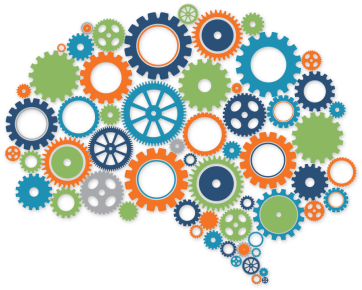


Critical Thinking & Innovation

Jake Chandler

Search Engines & Bias

LA TROBE UNIVERSITY

Introduction

Search

- We've focused on the analysis and evaluation of arguments bearing on a particular question
- We've said little about the process of search for arguments
- Search engines (SEs) now play a central role in this process for many people:

  In 2021, 63 billion Google searches in Australia!
- Gunn & Lynch (2019) note that "googling" exhibits 3 features
  - Primacy: It is our go-to search method for many questions
  - Seamlessness: We barely pay attention to the interface that provides us access
  - Trust: We take its outputs to be prima facie credible

Confidence in SE use

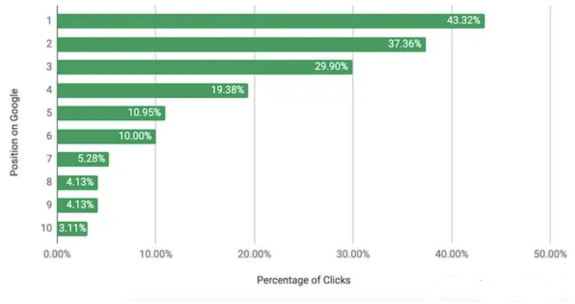
- Searchers have high levels of confidence in
  - their personal ability to effectively use search engines

    56% of 2012 US survey respondents reported being very confident and 37% somewhat confident (Purcell *et al* 2012)
  - the ability of search engines to provide credible and relevant data

    75% of 2012 US survey respondents considered search engines accurate and trustworthy

Confidence in SE use (ctd)

- 90% of searches don't even go past page 1 and rank strongly influences selection probability



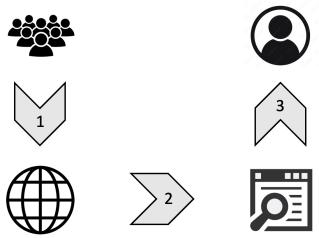
- Strong correlation persists when we control for relevance, etc

Concerns

- Still, we do have our suspicions…

"Do you really know what you're doing, or do you Google-search know?"

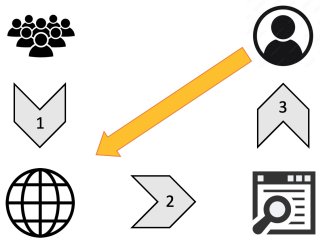
Concerns (ctd)

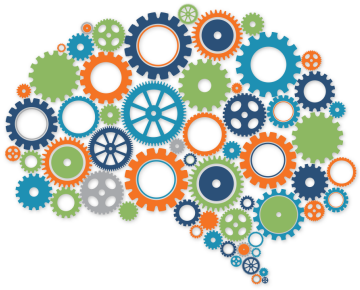


- Scope for problems at 3 points:
  (1) in the generation of the data
  (2) in the selection of the data presented in search results
  (3) in our inferences from the data presented

Concerns (ctd)

- These issues are amplified by the existence of feedback loops
- The main loop is due to the fact that users are also content creators: any biases in their views get fed back into the system





