The Filter Bubble: a threat for plural information?

The dangers of personalizing filters

Pietro De Nicolao

Computer Ethics (2016-2017), Politecnico di Milano

December 13, 2016

Introduction

What is the filter bubble?

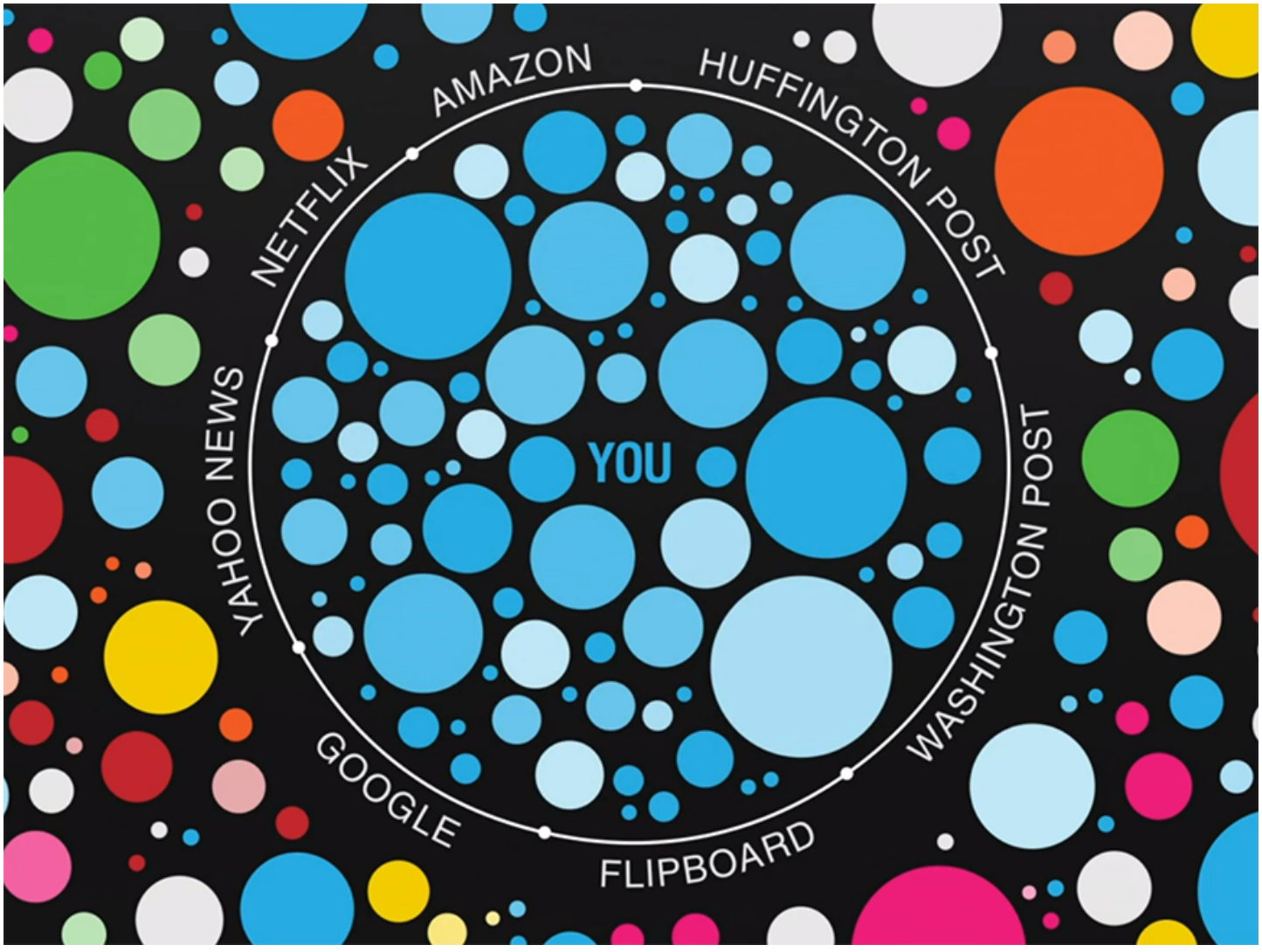


Figure 1: *A unique universe of information for each of us* (Pariser 2011, p. 10)

