

The Microtargeting Manipulation Machine

Stephan Lewandowsky (University of Bristol)

Fabio Carrella (University of Bristol)

Matthew Edwards (University of Bristol)

Almog Simchon (Ben-Gurion University of the Negev)

Advertising: Audience Segmentation vs. Microtargeting

- The best ad is the one you only show to people who then buy your product
- Showing ads to millions who do not need your product is wasteful and drives up consumer prices
- Audience segmentation
 - few cosmetic ads in motorcycle magazines
 - few toys advertised on TV after bedtime

But there is a darker side to audience segmentation and targeting

Targeting the Vulnerable With Pseudoscience

- Advertisers can use Google keywords to match ads to the searches of Google users
- “Alternative” cancer clinics in Mexico and Arizona targeted the sensitive queries of cancer patients ...
- ... thus giving prominence to pseudoscientific treatment options

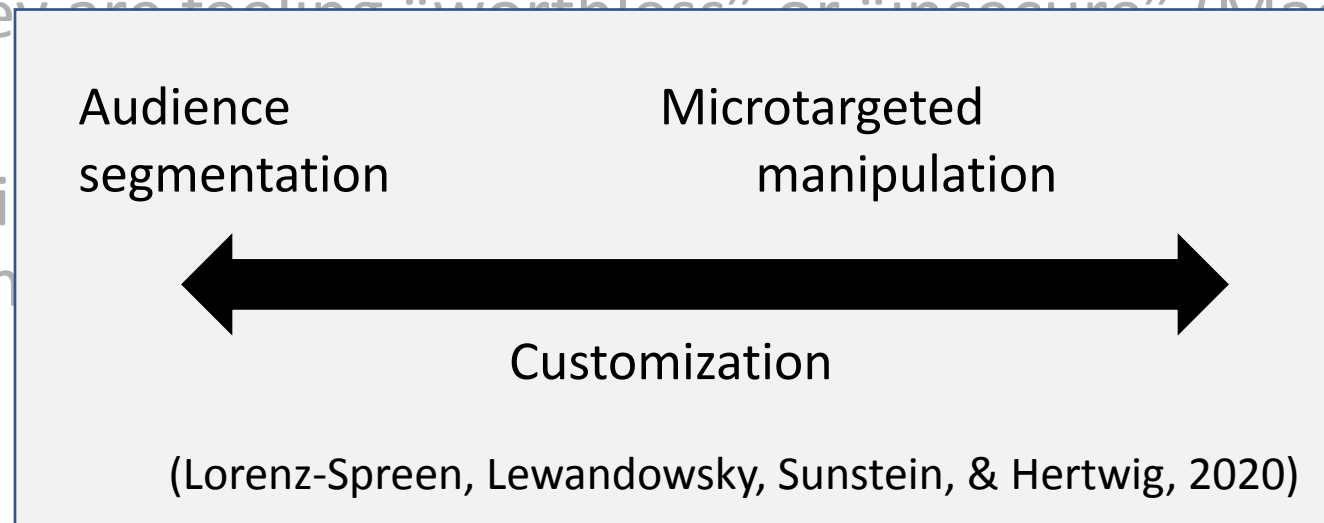
Names of world-leading cancer centers used as keywords – thus boosting visibility of “alternative” treatments ahead of evidence-based cancer centers

Microtargeted Manipulation

- Facebook's algorithms can determine, and allow advertisers to pinpoint, "moments when young people need a confidence boost" because they are feeling "worthless" or "insecure" (Machkovetch, 2017).
- Facebook did not dispute the existence of the technology, but claimed that it was never made available to advertisers.

Microtargeted Manipulation

- Facebook's algorithms can determine, and allow advertisers to pinpoint, "moments when young people need a confidence boost" because they are feeling "worthless" or "insecure" (Moshkovetch, 2017).
- Facebook did not deny the use of microtargeting, but claimed that it was not manipulation.

