Information Bias in Search Engines

Frank Hopfgartner and Monica L Paramita
CyCAT (Cyprus Center for Algorithmic Transparency)

Frank Hopfgartner

MY POSITIONS

- Senior Lecturer in Data Science (University of Sheffield, UK)
- Visiting Lecturer (Northeastern University, China)
- Advanced Partner (Cy. Center for Algorithmic Transparency)

MY DUTIES

- Head of Information Retrieval Research Group
- Deputy Director of Research of Information School
- Coordinator of MSc in Data Science programme

MY INTERESTS

- Intersection of Information Access and Data Science
- Focus on personal data analysis, e.g., interaction log files, heterogeneous sensor data

MY BACKGROUND

- PhD in Computing Science (University of Glasgow)
- Previous positions in Glasgow, Berlin, Dublin, Berkeley, London

Monica L Paramita

MY POSITION

- Post-doctoral Researcher at the Information School, University of Sheffield

MY DUTIES

- Researcher at the CyCAT project and Rolls-Royce project
- Lecturer in Information Retrieval module

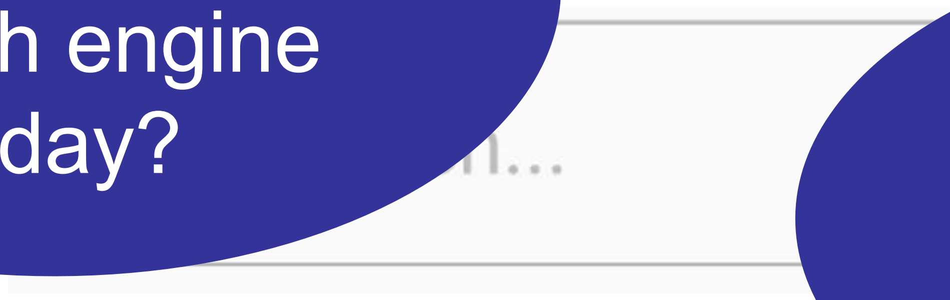
MY INTERESTS

- Cross-lingual similarity
- Bias in information retrieval

MY BACKGROUND

- PhD in Information Science (University of Sheffield)

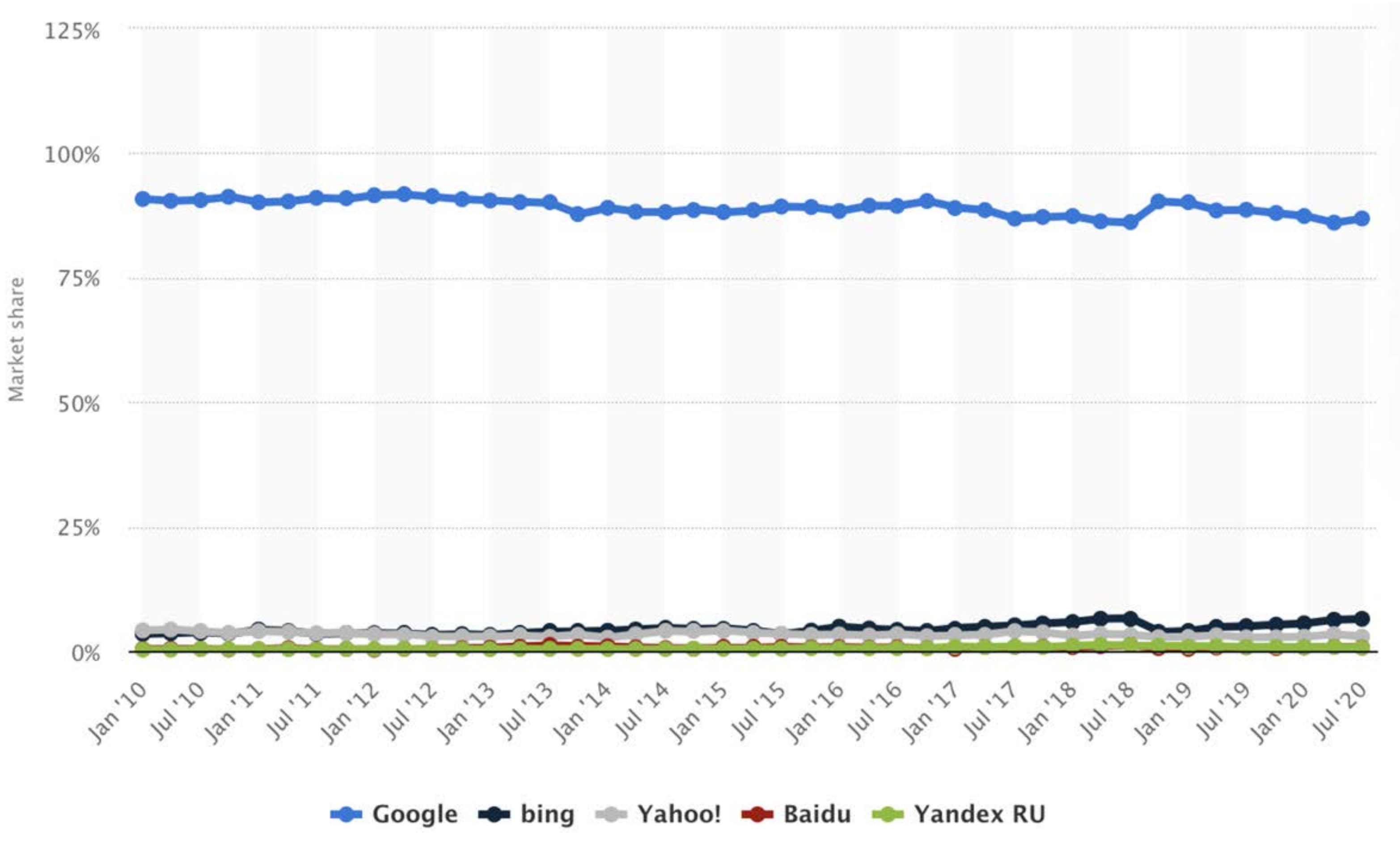
Web search engines



How often have *you* used a search engine today?

And which search engine did you use?