Definitions

Personalized search

Personalized search *refers to search experiences that are tailored specifically to an individual's interests by incorporating information about the individual beyond specific query provided. (Wikipedia)*

Filter bubble

A ***filter bubble*** *is the restriction of a user's perspective that can be created by personalized search technologies. (Haughn 2015)*

Political pluralism in the media

Political pluralism *in the media refers to the fair and diverse representation of and expression by various political and ideological groups, including minorities, in the media. (Leuven et al. 2009, p. 12)*

What I will show

In this presentation, I will show that:

- **Personalization reduces information pluralism** by giving users only what they like to see
- Personalizing filters **define our perception** of the world and are not *neutral intermediaries*
- Recommender systems are **relevance maximizers**
  - Important but non-relevant stories can be left out
  - Different point of views are shown less
- **Transparency** about the use of data and about the algorithms is needed
  - Users must know when personalization is active
  - Users should be able to control it

Concerns about the Filter Bubble

The dangers of personalization

The book *The Filter Bubble* (Pariser 2011) describes many risks associated with it:

- Data collection and privacy
- Democracy
- **Information** (I will focus on this)
- Freedom
- Creativity
- Censorship
- Serendipity

Importance vs. relevance of news stories

Two metrics can be defined for news stories:

- **Importance**: intrinsic "value" of a story with respect to society
- **Relevance**: probability that a story will be "liked" by the user; performance index of the recommender system

Recommender systems (personalizing filters) are **relevance maximizers**

Example

> *"A squirrel dying in front of your house may be more **relevant** to your interests right now than people dying in Africa."*
>
> *– Mark Zuckerberg (Facebook CEO)*

Concerns about information

Friendly world syndrome

Personalizing filters block *important*, but *unpleasant* things:

- Some **topics** will always be *not likable*: war, homelessness, poverty...
- Different **point of views** are less *relevant* to us

Autonomy

Autonomy of the readers is compromised, as they can't choose what's in or what's out their "bubble"

Worst case scenario

Deliberate use of filters to shape the public opinion, by governments or multinational companies

Case study: Facebook News Feed

Facebook is too friendly!

Suppose that you are a Facebook user and you identify as a **liberal**, and you have both liberals and conservatives friends.

- News Feed recommendation algorithm: **you get more posts which reflect what you like** (relevant to you)
- You may not see conservatives' stories at all, if you interact less with your conservative friends
- **Cross-cutting stories** (those different from our viewpoint) are less likely to reach us
  - . . . but how much?

89.4% of under-30 Italians uses Facebook (CENSIS 2016)

- The issue of biased content is certainly important!

Facebook: Exposure to ideologically diverse content

Facebook published a study (Bakshy, Messing, and Adamic 2015) on *Science* about how likely are users to **view and interact with *cross-cutting content***.