Critical Thinking & Innovation

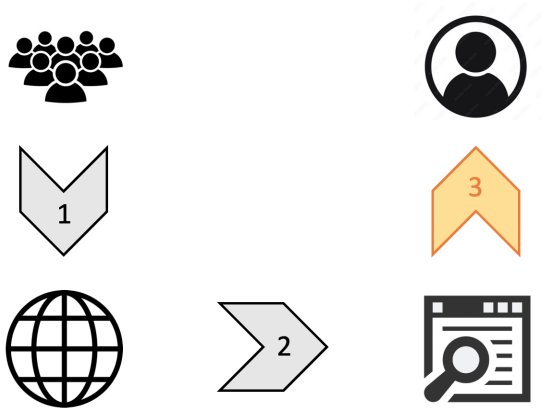
Jake Chandler

Search Engines & Bias

LA TROBE UNIVERSITY

User Biases and Content Biases

Concern: our inferences from the data presented



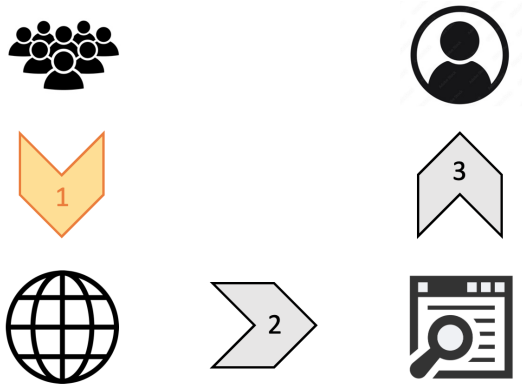
Evidence of user biases

- Like everywhere, cognitive biases show up in web-assisted inquiry (user bias; see Azzopardi 2021)
- Confirmation bias is one case in point:
  - Searchers viewed articles inconsistent with prior beliefs for significantly less time (Knobloch-Westerwick *et al* 2015)
  - Searchers observed for longer and were more likely to select links whose content aligned with their own political views (Ekstrom *et al* 2022)

Evidence of user biases (ctd)

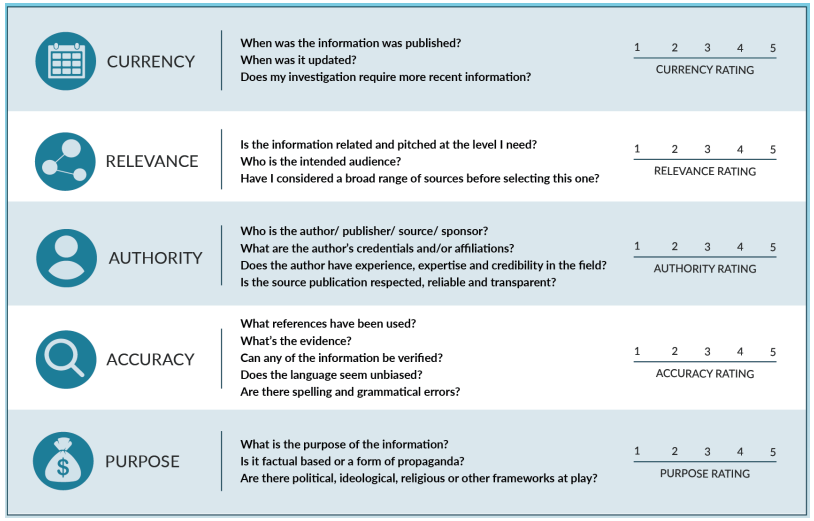
- Assessing reliability of sources is much harder than for traditional media (credibility uncertainty)
- This leads to reliance on heuristics for assessing trustworthiness (see Metzger & Flanagin 2013), a *potential* source of bias
- Examples:
  - Visual design
  - Name familiarity
  - Absence of advertising

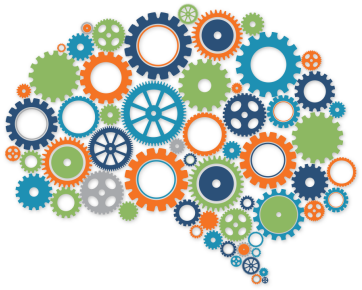
Concern: the generation of data



Concern: the generation of data

- The information *available in principle* online is skewed due to the content creators (content bias)
  - Cognitive biases (see previous slides)
  - Ulterior motives (e.g. political or commercial agendas)
  - Demographics (e.g. internet access or tech skills)
- Same for traditional media but less so, due to stricter checks
- These are obvious reasons for increased vigilance and we can mitigate this by
  - Using fact-checking sites
  - Following guidelines like the CRAAP test





Critical Thinking & Innovation

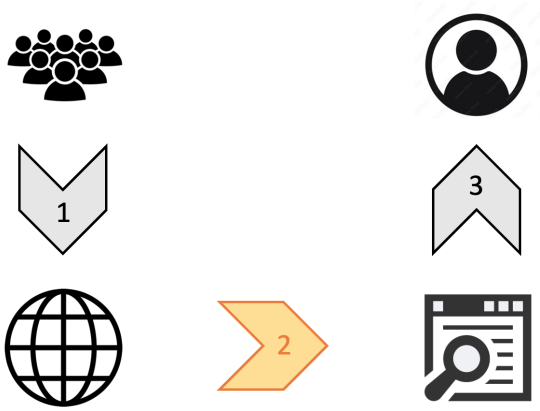
Jake Chandler

Search Engines & Bias

LA TROBE UNIVERSITY

Search Engine Biases

Concern: the selection of data in search results



Two issues with SEs

- The use of SEs to filter information adds a layer of issues, including
  (1) manipulability via search engine optimisation (SEO), the process of tweaking a page to improve search rank
  (2) potential exacerbation of other biases
- Let's say a little more about (2)