Some search engine stats



Worldwide Desktop market share of leading search engines (2010-2020)

- Nearly 93% of all web traffic comes through search engines
- Google processes 2 trillion searches per year

https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

In Search we trust

- “Search has assumed a position of central importance in the way that people access and use online information and services every day”
- “By shaping both what people know and how people know it, search engines and their organisations are able to wield an immense amount of social power”

Noel Carroll. In Search we Trust: Exploring How Search Engines are Shaping Society. *Int. Journal of Knowledge Society Research*, 5(1):12-27, 2014. DOI: 10.4018/ijksr.2014010102

Aims and Objectives

- The aim of this seminar is to introduce the notion of information bias in search engines.
- In particular, we focus on cultural biases.
- By the end of this lecture you should be able to
  - identify different types of cultural information biases in search engines
  - reflect on how a search engine could make users aware of biased search results

Outline

- Background
- **Search engine basics**
- Examples of cultural bias
  - Political bias
  - Gender bias
- Group activity

The entire search process

! "#$

%&"'# (

Peters, C., Braschler, M. and Clough, P. (2012) Multilingual Information Retrieval: From Research to Practice, Springer: Heidelberg, Germany, ISBN 978-3-642-23007-3, 217 pages.

IR process

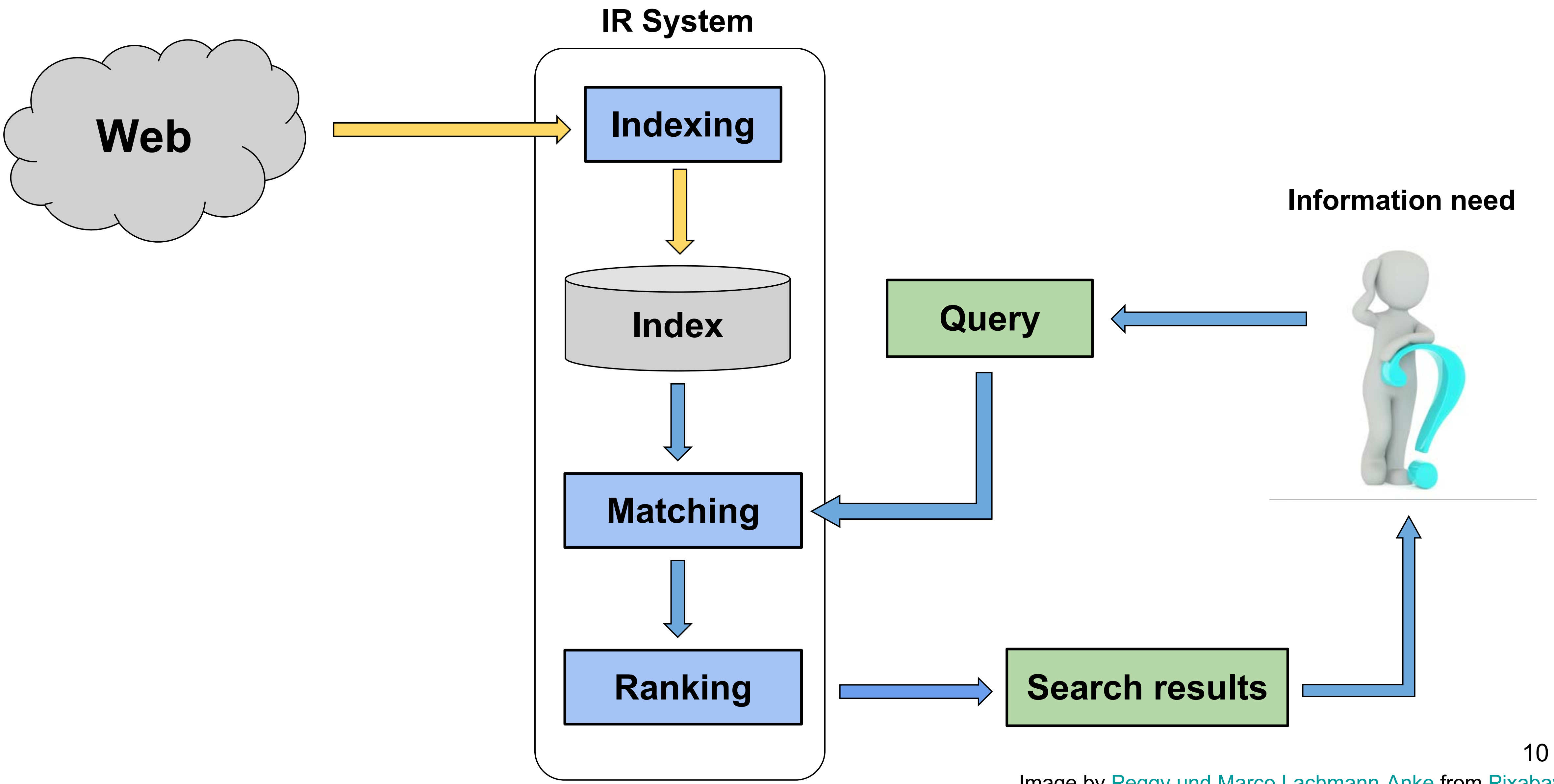


Image by Peggy und Marco Lachmann-Anke from Pixabay

Personalisation

- Personalised Information Retrieval (IR) systems rank items that match users' interests higher in the search results

- Personalised IR uses user behaviour to build information regarding users' interests
  - Previously used queries
  - Previously clicked results (implicit relevance feedback)
    - This assumes that a document clicked by user is relevant to their query
  - Users' location

Outline

- Background
- Search engine basics
- **Examples of cultural bias**
  - **Political bias**
  - **Gender bias**
- Group activity

Problem: Filter bubbles

Personalised search engines *tailor content presentation* based on our interests

This results in *'filter bubbles'* that amplify **confirmation bias** as it filters out alternative viewpoints