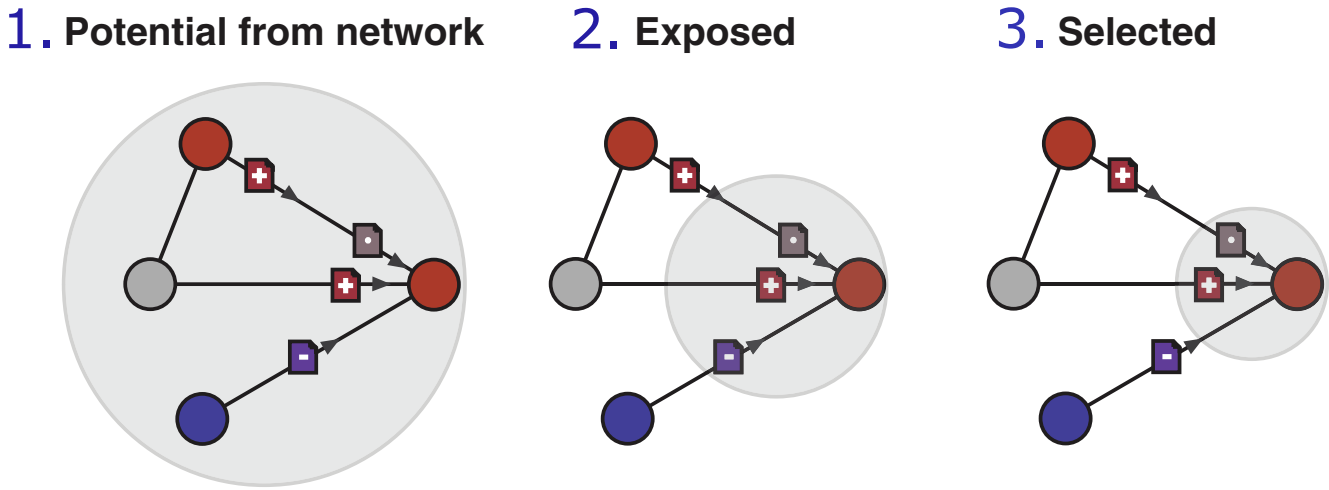
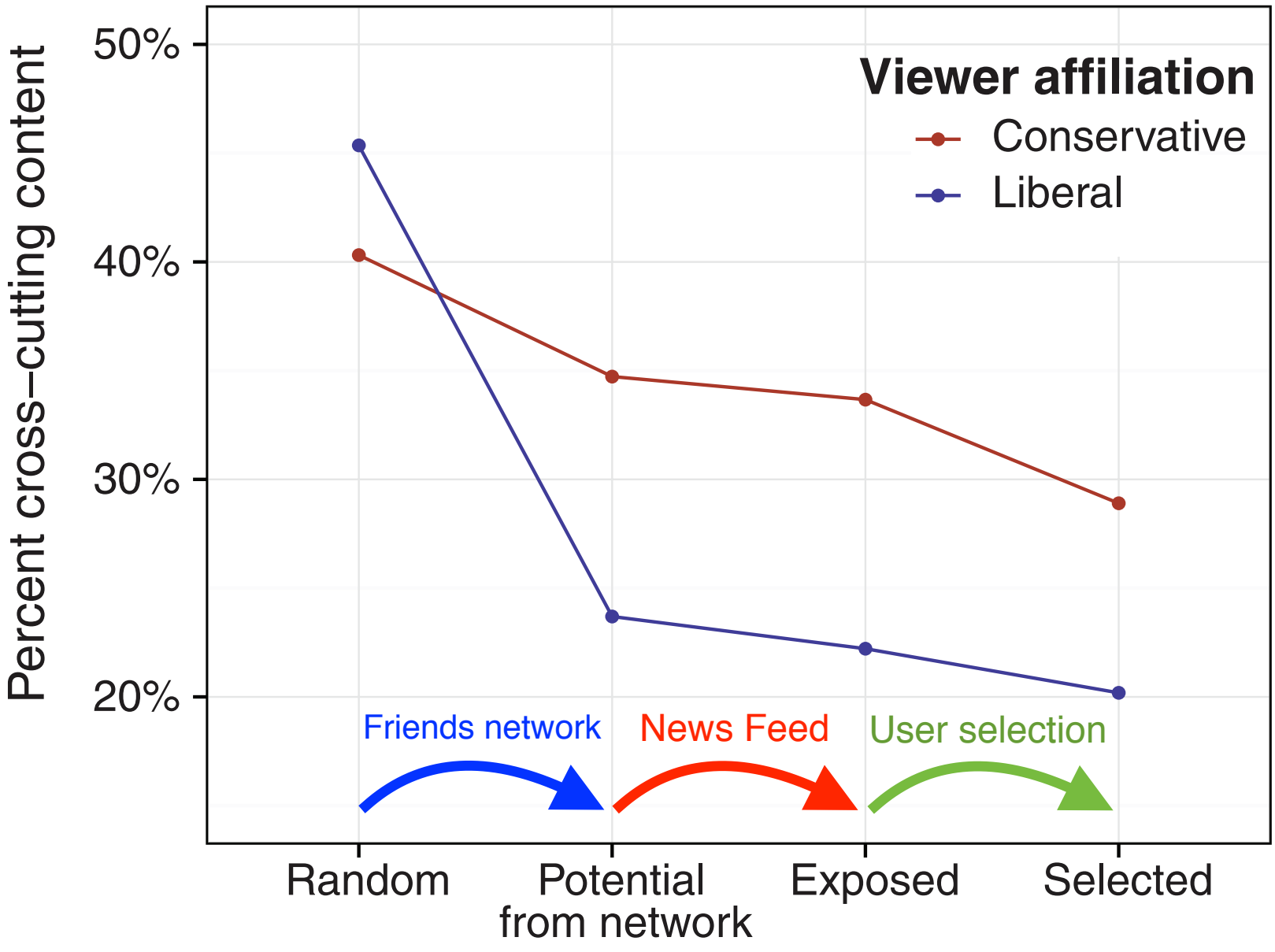