A problem of incentives

- Key issue: SEs only have indirect incentives to provide results that help find the truth
- Simpson (2012) notes:
  - SEs gain revenue from people using the service
  - People use the service when they tend to find something they are happy with
  - The things people are happy to find are not necessarily those that will objectively best assist them with inquiry

Google's ranking system

- We'll focus on Google (92% worldwide 2022 market share)
- It's algorithmic system is complicated, constantly changing and secret (partly to combat SEO manipulation)
- Over 200 factors influence ranking
- 2 of these, linked to Google's incentives, might cause problems:
  - PageRank: the key distinguishing feature of Google search, present since inception in 1998
  - Personalisation: rolled out for all users in 2009

PageRank

- A page's PageRank is increased by
  (a) the number of pages that link to it
  (b) the PageRank of the pages that link to it
- A page's PageRank is decreased by
  (c) the number of links leading out of the pages that link to it

- Two justifications for using PageRank (Brin & Page 1998)

Why PageRank? Random Surfer

- Random Surfer: Imagine a surfer who starts on a random page and follows links at random, before eventually getting bored and starting on another random page. The PageRank of a page is the proportion of the surfer's total visits that are visits to that page.
- The chances of the surfer visiting a page is increased by
  (a) the number of pages that link to it (more opportunities to visit)
  (b) the PageRank of the pages that link to it (greater chances to visit places from which the page can be accessed)
- The chances of the surfer visiting a page is decreased by
  (c) the number of links leading out of the pages that link to it (more opportunities to go elsewhere instead)

Why PageRank? Voting

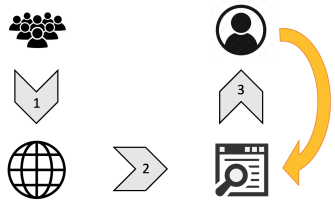
- Voting: Imagine that each page can vote for other content it thinks is important, dividing this vote equally among pages it votes for, and that pages that receive more votes get a greater say. The PageRank of a page corresponds to its share of the votes.
- The perceived importance of a page, according to the other pages is increased by
  (a) the number of pages that link to it (more votes for the page)
  (b) the PageRank of the pages that link to it (votes from more important pages)
- The perceived importance of a page is decreased by
  (c) the number of links leading out of the pages that link to it (votes that are more divided among more alternative pages)

The rich get richer

- PageRank tracks
  - perceived importance among content creators
  - traffic among users
- Ranking by PageRank amplifies the influence of content creators
- Information that already receives more attention because of the hyperlinks they provide gets additional attention because of search

Personalisation

- Personalisation involves the use of user info in determining search ranking, e.g.:
  - click/browsing/search histories
  - location
  - language
- It could *potentially* create a second kind of feedback loop in which people's beliefs affect the evidential basis from which they form those beliefs



Filter bubbles

- Eliser (2012): SE personalisation amplifies confirmation bias, creating filter bubbles

> "We thought that the Internet was going to connect us all together…What it's looking like increasingly is that the Web is connecting us back to ourselves. There's a looping going on where if you have an interest, you're going to learn a lot about that interest. But you're not going to learn about the very next thing over. And you certainly won't learn about the opposite view…If you Google some sites about the link between vaccines and autism, you can very quickly find that Google is repeating back to you your view about whether that link exists and not what scientists know, which is that there isn't a link between vaccines and autism. It's a feedback loop that's invisible. You can't witness it happening because it's baked into the fabric of the information environment."

What's the evidence?