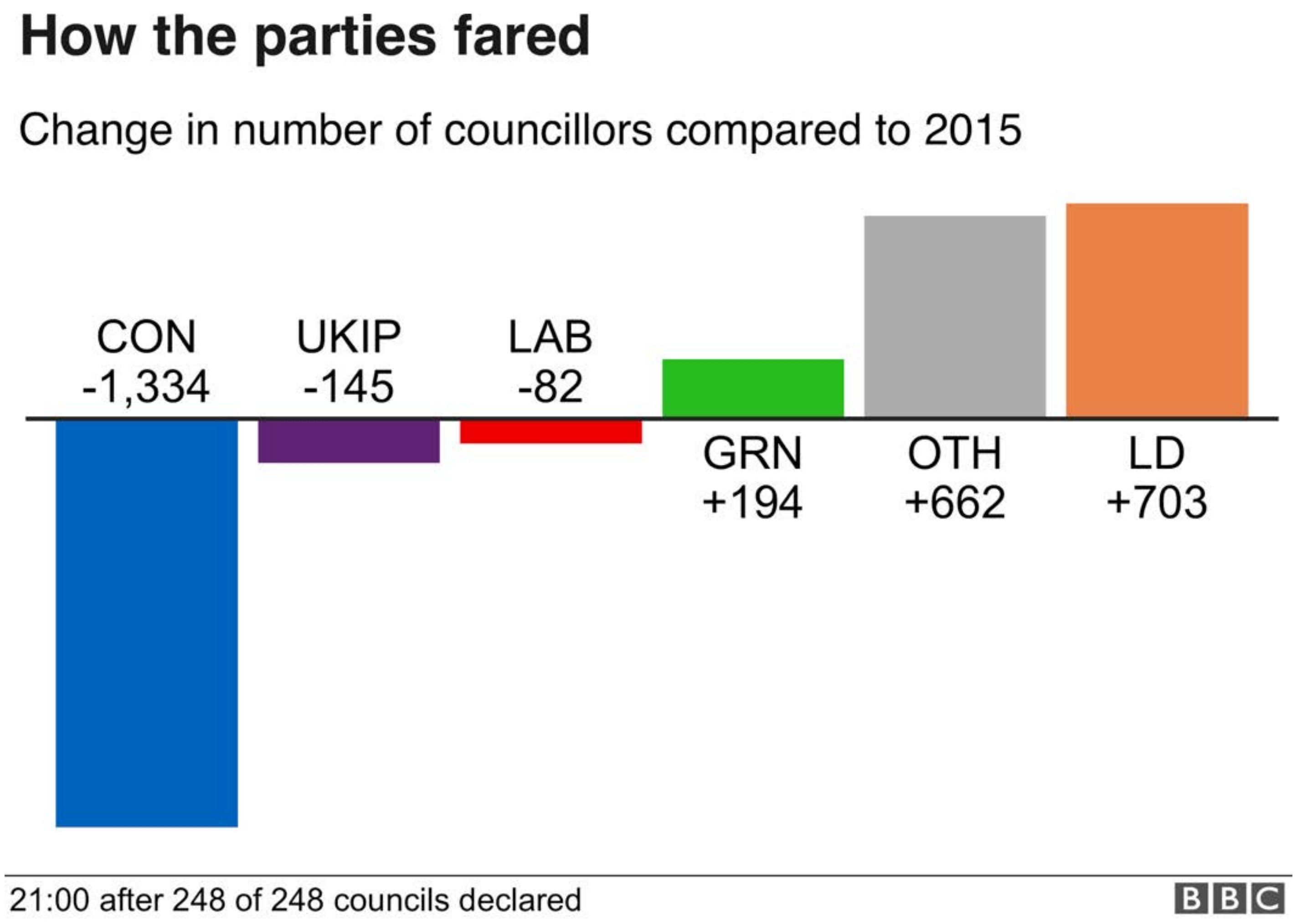
Closely linked with *political polarization*

"*[Personalisation] moves us very quickly toward a world in which the internet is showing us what it thinks we want to see, but not necessarily what we need to see.*

Eli Pariser, internet activist, 2011

In his TED Talk, Eli Pariser raises awareness of the negative effect of online filter bubbles

Political bias case study: 2019 Local Elections in England



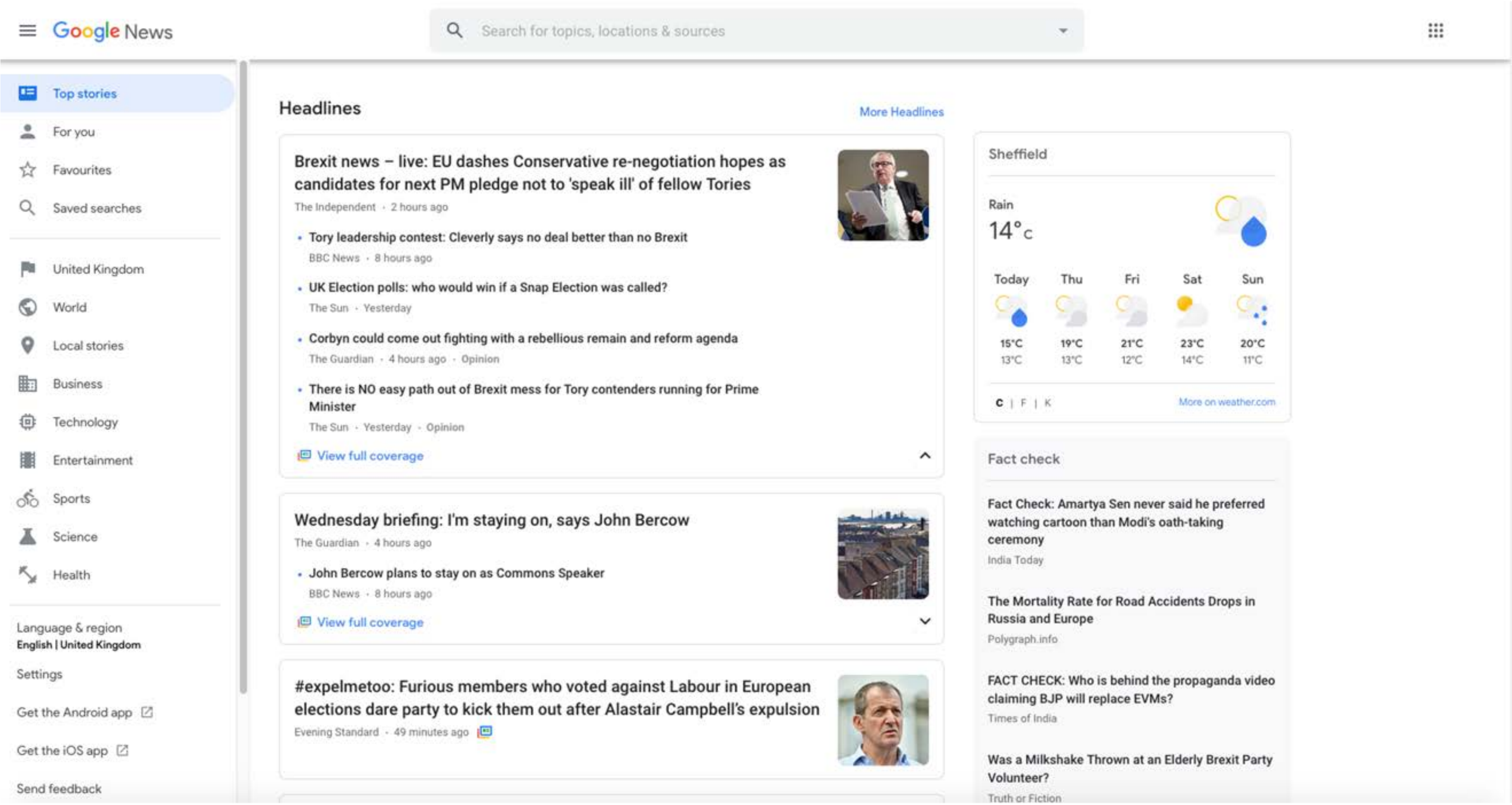
https://www.bbc.co.uk/news/uk-politics-48091592

