Seminar 8

Publication Bias, AI in Research, and Misinformation in Science

James Rice

UCL Department of Political Science

Monday, March 9, 2026

Today

1. Lecture recap:
   - AI in research: what it can now do, and where caution is still needed
   - Misinformation in science: from misleading claims to distorted evidence
   - publication bias, p-hacking, and p-screening
2. Reading summary:
   - West & Bergstrom (2021)
   - *Thinking Clearly with Data*, Chapter 7
3. Seminar task: searching for literature and comparing search tools
4. Discussion

Question for discussion:

How do we find reliable knowledge when the research environment itself can be biased?

AI in research

- AI is now genuinely useful for real research work, not only brainstorming.
- Strong uses include:
  - topic-based literature search across large corpora
  - rapid screening and clustering of papers
  - summarizing wide-ranging debates and identifying gaps in the literature
  - coding text and analyzing and cleaning data
- But it does **not** remove the need for human judgment:
  - it may hallucinate citations or overstate consensus
  - it cannot by itself judge scholarly quality, causal identification, or field significance

My view: AI is now a powerful research assistant and, in some tasks, a partial substitute for manual review. But it is not a substitute for critical thinking.

How to use AI well in academic research

Good practice

- use AI to expand search terms and keywords
- ask for alternative explanations and rival literatures
- verify every important citation manually
- pair AI search with Google Scholar, Web of Science, and library databases

Poor practice

- copying a generated literature review directly
- trusting references without checking them
- using AI summaries as evidence
- allowing one tool to define the whole field

Rule of thumb: use AI for *speed, coverage, and organisation*; use human judgement for *verification, interpretation, and argument*.

Misinformation in science

- We are not only worried about misinformation *about* science (e.g. climate denial, vaccine myths).
- We are also worried about misinformation *within* the scientific information system:
  - exaggerated claims from weak evidence
  - publication bias toward positive and novel findings
  - algorithmic ranking systems that amplify what is already visible (echo-chambers)
- If the visible literature is distorted, then students, journalists, policymakers, and researchers may all develop a biased view of what is actually known.

Misinformation is not only false content spreading online; it is also built into the institutions that produce and circulate knowledge.

Publication bias

- Publication bias occurs when statistically significant or dramatic findings are more likely to be published than null or mixed findings.
- This creates the classic file drawer problem: studies with weak or non-significant results disappear from view.
- The visible literature then overstates certainty, effect sizes, and sometimes even whether a relationship exists at all.
- This is one reason discussions of a broader **replication crisis** emerged across the natural and social sciences.

p-hacking and p-screening

P-hacking

- trying multiple specifications, outcomes, samples, or models
- reporting the version that "works"
- can produce p-values just below 0.05 even when no real relationship exists

P-screening

- many researchers may run related studies
- only the lucky significant result gets published
- bias can emerge even without deliberate bad faith

Why multiple testing is dangerous: if a true null is tested repeatedly, the probability of at least one false positive rises. So with many tries, "significance" becomes much less impressive than it first appears.

Reading summaries

West & Bergstrom (2021)

- Misinformation in science is **systemic**, not just individual error.
- Main mechanisms:
  - citation misdirection
  - predatory publishing
  - publication bias
- Key point: scientific literacy now requires understanding how search, ranking, and publishing shape what we read.

***Thinking Clearly with Data*, Ch. 7**

- Evidence becomes misleading when researchers over-compare and under-report.
- Main problems include forms of p-hacking
- Pre-registration is being rapidly adopted across publishing
- Key point: do not focus only on the p-value — focus on the research processes that generated the evidence.

Key takeaway:

Research findings are shaped by systems of selection, visibility, and reporting.

Why this matters beyond academia

- Most of you will consume research more often than produce it.
- News stories, policy briefs, workplace reports, and AI tools all summarise published evidence.
- If the visible evidence is biased, your view of "what the research says" may be biased too.
- So today's topic is not just about publishing papers:
  - it is about judging evidence well
  - and knowing what may be missing from view

Seminar task

Are economic sanctions more likely to succeed when they are imposed by broad international coalitions or by a single state?

- **Find literature on the research question using:**
  - Google Scholar;
  - UCL library database;
  - NotebookLM or another academic AI/search tool (Perplexity);
  - Out-of-the-box AI assistant (ChatGPT, Claude)
- **Seminar questions:**
  - What are the types of misinformation and bad research practices that lead to publication bias? What are some examples of each?
  - What are possible remedies for tackling these practices? Are these remedies practical? Do the remedies have any drawbacks?
  - How do you do background research for your own work?
  - What are ways we can try to ensure that the sources are reliable in our own work?

Whole-class discussion

Takeaway

A good researcher does not just ask whether a finding is interesting. They ask how the finding became visible in the first place. Research is shaped by incentives, search systems, and selection effects, so rigorous scholarship requires both statistical reasoning and information literacy.