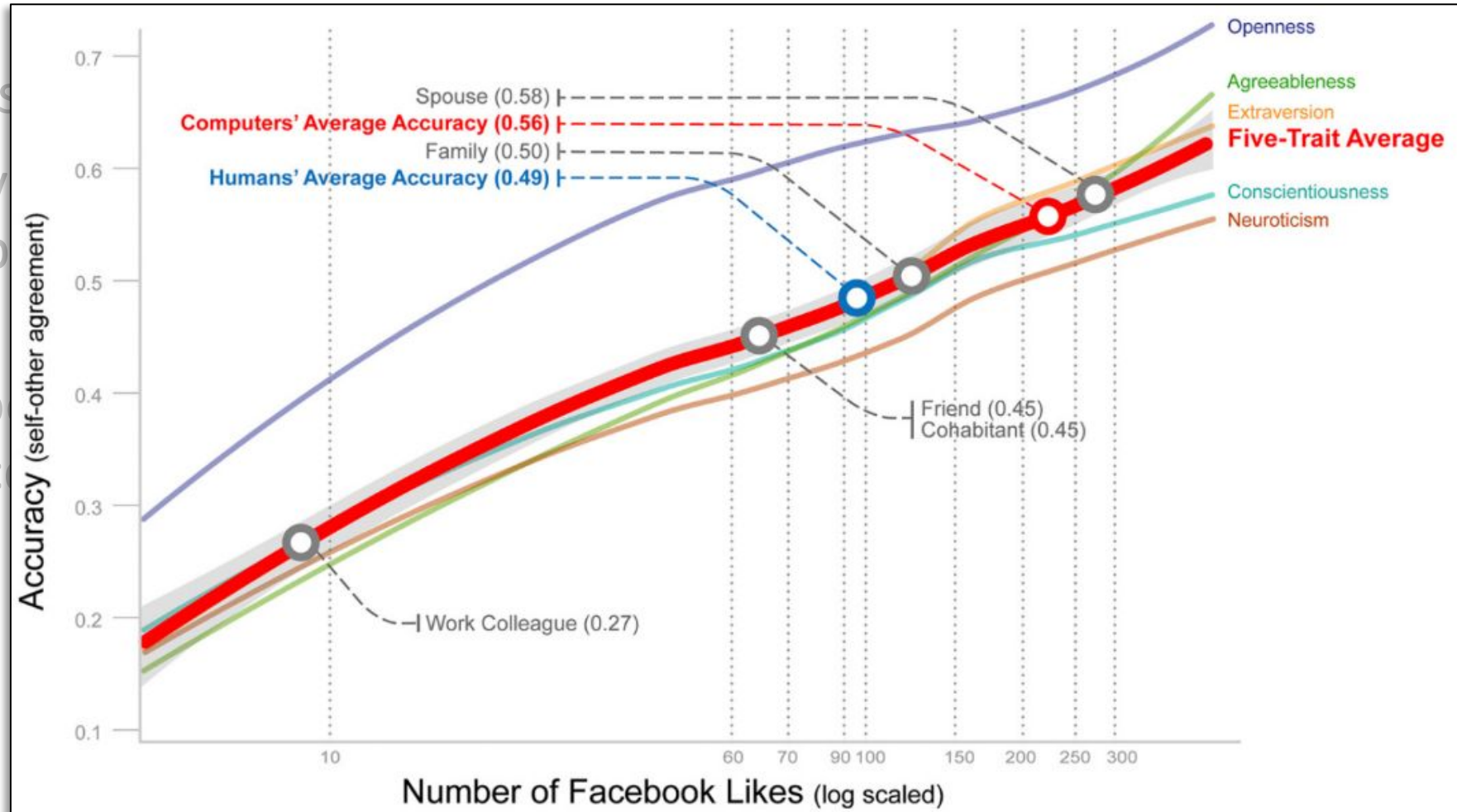


Psychological Microtargeting Raises Several Concerns (I.)

- Individuals rarely consent to personal data use
- Personal vulnerabilities can be exploited because users can be targeted based on their personality which can be inferred from their “likes”
- 300 Facebook “likes” permit identification of one’s personality with greater accuracy than one’s spouse (Youyou et al., 2015)

Psychological Microtargeting Raises Several Concerns (I.)

- Individuals
- Personal v targeted b “likes”
- 300 Faceb greater ac



Psychological Microtargeting Raises Several Concerns (II.)

- Intent and true nature of messages concealed
- No rebuttal of political (mis-)information is possible
- Allows politicians to make mutually incompatible promises to different segments of the electorate



FDP

Gesponsert • Finanziert von FDP

Wir brauchen mehr Klimaschutz und müssen die #Klimaschutzziele einhalten. Wie wir das erreichen wollen? Mit einem einzigen harten Mittel: Einem klaren CO2-Limit!



FDP

Gesponsert • Finanziert von FDP

Wie meistern wir die großen Herausforderungen wie den Klimawandel oder die Corona-Pandemie? Bei unseren politischen Mitbewerbern ist häufig die Rede von restriktiven staatlichen Maßnahmen, Freiheitseinschränkungen und Verboten. Wir sehen das anders: Natürlich muss und wird sich vieles verändern. Nutzen wir also unsere Kreativität und unseren Erfindergeist, haben wir den Mut für innovative Ideen. 🚀 ...



CHRISTIAN LINDNER

FÜR MEHR FREUDE
AM ERFINDEN ALS AM
VERBIETEN.

Freie
Demokraten
FDP



FDP

Gesponsert • Finanziert von FDP

We need more climate protection and must meet the #climateprotectiongoals. How do we want to achieve this? With one hard measure: a clear CO2 limit!



CHRISTIAN LINDNER:

**Organic food
WWF
Climate protection**



FDP

Gesponsert • Finanziert von FDP

How do we master the big challenges like climate change or the Corona pandemic? Among our political competitors, there is often talk of restrictive government measures, restrictions on freedom and bans. We see it differently: Of course, many things must and will change. So let's use our creativity and inventive spirit, let's have the courage for innovative ideas.