Media bias

“Get on with it”

“Stop Brexit”

Now let's add a news search engine



Effect of the filter bubble

Check for updates

Article

Communication Research
2014, Vol. 41(8) 1019–1041

DOI: 10.1177/0093650213497979
crx.sagepub.com
SAGE

Automating the News: How Personalized News Recommender System Design Choices Impact News Reception

Michael A. Beam[1]

Abstract

This study investigates the impact of personalized news recommender system design on selective exposure, elaboration, and knowledge. Scholars have worried that proliferation of personalization technologies will degrade public opinion by isolating people from challenging perspectives. Informed by selective exposure research, this study examines personalized news recommender system designs using a communication mediation model. Recommender system design choices examined include computer-generated personalized recommendations, user-customized recommendations, and full or limited news information environments based on recommendations. Results from an online mock election experiment with Ohio adult Internet users indicate increased selective exposure when using personalized news systems. However, portals recommending news based on explicit user customization result in significantly higher counterattitudinal news exposure. Expected positive effects on elaboration and indirect effects on knowledge through elaboration are found only in personalized news recommender systems that display only recommended headlines. Lastly, personalized news recommender system use has a negative direct effect on knowledge.

Keywords

Internet news, personalization, selective exposure, news information processing, news knowledge

[1]Washington State University, Pullman, USA

Corresponding Author:
Michael A. Beam, Washington State University, 106A Murrow Complex East, P.O. Box 642520, Pullman, WA 99164-2520, USA.
Email: michael.beam@wsu.edu

The authors studied impact of personalized news recommender systems on users.

"Users more often choose news stories that align with their own preferences."

"Personalised news systems usage had negative direct effect on knowledge gain."

Filter bubbles

Check for updates

Article

BIR

Business Information Review
2017, Vol. 34(3) 150–160

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0266382117722446
journals.sagepub.com/home/bir
SAGE

Fake news and ideological polarization: Filter bubbles and selective exposure on social media

Dominic Spohr
Independent Scholar

Abstract
This article addresses questions of ideological polarization and the filter bubble in social media. It develops a theoretical analysis of ideological polarization on social media by considering a range of relevant factors. Over recent years, fake news and the effect of the social media filter bubble have become of increasing importance both in academic and general discourse. The article reviews the assumption that algorithmic curation and personalization systems place users in a filter bubble of content that decreases their likelihood of encountering ideologically cross-cutting news content. At the intersection of new media, politics and behavioural science, the article establishes a theoretical framework for further research and future actions by society, policymakers and industries.

Keywords
Facebook, fake news, ideological polarization, misinformation, social media

Introduction

This article addresses questions of ideological polarization and the filter bubble in social media. Over recent years, fake news and the effect of the social media filter bubble have become of increasing importance both in academic and general discourse. This has been exacerbated by the perceived role of fake news and selective news filtering in echo chambers of our own beliefs and is the stronger cause of polarization (Pariser, 2011; Rader and Grey, 2015) and a critical factor in the growing importance of fake news. The other source of polarization discussed has been around for much longer and has its roots in psychology and behavioural economics. By this argument, selective exposure behaviour, confirmation bias

Ideological polarization and information consumption are intertwined

Personalization systems make things worse

D. Spohr. *"Fake news and ideological polarization. Filter bubbles and selective exposure on social media", In Business Information Review, 34(3):150-160, 2017.*