- The empirical evidence is very inconclusive
- *Question 1:* To what extent does personalisation happen?
  - Hannak *et al* (2013): on average, 11.7% of results show differences due to personalisation on Google with strong variation by search query result ranking; this variation was largely tied to user accounts
  - DuckDuckGo (2018): significant differences in ranking results for politically sensitive topics (e.g. vaccination, gun control) between users, controlling for factors such as time and location
  - Krafft *et al* (2019): no substantial differences due to personalisation beyond those due to regional variation

What's the evidence? (ctd)

- *Question 2:* To what extent does personalisation happen *in a way that amplifies biases*?
  - Barely any empirical data (that I am aware of)
  - Fletcher & Nielsen (2018): evidence that people who use SEs to search for news have a *more* politically balanced range of sources!
- In sum:
  - Google has the motive and means to personalise
  - Personalisation has the potential to amplify confirmation bias
  - We have no real sense of the scale of the problem, if any
- Advice:
  - Search for information both supporting and undermining one's preferred position
  - Consider using more neutral SEs (e.g DuckDuckGo)

Intellectual virtues and web search

- Heersmink (2019): search can be improved by consciously cultivating character traits known as intellectual virtues
- Examples (from Baehr 2021):
  - Curiosity: "A disposition to wonder, ponder, and ask why. A thirst for understanding and a desire to explore."
  - Humility: "A willingness to "own" one's intellectual limitations and mistakes."
  - Autonomy: "A capacity for active, self-directed thinking. An ability to think and reason for oneself."
  - Open-mindedness: "An ability to "think outside the box." Gives a fair and honest hearing to competing perspectives."

Intellectual virtues and web search (ctd)

- This recommendation is arguably too vague to be very effective in this context
- But it does highlight an often overlooked part of an intellectual education: the development of a good mindset

  > "Many highly intelligent people are poor thinkers. Many people of average intelligence are skilled thinkers. The power of a car is separate from the way the car is driven." (Edward de Bono)

- I hope this subject will prove to have made some contribution to your learning to become good drivers

References

- Azzopardi, L. (2021). Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21)*. Association for Computing Machinery, New York, NY, USA, 27–37. https://doi.org/10.1145/3406522.3446023
- Baehr, J. (2021) *Deep in Thought: A Practical Guide to Teaching for Intellectual Virtues*. Harvard Education Press.
- Brin S, & Page L (1998) The anatomy of a large-scale hypertextual web search engine, WWW 1998 (Seventh International World-Wide Web Conference), Brisbane, Australia

References (ctd)

- DuckDuckGo (2018). Measuring the "Filter Bubble": How Google is Influencing What You Click, December 4, 2018, https://spreadprivacy.com/google-filter-bubble-study/
- Fletcher, F. & Nielsen, R.K. (2018) Automated Serendipity: The effect of using search engines on news repertoire balance and diversity. *Digital Journalism*, 6:8, 976-98
- Gunn. H. & M.P. Lynch (2019). Googling. In David Coady & James Chase (eds.), *Routledge Handbook of Applied Epistemology*. New York, NY, USA: pp. 41-53.

References (ctd)

- Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring personalization of web search. In *Proceedings of the 22Nd International Conference on World Wide Web* (pp. 527–538). Geneva, Switzerland: International World Wide Web Conferences Steering Committee
- Heersmink, R. (2018). A Virtue Epistemology of the Internet: Search Engines, Intellectual Virtues and Education, *Social Epistemology*, 32:1, 1-12,
- Knobloch-Westerwick, S., B.K. Johnson, & A. Westerwick (2015). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication* 20, 2 (2015), 171–187.

References (ctd)

- Krafft, T.D., Gamer, M. & Zweig, K.A. (2019) What did you see? A study to measure personalization in Google's search engine. *EPJ Data Sci.* 8, 38.
- Masterton, G., & Olsson, E. J. (2018). From Impact to Importance: the Current State of the Wisdom-of-Crowds Justification of Link-based Ranking Algorithms. *Philosophy & Technology*, 31(4), 593-609.
- Metzger, M.J. & A.J. Flanagin (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics* 59: 210-220.
- Purcell K, Brenner J, & Rainie L. (2012). Search engine use 2012. Washington, DC, https://www.eff.org/files/pew_2012_0.pdf

References (ctd)

- Simpson, T.W. (2012) Evaluating Google as an Epistemic Tool. *Metaphilosophy* 43 (4):426-445.
- Stanovich, K. E., West, R. F., and Toplak, M. E. (2016). *The Rationality Quotient - Toward a Test of Rational Thinking.* Cambridge, MA: MIT Press.
- White, R.W. (2014). Belief dynamics in web search. *Journal of the Association for Information Science and Technology* 65, 11 (2014), 2165–2178.