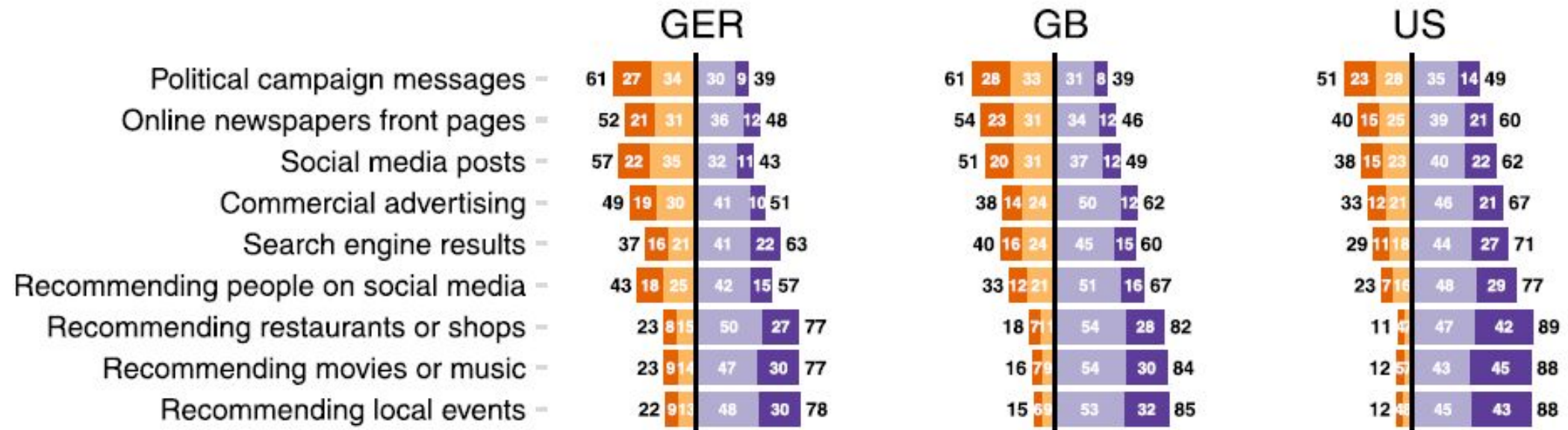


Frequent traveler

Psychological Microtargeting And The Public

- Public rejects political microtargeting in U.S., U.K., and Germany (Kozyreva, Lorenz-Spreen, Hertwig, Lewandowsky, & Herzog, 2021)

Acceptability of personalizing a service



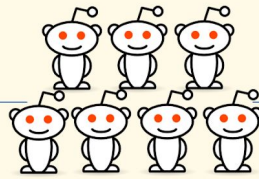
Three Questions

1. Can we detect microtargeting?
2. Is it effective?
3. Can we defang it?

1. Detection (Simchon et al. 2023)

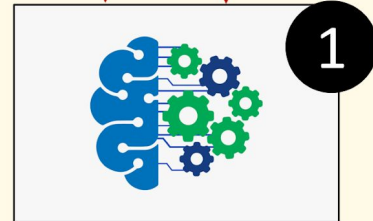
Model 1

Sample of Reddit participants



Personality inventory

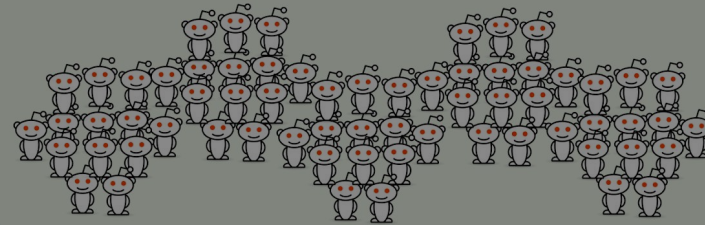
Comments (text)



Machine learning

Model 2

Sample of Reddit users

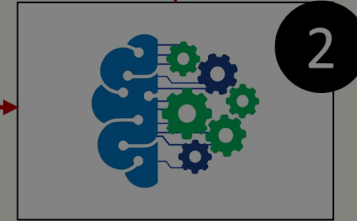


Comments (text)

Posts (text)

1

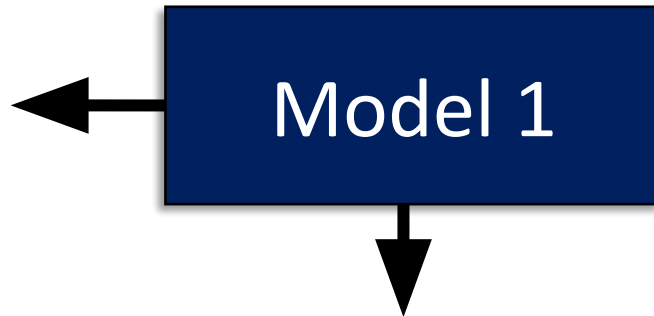
Predicted personality



2

Model 1: Method

- Reddit collection
 - Communities of fiction writing
 - 32,344 Potential Participants
 - 9,244 received personalized invitations to participate
 - 1,100 participants
 - 650,000 comments
- Measures
 - “Big 5” personality: BFI-2 (Soto & John, 2017; *JPSP*)
 - Values: SVS-PVQ (Schwartz, 1992; 2012)



Detection (Simchon et al. 2023)

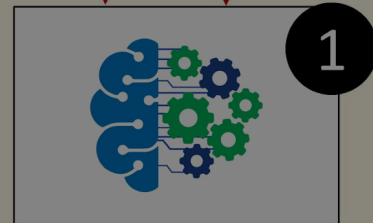
Model 1

Sample of Reddit participants



Personality inventory

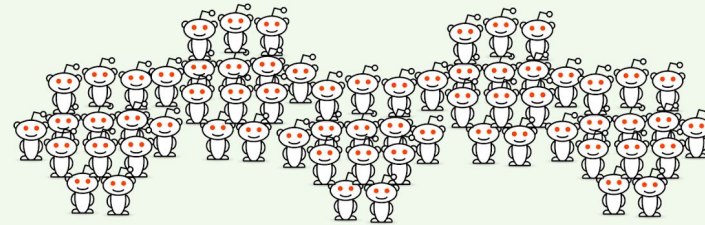
Comments (text)



Machine learning

Model 2

Sample of Reddit users

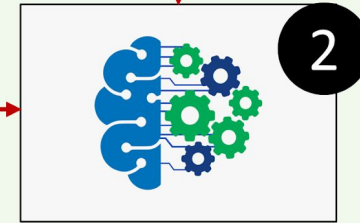


Comments (text)

Posts (text)