The Making of a YouTube radical

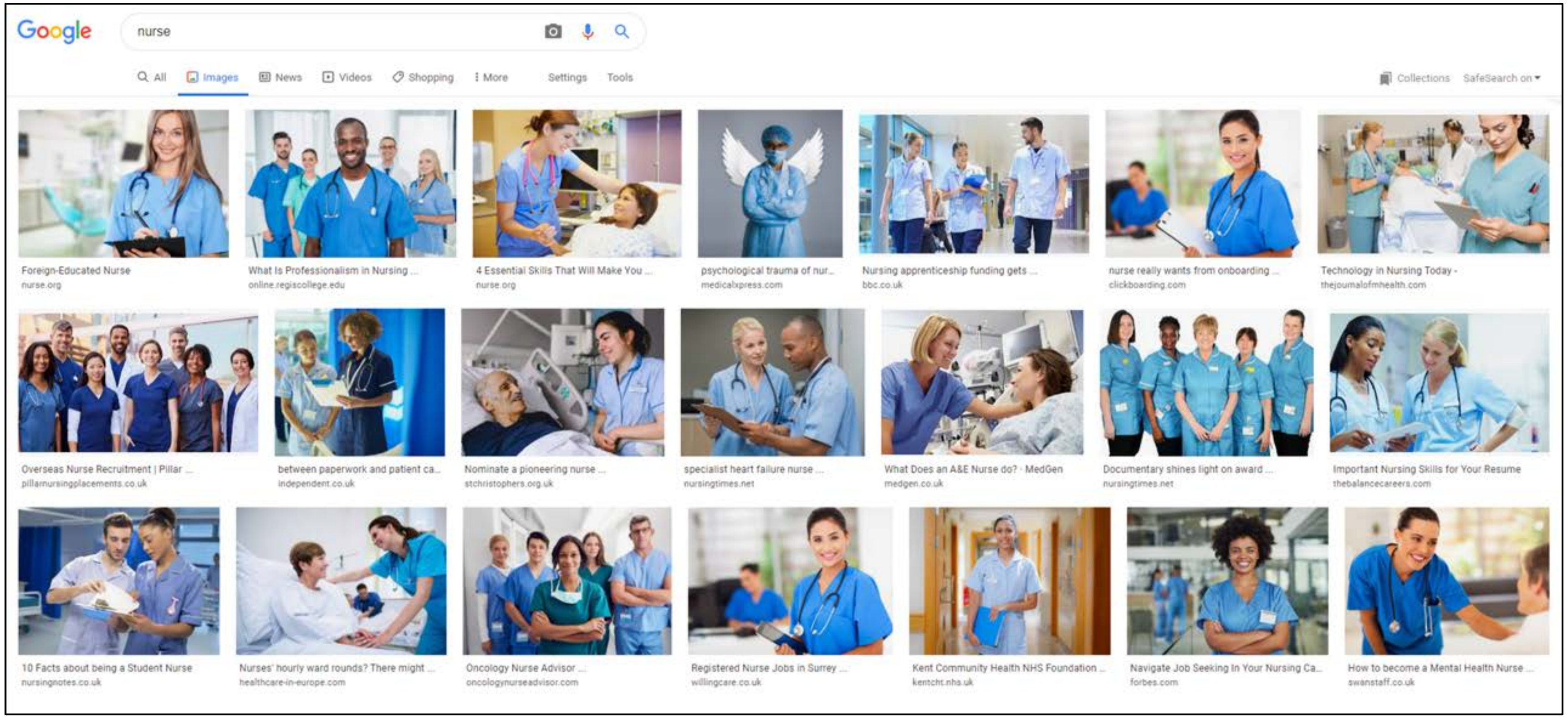
"Soon, he was pulled into a far-right universe, watching thousands of videos filled with conspiracy theories, misogyny and racism."

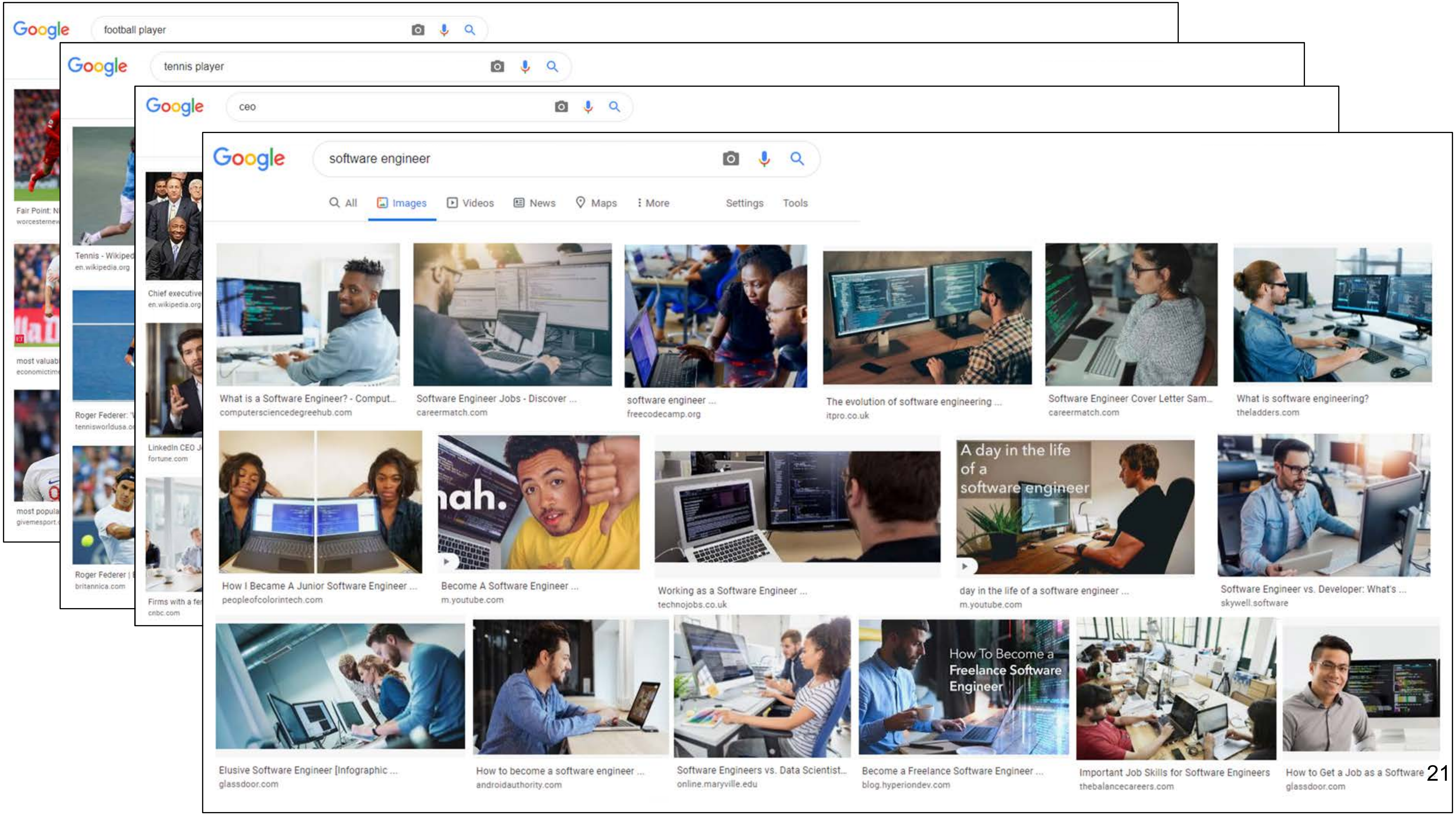
https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

Gender bias (1/3)

- Kay et al. (2015) studied Google image results depicting different occupations
  - E.g. “police officer”, “nurse”, “construction worker”, etc.