Figure 2: Exposure stages of news stories

1. *Potential from network*: all the content shared by friends
2. *Exposed*: content effectively shown in users' News Feeds
3. *Selected*: content clicked by the user

% cross-cutting content vs. exposure stage on Facebook



Facebook study: conclusions

1. The **friendship network** is the most important factor limiting the mix of content encountered in social media
   - if I have only friends of the same political affiliation, the filter bubble is obvious
2. The effect of **News Feed ranking** on cross-cutting content is limited:
   - -5% for liberals
   - -8% for conservatives
3. **Individual choice** influences the exposure to cross-cutting content more than the News Feed filtering

[. . .] we conclusively establish that on average in the context of Facebook, ***individual choices more than algorithms limit exposure to attitude-challenging content*** *(Bakshy, Messing, and Adamic 2015)*

Facebook study: criticism

Limitations of the study

- Underlying (false) assumption: the building of the **friendship network** is independent from Facebook's algorithms
  - Friends are only partly from "offline" connections
  - Facebook suggests both pages to like and new friends
- What about **sponsored content**?

Methodological issues

- **Sample** of the study: people which declare their political affiliation
  - may not be representative of the entire Facebook community
- Independent researchers can't access Facebook data and analyze it

Ranking = visibility

The **position** (rank) of a story in the News Feed is very important!

- the position in the News Feed may be used to promote some stories and not others
- money can buy rankings!
- even if the algorithm is "fair" *now*, what about the future?