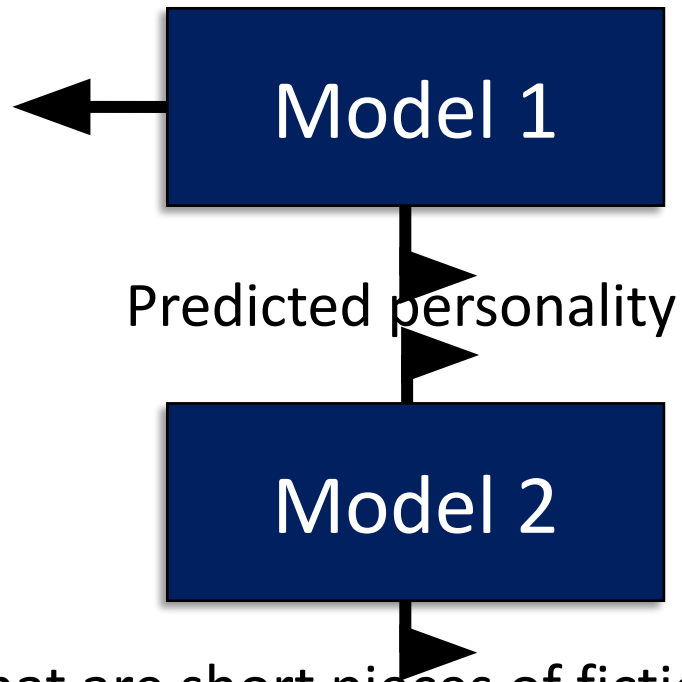
1

Predicted personality



Model 2: Method

- Reddit collection
 - communities of fiction writing
 - 10,050 users
 - 2.5 million comments



- 10,050 unique posts that are short pieces of fiction (“consumed content”)

Detection (Simchon et al. 2023)

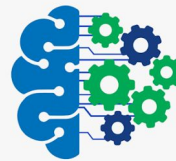
Model 1

Sample of Reddit participants



Comments (text)

Pearson's r
(out of sample test):
.26 – .39



1

Machine learning

Model 2

Sample of Reddit users



- Pearson's r : .08 – .19
- Focus on openness

- Model 2 more sensitive to “how” rather than “what”
- Style rather than content

1

Predicted personality



2

Effect Size?

- Isn't an effect size of $r = 0.1$ negligible?
 - yes, for a particular event, but not at the aggregate or at scale
- Real world examples:
 - the effect of antihistamines on runny nose and sneezing: $r = 0.11$
 - the effect of ibuprofen on pain relief: $r = .14$
 - the correlation between extraversion and money spent on holiday shopping: $r = .09$

(Funder & Ozer, 2019)

Examples (from 1,500 Facebook ads)

Item 5: lowest on Openness



UK government

Sponsored

Avoid travelling unless essential for work, education or other permitted exemption