- They investigated whether the gender depicted in the images represent the proportion of those in the real world




football player
tennis player
ceo
software engineer
All
Images
Videos
News
Maps
More
Settings
Tools
What is a Software Engineer? - Comput...
computersciencedegreehub.com
Software Engineer Jobs - Discover ...
careermatch.com
software engineer ...
freecodecamp.org
The evolution of software engineering ...
itpro.co.uk
Software Engineer Cover Letter Sam...
careermatch.com
What is software engineering?
theladders.com
How I Became A Junior Software Engineer ...
peopleofcolorintech.com
Become A Software Engineer ...
m.youtube.com
Working as a Software Engineer ...
technojobs.co.uk
A day in the life of a software engineer
day in the life of a software engineer ...
m.youtube.com
Software Engineer vs. Developer: What's ...
skywell.software
Elusive Software Engineer [Infographic ...
glassdoor.com
How to become a software engineer ...
androidauthority.com
Software Engineers vs. Data Scientist...
online.maryville.edu
How To Become a Freelance Software Engineer
Become a Freelance Software Engineer ...
blog.hyperiondev.com
Important Job Skills for Software Engineers
thebalancecareers.com
How to Get a Job as a Software
glassdoor.com
Tennis - Wikiped
en.wikipedia.org
Roger Federer
tennisworldusa.o
Roger Federer |
britannica.com
Chief executive
en.wikipedia.org
LinkedIn CEO J
fortune.com
Firms with a fe
cnbc.com
Fair Point: N
worcesternew
most valuab
economictim
most popula
givemesport.

Gender bias (2/3)

- Kay et al. (2015) found that image search results slightly exaggerated gender stereotypes, compared to the US Bureau of Labor and Statistics data
  - Under-representations of images of women in a number of roles
    - e.g. 11% of images of CEO women compared to the real world (27%).
  - The minority gender (i.e., women) are represented less professionally
    - Provocative images of “female construction worker”
  - Roles such as “telemarketers” retrieved mostly images of women (although it’s 50% men, and 50% women)

Gender bias (3/3)

- Otterbacher et al. (2017) also analysed bias in image search results and found:
    - Query containing competence traits (e.g., “rational”, “intelligent”) retrieve more images of men
    - Query containing warmth and communality traits (e.g., “emotional”) retrieve more images of women
    - Backlash effect for women who have “competency” traits

Impacts of bias

- **Change user perception of a particular aspect**
  - Distribution of gender in a particular occupation (Kay et al., 2015)
- **Inaccurate representations of people / groups of people**
  - Female are represented as provocative/sexy subject more than men (Kay et al., 2015)
  - Black men are often suggested to be criminals (Woodruff et al., 2018)
- **Influence on decision making**
  - Manipulate user understanding of an unknown topic (Novin & Meyers, 2017)
  - Influence undecided users in voting decisions (Epstein & Robertson, 2015; Kulshrestha et al., 2017)

How should a search engine address these biases?

- Approach 1: A search engine should re-rank the results to reduce unfairness and bias and achieve the “ideal” results.
    - But, what is the “perfect” or “ideal” results?
        - Representative of those in the real world (Gao & Shah, 2020)
        - Representative of what the real world should look like
            - Balanced model (Kay et al., 2015): proportion of groups to be equal or closer to equal
- Approach 2: Ranking should stay the same but users should be informed if there is a certain bias in the results
    - Increasing transparency of search engine results (Snow, 2018)

Outline

- Background
- Search engine basics
- Examples of cultural bias
  - Political bias
  - Gender bias
- **Group activity**

Group Activity

We will ask you to work in a group to reflect on your knowledge about bias:

- Pre-questionnaire (individual)
- Activity 1 (group): 10 minutes
- Activity 2 (group): 30 minutes

Activity 1

> **“Imagine that you are using a news search engine to look for the topic: Covid-19”**

