The behavioural science of online harm and manipulation and what to do about it

Making Sense of Media - 18 July 2019
Aisling Ní Chonaire

How do our biases manifest online?

""It turns out that the environmental effects on behavior are a lot stronger than most people expect"* Daniel Kahneman*

Three themes for today

1. Leveraging data and AI on the side of the consumer

2. Using reputation as a regulatory tool

3. Building the evidence base
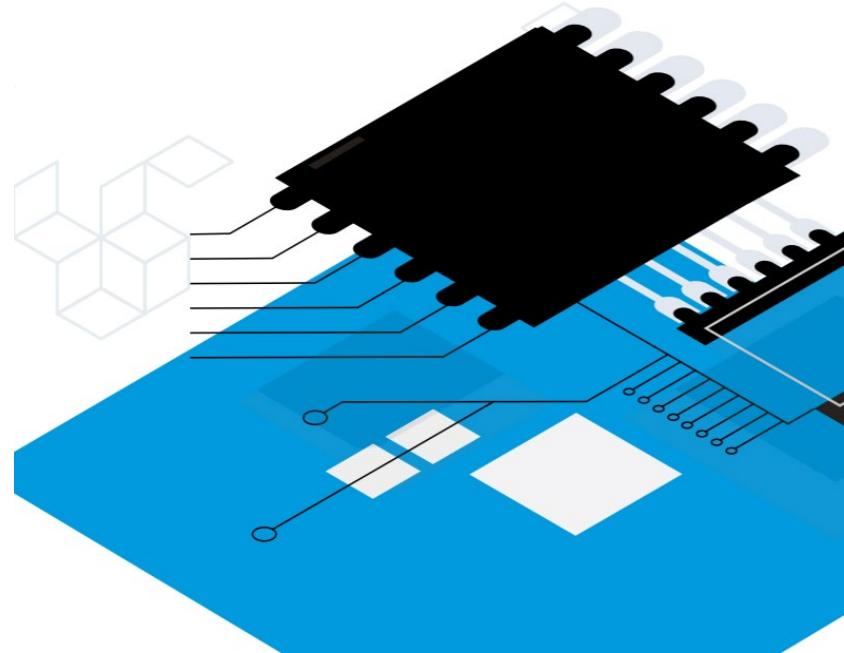
Theme 1. Leveraging data and AI on the side of the consumer

Terms are not set up for people to understand

Opinion | THE PRIVACY PROJECT

We Read 150 Privacy Policies. They Were an Incomprehensible Disaster.