Item 22: highest on Openness



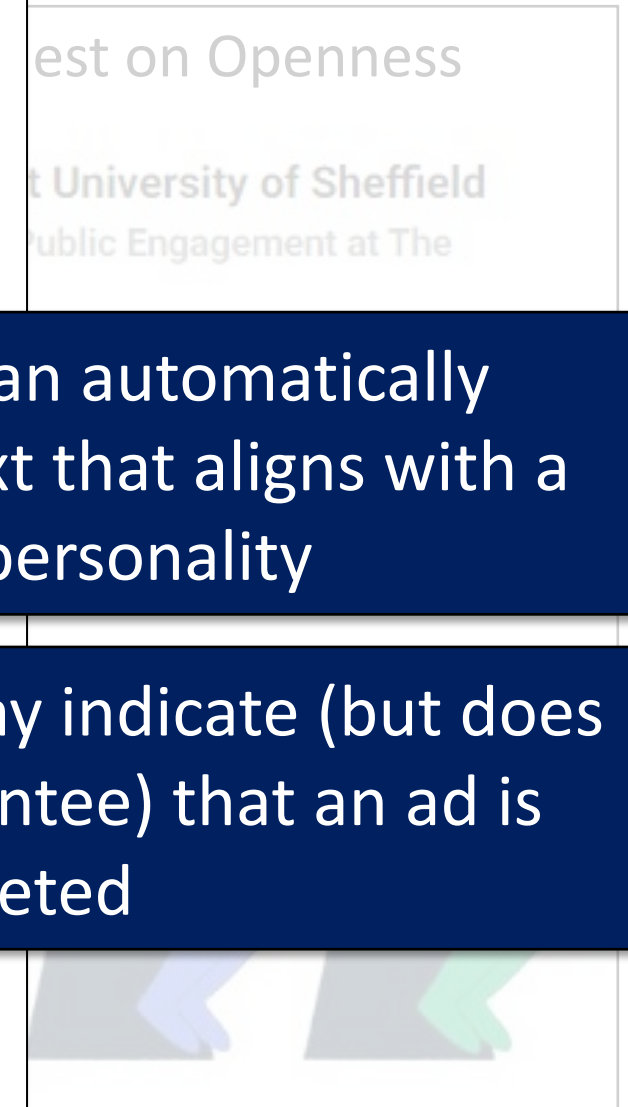
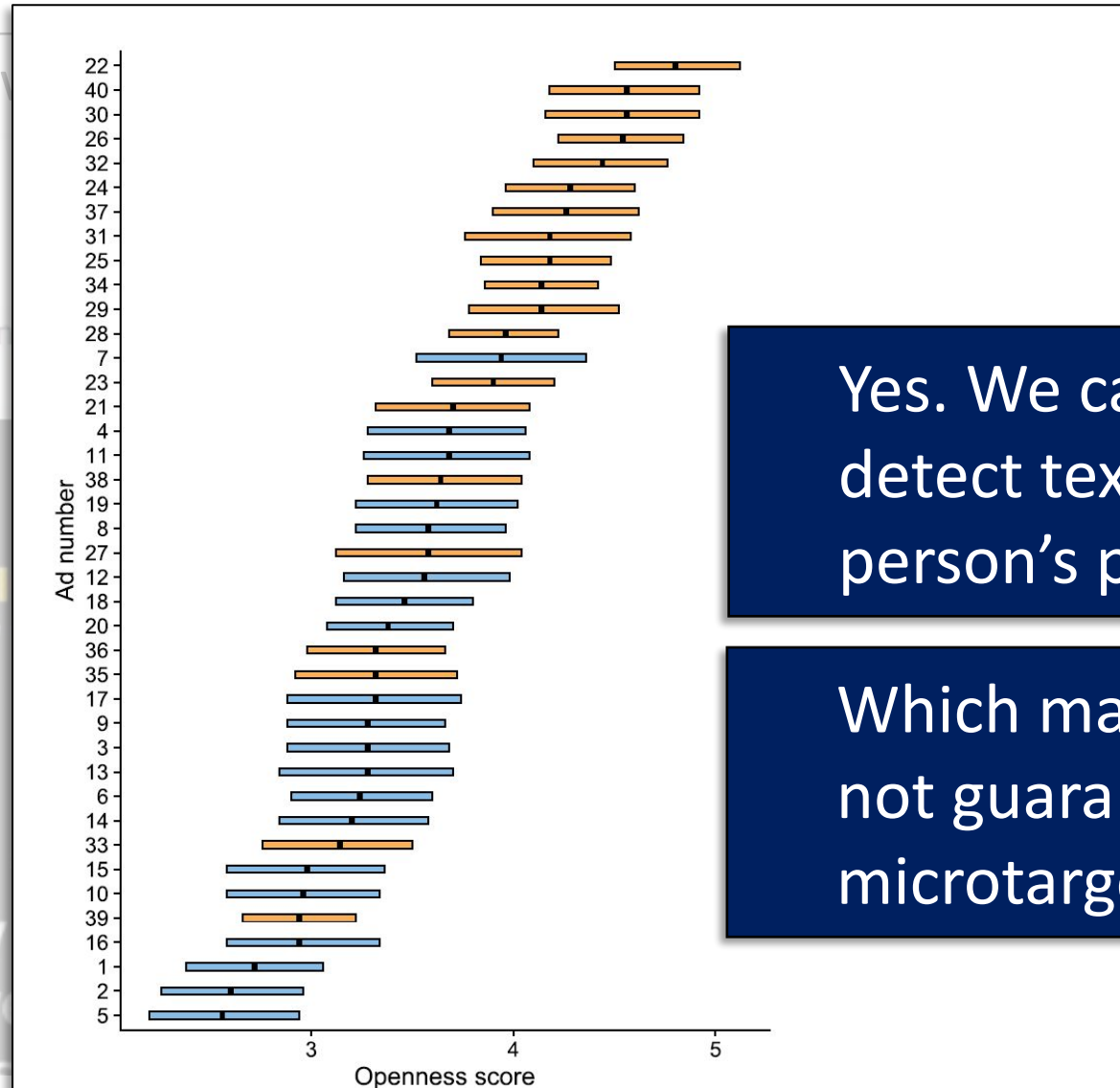
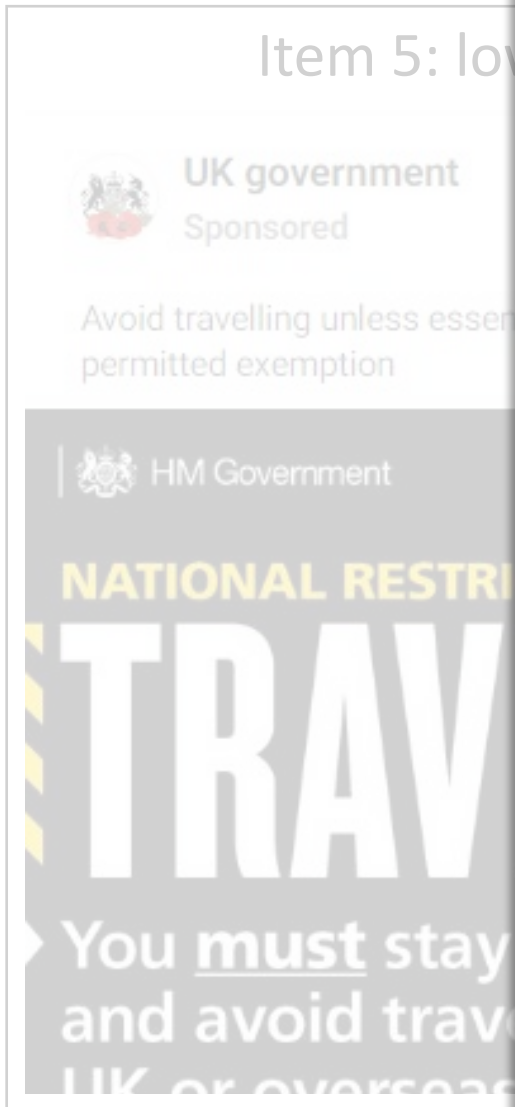
Public Engagement at University of Sheffield

Sponsored • Paid for by Public Engagement at The University of Sheffield

Check out Dr Susan Oman's visually stunning digital exhibition exploring the concept of well-being data and its power and potential to shape society.