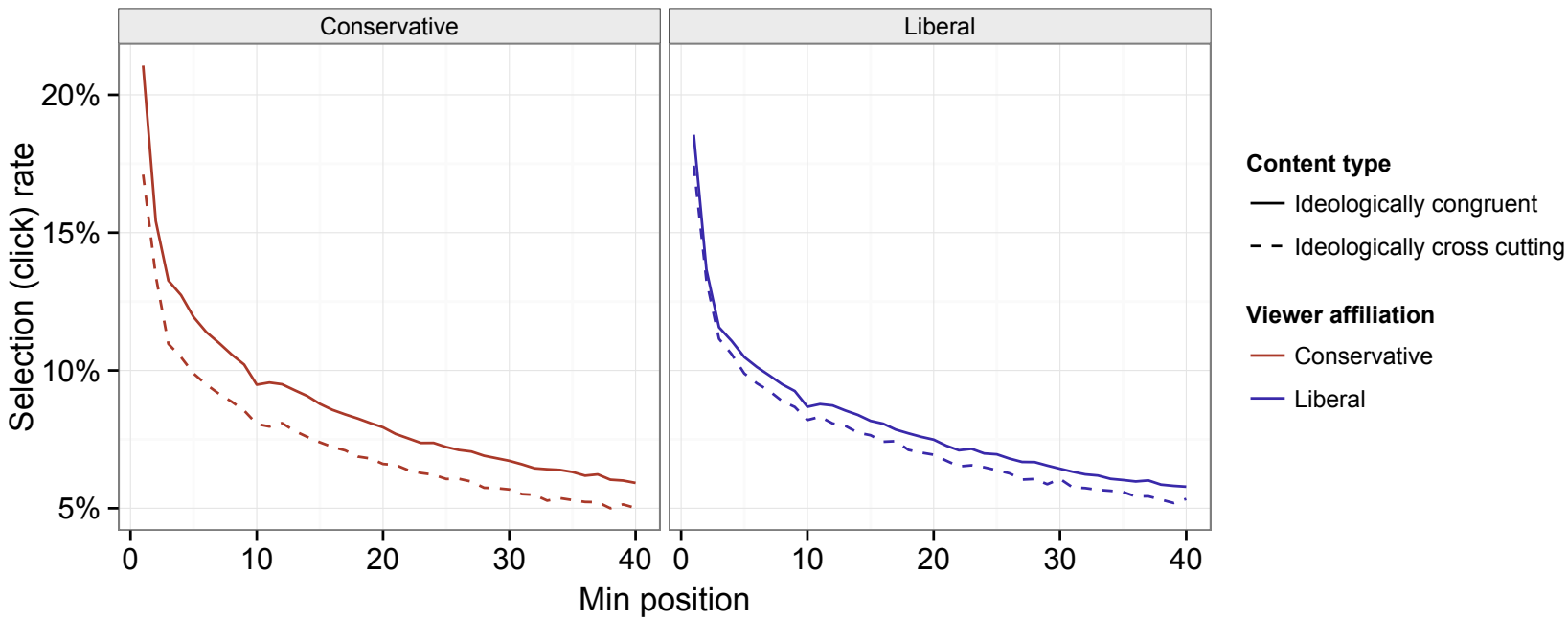


Figure 4: Click rate depends on the position of the story in the News Feed.

Proposed remedies and counter-objections

Moralizing filters

Problem: the Internet is showing off what we *want* to see, but not what we *need* to see

- Algorithms cannot compute "what should be seen" (Morozov 2011)
- What if one day Google could urge us to stop obsessing over Lady Gaga's videos and instead pay attention to Darfur?

Let's introduce "moralizing" filters!

- Would it be a good idea to make multinational companies **moralizing agents**?
  - Paternalistic, technocratic approach
- Active, educated citizens should be able to **autonomously search and retrieve information**
  - not just "ingest" whatever is thrown at them

Make the algorithms transparent

What if the **algorithms** and/or some of the **data** were **public**?

- The inner working of complex neural networks and machine learning agents is not intuitively understandable
  - Even if published, we may not understand those algorithms
- They are often **trade secrets**
- Knowing at least **which personal data** is used to make the recommendation may prove useful