By Kevin Litman-Navarro

In the background here are several privacy policies from major tech and media platforms. Like most privacy policies, they're verbose and full of legal jargon — and opaquely establish companies' justifications for collecting and selling your data. The data market has become the engine of the internet, and these privacy policies we agree to but don't fully understand help fuel it.
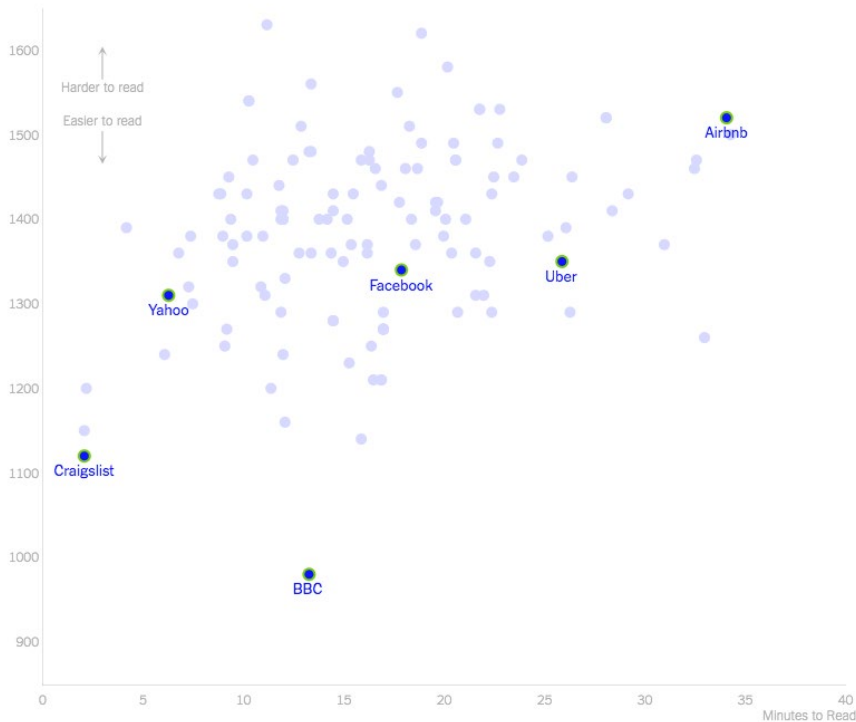
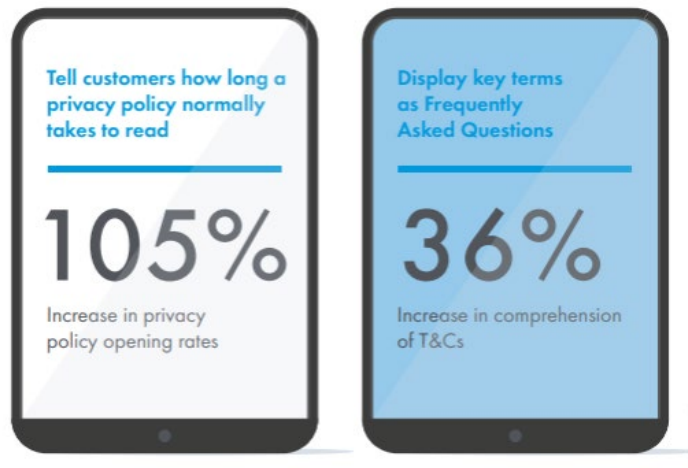# Intermediaries that help consumers navigate

Figure 5: Improving engagement with and comprehension of terms and conditions and data privacy notices

Tell customers how long a privacy policy normally takes to read

105%

Increase in privacy policy opening rates

Display key terms as Frequently Asked Questions

36%

Increase in comprehension of T&Cs

Terms of Service Didn't Read

"I have read and agree to the Terms" is the biggest lie on the web. We aim to fix that.

Giving back individual control

"**Nearly half of the 12 million** UK consumers with mental health problems surveyed by the Money and Mental Health Policy Institute stated **they would like to set controls** like spending limits in online shops."
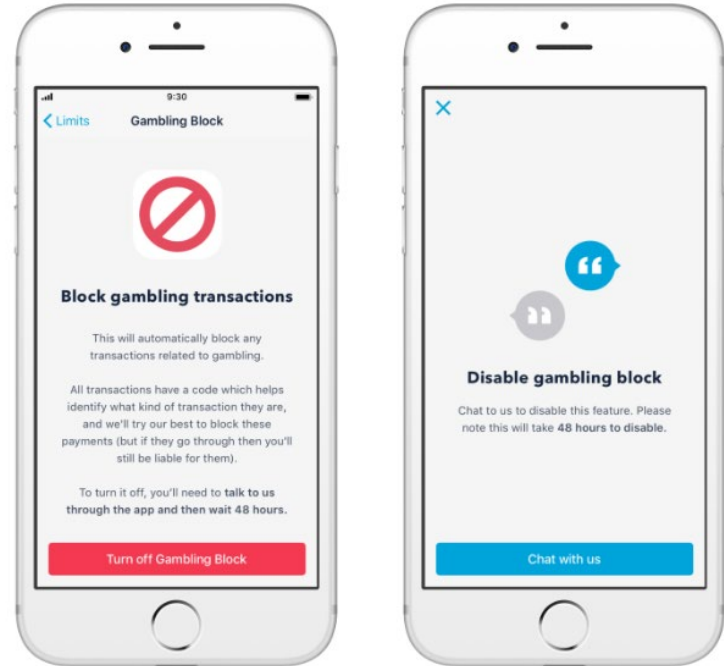
Prompting reflection in the heat of the moment

Instagram now asks bullies: 'Are you sure?'