Examples: Human Validation



2. Is Microtargeting Persuasive? (Simchon et al. 2024)

- Four preregistered studies
- Collected Openness scores on BFI-2
- Asked for persuasiveness ratings for top 10 ads

I find this ad to be persuasive

This is an effective ad

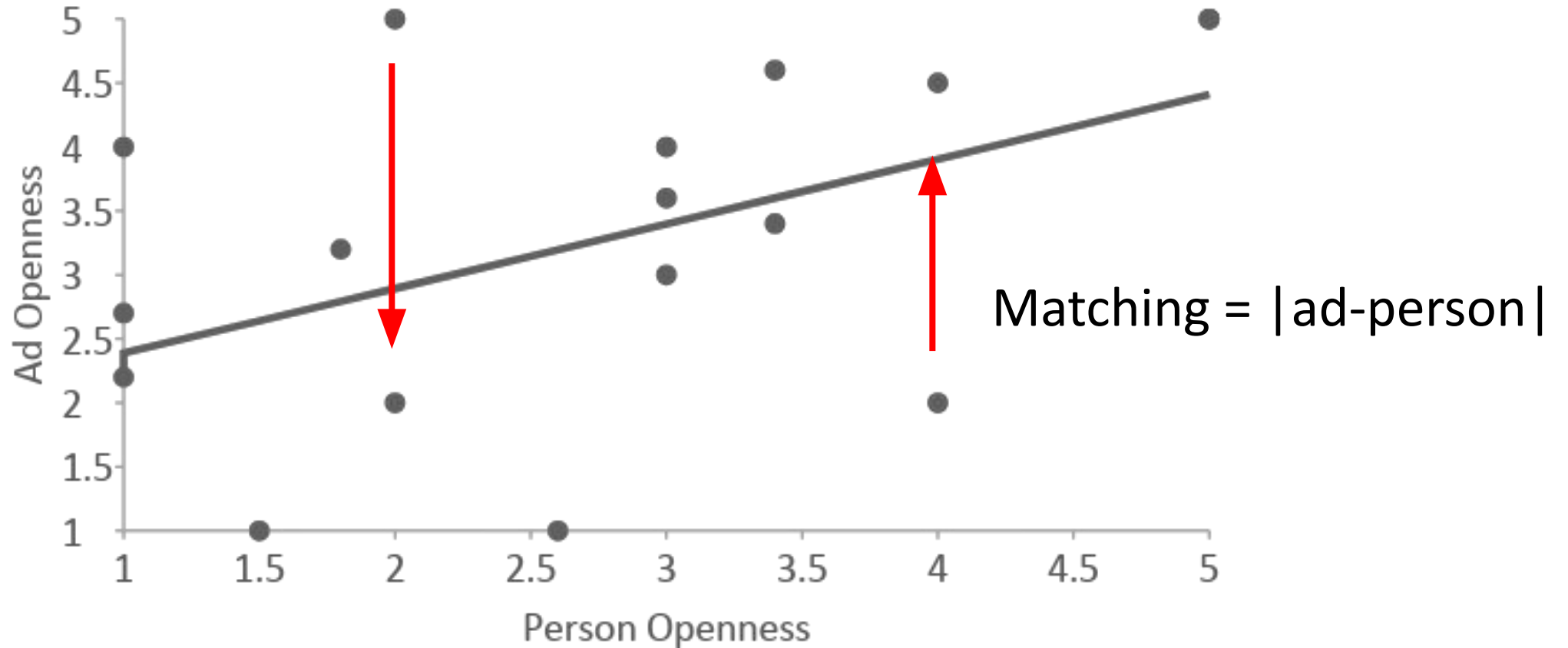
I would click on this link after seeing this ad

Overall, I like this ad

This ad has made me more interested in the topic

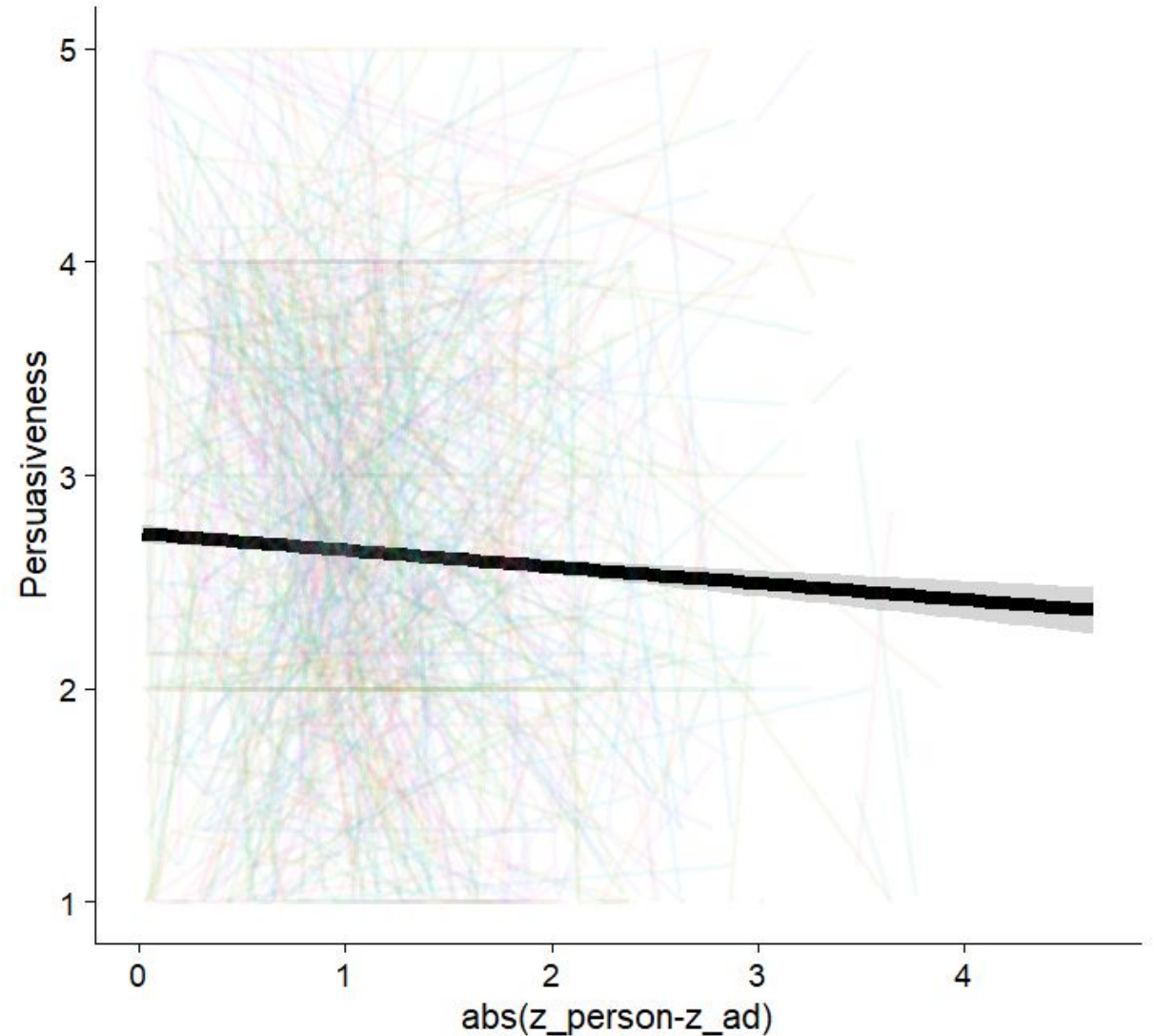
I am interested in learning more about this topic after seeing this ad

