will be obvious when one sentence negates or affirms another.”

Style: Lessons in Clarity and Grace:

“Some writers try to fake coherence by lacing their prose with conjunctions like *thus*, *therefore*, *however*, and so on, regardless of whether they signal real logical connections.”

e.g. *The scale of the problem is large. The available resources are limited.*

This is better without a “ ; however, ” intruding in the middle.

Several short sentences about writing

"The point of learning the fundamental language of grammar and syntax
Isn't correctness or obeying the rules.
It's keeping the rules from obtruding themselves upon the reader
Because you've ignored them.

The reader is just like you,
Full of subtle, distracting feelings when things are going wrong in a sentence."

Several short sentences about writing

“*Imagine* sentences instead of writing them.
Keep them imaginary until you're happy with them.
An imaginary sentence somehow feels less *bound* than one you've written down.”

Writing a sentence before we’re happy with it leaves us vulnerable to status quo bias.
And what’s easier than going back to the drawing board?
Not leaving it in the first place till the drawing is right.

Several short sentences about writing

While writing, whole sentences or long strings of words may volunteer themselves.
Distrust these: they’re probably overused elsewhere.
Revise them.
Build sentences word by word.

Several short sentences about writing

The *result* of good writing may be something that seems to flow, as experienced by the reader.
The *process* of good writing is not one of flow.
This misconception is behind the idea of writers block.
Writing feels hard because it is.

“Your labor isn't a sign of defeat.
It's a sign of engagement.
The difference is all in your mind, but *what* a difference."