Dave Lee
North America technology reporter

🕐 8 July 2019

f 💬 🐦 ✉️ ◁ Share

francescofogu Amazing
1w   Reply   ♡

divdivk You are so ugly and stupid   Undo
Posting...

Are you sure you want to post this? Learn More

👏 🤔 😍 😮 😂 🤪 😎 🤩

INSTAGRAM

"Instagram believes its new anti-bullying tool, which prompts users to **pause and consider what they are saying**, could help curb abuse on the platform.

It will also soon offer the targets of bullying the ability to **restrict interactions** with users who are causing them distress."

Theme 2. Using reputation as a regulatory tool

Levers that target reputation

Changes We're Making to Do More to Support and Protect the Most Vulnerable People who Use Instagram

Posted on February 07, 2019   0 comments

TECH

Facebook has struggled to hire talent since the Cambridge Analytica scandal, according to recruiters who worked there

PUBLISHED THU, MAY 16 2019 · 2:58 PM EDT | UPDATED FRI, MAY 17 2019 · 1:22 PM EDT

- Publicly urge specific companies to improve, and signal willingness to act

- Harness the power of transparency reports by leveraging key stakeholders to act as messengers
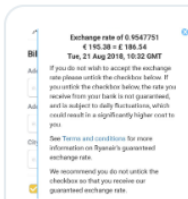
Calling out 'Sludge'

"Sludge can take two forms. It can discourage behavior that is in a person's best interest such as claiming a rebate or tax credit, and it can encourage self-defeating behavior such as investing in a deal that is too good to be true." - Richard Thaler
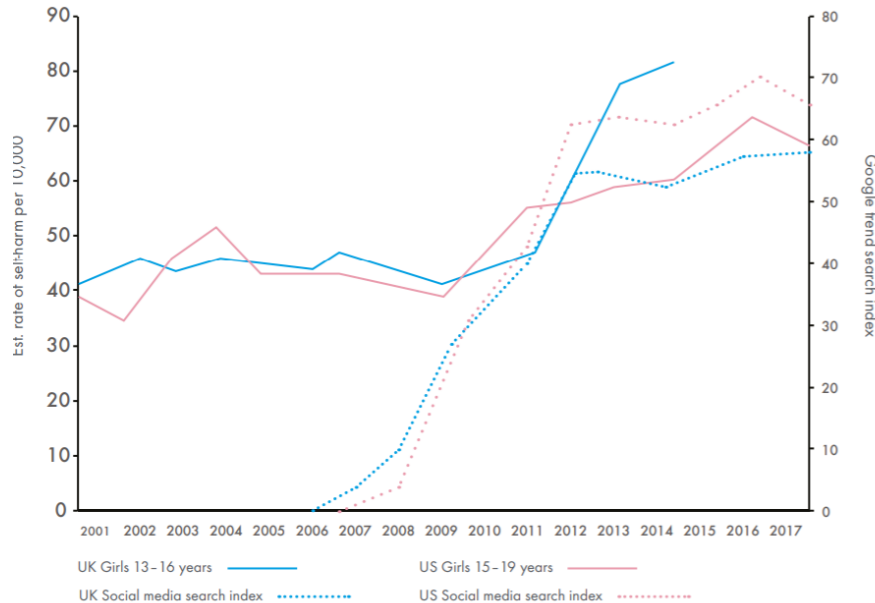
Theme 3. Building the evidence base