Analytical Procedure



Study 1: Are Targeted Ads Persuasive? Yes.

- Persuasiveness as a function of discrepancy between “personality” of ad and personality of person



Studies 2 and 3: AI Automation (GPT3 & ChatGPT)

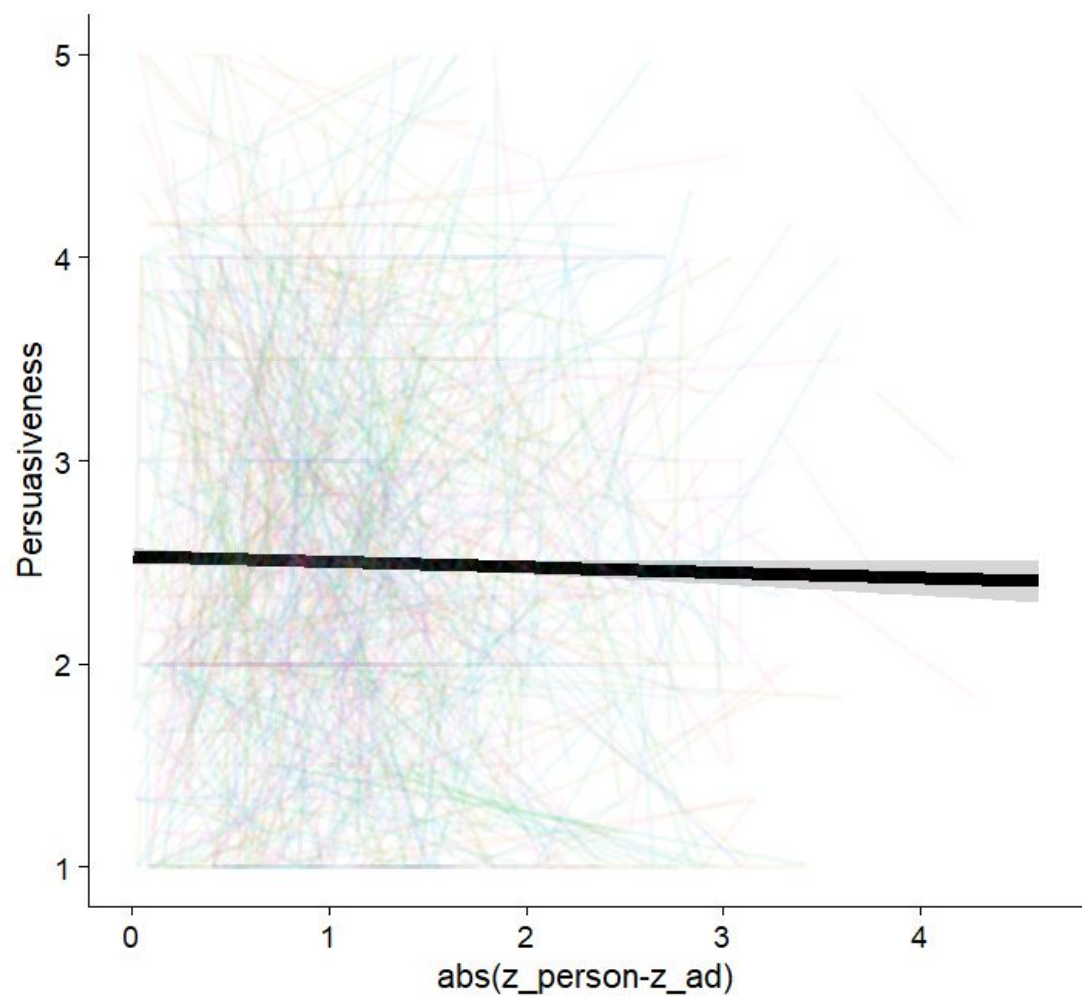
High Openness: now! Experience the extraordinary and join the global movement for universal access to vaccines! Sign up now and help make sure everyone, everywhere can benefit from the power of

Original: Vaccines should be available to everyone, everywhere. Tell Boris Johnson to take action.