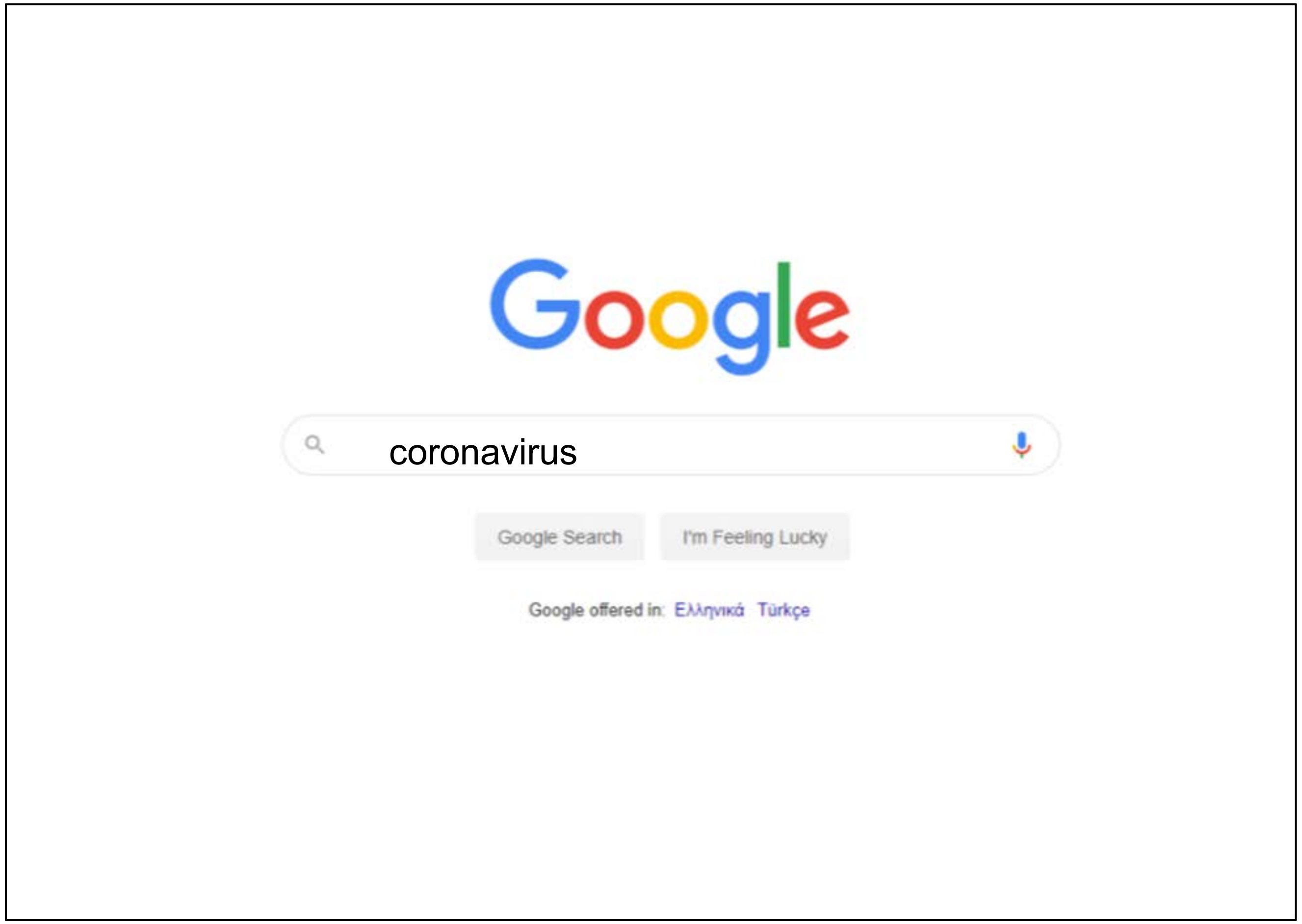
Activity 1:

- What types of information bias would be problematic in this situation (e.g., political bias, gender bias)?
  - Please identify at least 5 types of bias
- Please rank each bias based on its impact to users
  - I.e., bias that is the most important for a search engine to highlight should be rank the highest

Activity 2

- You have identified a number of biases that users should be informed of when using a news search engine
- How should these biases be shown to the users?
- In the next activity, you are asked to work in a group to design a mock-up search engine to visualise this information
  - Your mock-up should include **different stages of the interactions**
  - You should add **annotations** on each stage to describe these interactions. I.e., what the user/system does, and the interaction between the user and the system

Activity 2: Example

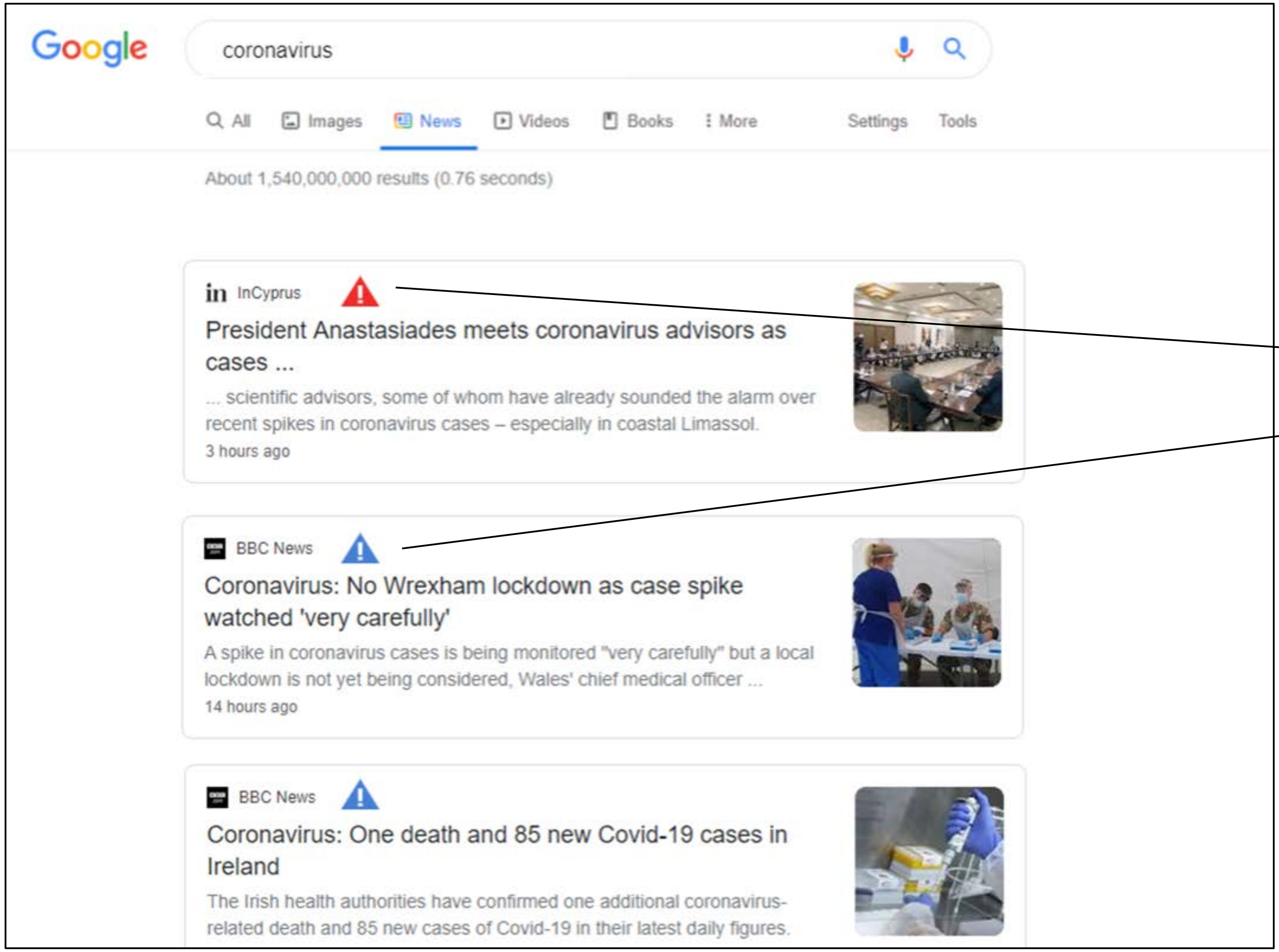


Annotations - STAGE 1

A user enters their search keywords in the search box.