We need to improve the evidence base

Figure 3: Self harm among teenage girls and social media use

The Welfare Effects of Social Media

Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow*
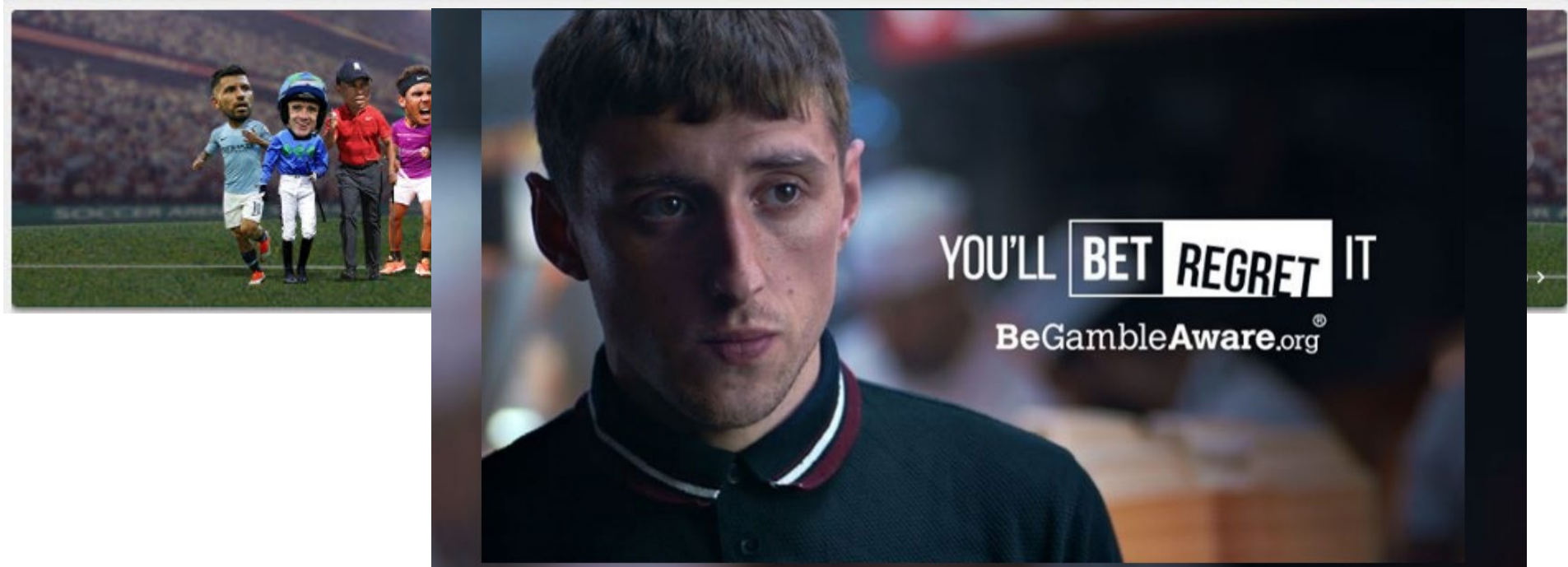
April 1, 2019

Abstract

The rise of social media has provoked both optimism about potential societal benefits and concern about harms such as addiction, depression, and political polarization. We present a randomized evaluation of the welfare effects of Facebook, focusing on US users in the run-up to the 2018 midterm election. We measured the willingness-to-accept of 2,743 Facebook users to deactivate their Facebook accounts for four weeks, then randomly assigned a subset to actually do so in a way that we verified. Using a suite of outcomes from both surveys and direct measurement, we show that Facebook deactivation (i) reduced online activity, including other social media, while increasing offline activities such as watching TV alone and socializing with family and friends; (ii) reduced both factual news knowledge and political polarization; (iii) increased subjective well-being; and (iv) caused a large persistent reduction in Facebook use after the experiment. Deactivation reduced post-experiment valuations of Facebook, but valuations still imply that Facebook generates substantial consumer surplus.

JEL Codes: D12, D90, I31, L86, O33.

Keywords: Social media, political polarization, subjective well-being, consumer surplus from digital technologies.

We need to measure the right metrics

Thank you

aisling@bi.team