Facebook News Feed settings

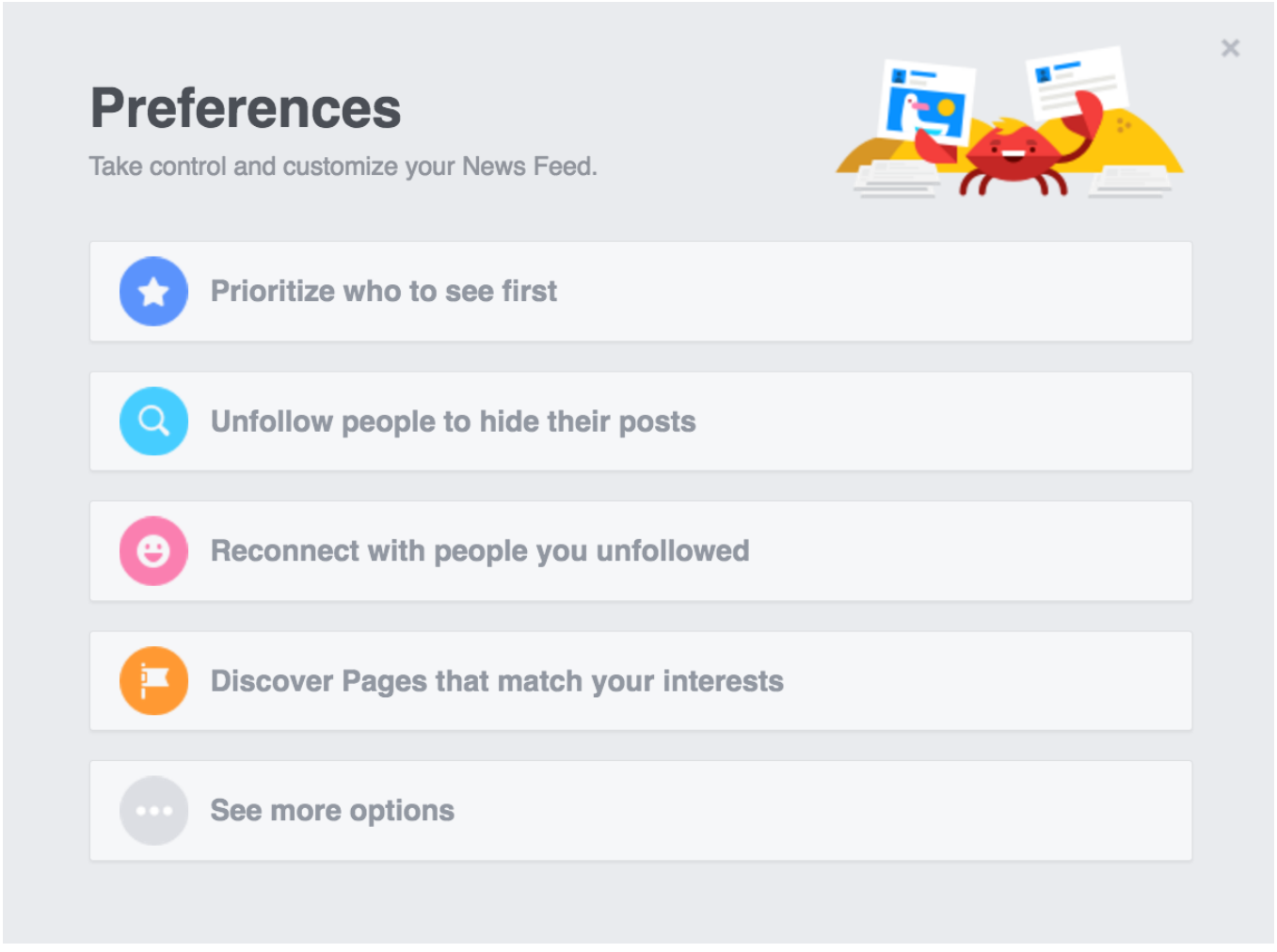


Figure 5: A rather good solution: Facebook lets users **see and customize** some parameters of the News Feed algorithm

Turn off the personalization!

- What if we could **turn off the personalization**?
- Personalization is the key feature of some services
  - Facebook without personalization would be... Twitter?
  - For other services, this would be a feasible solution
- Without personalization **ads would be less relevant** and profitable: no economic incentive to do so
- Users should at least know **whether** personalization is enabled or not

Conclusions

Conclusions

- **Personalization reduces information pluralism** by giving users only what they like to see
- Recommender systems privilege **relevance** over importance
- These technologies and their implementations are **not neutral**
- **Transparency** about the use of data and about the algorithms is needed
- Always use those services with a **critical eye**!

References I

Eytan Bakshy, Solomon Messing, and Lada A Adamic. "Exposure to ideologically diverse news and opinion on Facebook". In: *Science* 348.6239 (2015).

CENSIS. *13° Rapporto Censis-Ucsi sulla comunicazione*. http://www.censis.it/7?shadow_comunicato_stampa=121073. 2016.

Matthew Haughn. *"filter bubble" - WhatIs.com*. URL: http://whatis.techtarget.com/definition/filter-bubble (visited on 2015).

Katholieke Universitet Leuven et al. "Independent study on indicators for media pluralism in the member states—Towards a risk-based approach". In: *Prepared for the European Commission Directorate-General Information Society and Media.* 19 (2009).

Evgeny Morozov. "Your Own Facts". In: *New York Times Sunday Book Review* (2011).

Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.