For example, they might type “coronavirus” as a query as they were interested to see the latest news related to coronavirus.

Activity 2: Example



Annotations - STAGE 2

Step 2 shows the results for news articles relevant to the query (“coronavirus”).

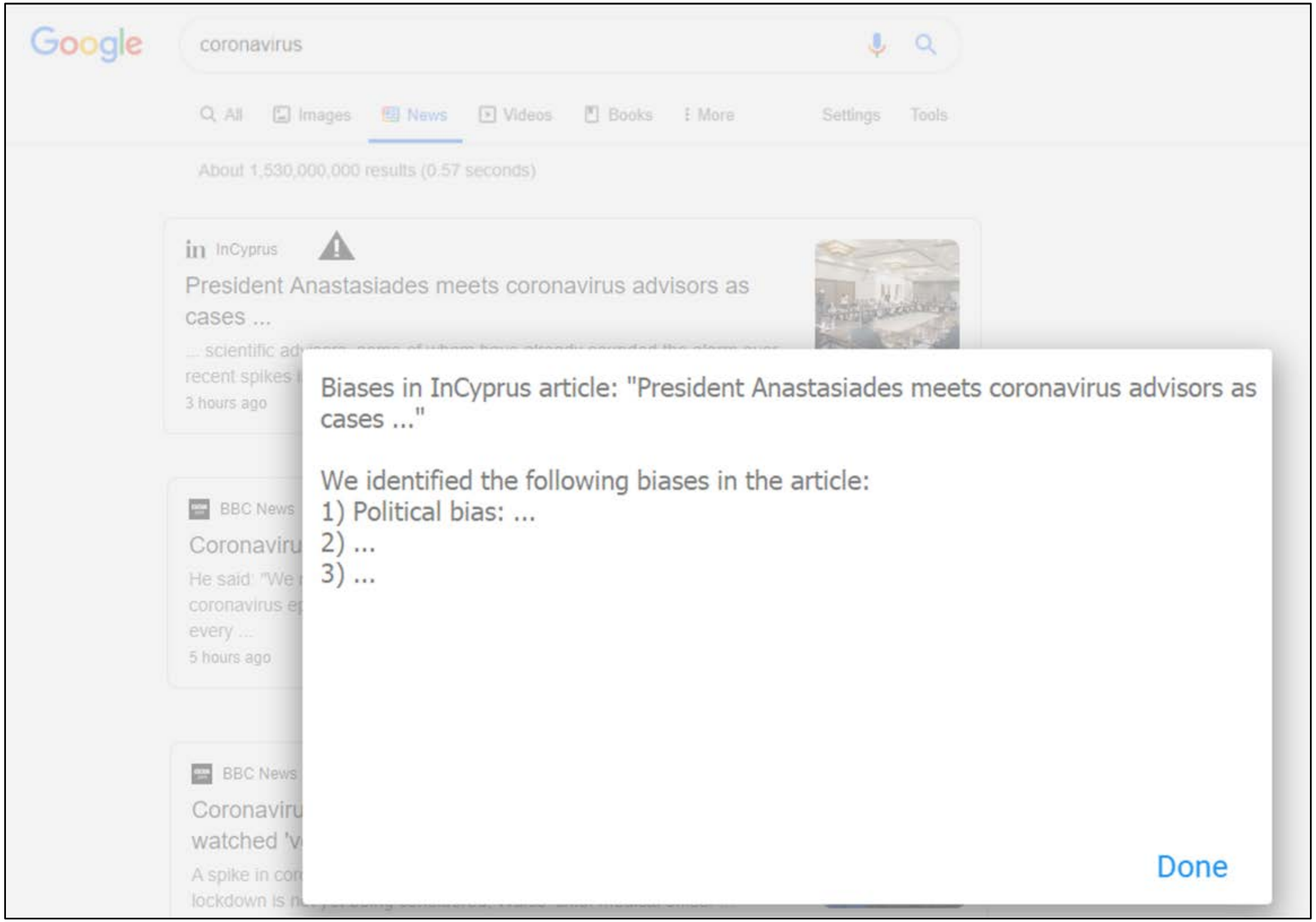
In each result, an icon is displayed to visualise the level of bias identified in the article.

If some bias is identified, the icon is shown in **red**.

If no bias is identified, the icon is shown in **blue**.

Users can click on the icon to see more information (see Stage 3).

Activity 2: Example



Annotations - STAGE 3

When a user clicks on the red icon, a pop-up window will appear.

It describes the **types of bias** identified in the article and other relevant information.

Note: this is a very basic example.

How would *you* design your system to inform relevant biases to users?

Activity 2: “How should a search engine visualise these biases to users?”

- Please design a mock-up search engine to include this information in Google Slides
- Your mock-up should include different stages of the interactions
- You should add annotations on the slides to describe these interactions

Evaluation of prototype

- **Thank you for your contribution in this seminar**
- What’s next?
  - We will run similar studies in other universities.
  - We will create some prototypes of the bias intervention tools based on feedback provided by you and others.
  - You will be invited to give feedback on these prototypes.
- If you have any questions about this study, please contact:
  - Frank Hopfgartner (f.hopfgartner@sheffield.ac.uk)
  - Monica Paramita (m.paramita@sheffield.ac.uk)