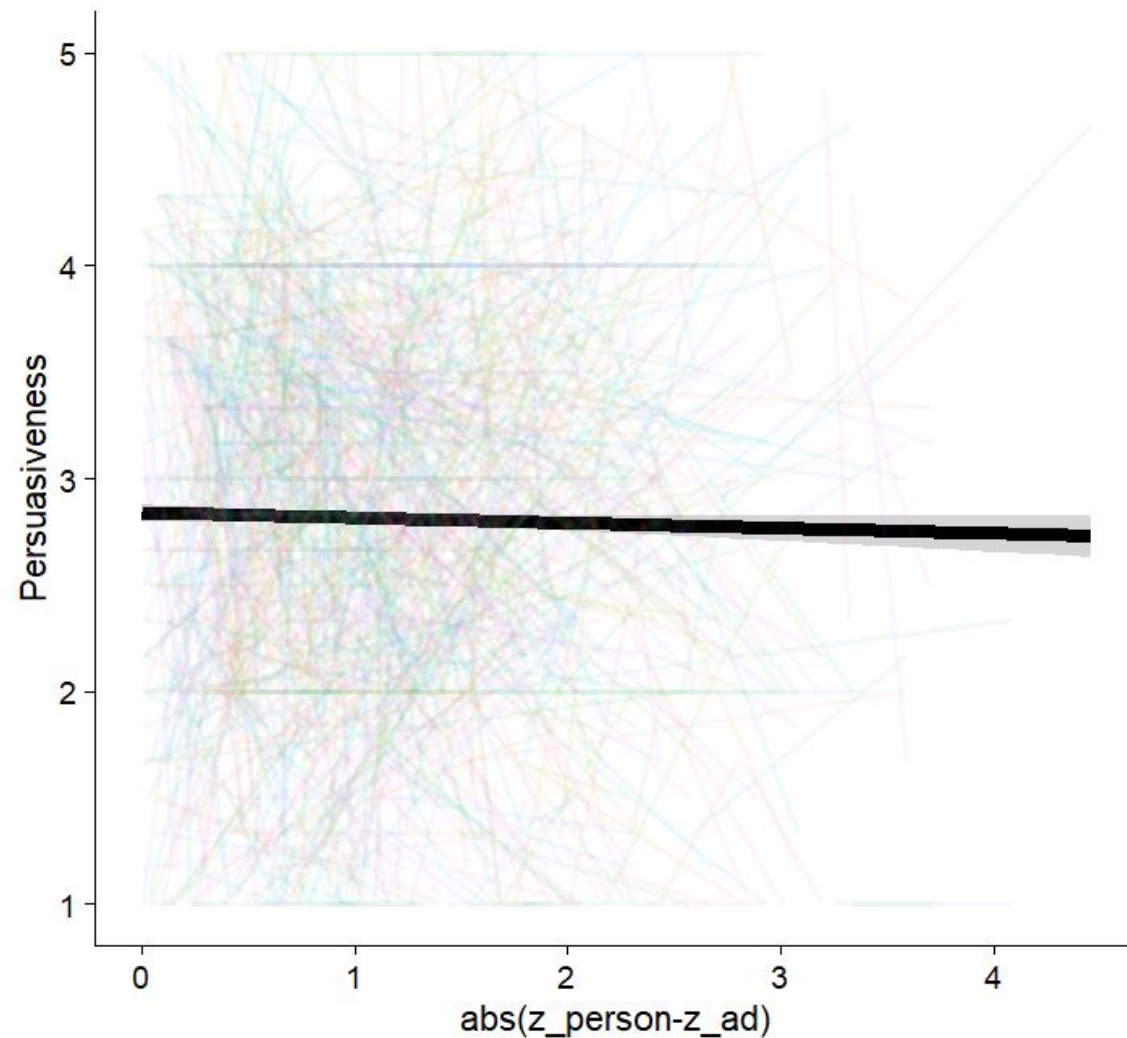
Low Openness: Protect yourself and your family. Get your vaccines and stay safe. Take the traditional approach and join the fight against disease. Tell Boris Johnson to take action now!

Studies 2 and 3: Results

GPT-3



ChatGPT



But How Would This Work in Practice? Advertising on Facebook

- Advertisers can select a highly-specific target audience based on many attributes

Location

Advertise in the cities, communities and countries where you want to do business.

Demographics

Choose your audience based on age, gender, education, job title and more. You can keep track of the types of people your ads are reaching, but Facebook will never share personally-identifiable information about them.

Interests

Add interests and hobbies of the people you want your ad to reach – from organic food to action films – and make your targeted ads more relevant.

behaviour

Target your ads based on consumer behaviours such as prior purchases and device usage.

“Interests” are a little more pernicious than simply being interests

(12) **United States Patent**
Nowak et al.

(10) **Patent No.:** **US 8,825,764 B2**
(45) **Date of Patent:** **Sep. 2, 2014**

(54) **DETERMINING USER PERSONALITY
CHARACTERISTICS FROM SOCIAL
NETWORKING SYSTEM
COMMUNICATIONS AND
CHARACTERISTICS**

(75) Inventors: **Michael Nowak**, San Francisco, CA
(US); **Dean Eckles**, Palo Alto, CA (US)

(73) Assignee: **Facebook, Inc.**, Menlo Park, CA (US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 185 days.

(21) Appl. No.: **13/608,943**

(22) Filed: **Sep. 10, 2012**

(65) **Prior Publication Data**
US 2014/0074920 A1 Mar. 13, 2014

(51) **Int. Cl.**
G06F 15/16 (2006.01)

(52) **U.S. Cl.**
CPC **G06F 15/16** (2013.01)
USPC **709/204**

(58) **Field of Classification Search**
CPC ... G06Q 30/00; G06Q 30/0241; G06Q 50/01;
G06Q 10/10; G06F 17/30867; G06N 5/043
USPC 709/204
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2011/0055098	A1 *	3/2011	Stewart	705/321
2012/0102050	A1 *	4/2012	Button et al.	707/749
2012/0166520	A1 *	6/2012	Lindsay et al.	709/203
2012/0254333	A1 *	10/2012	Chandramouli et al.	709/206
2013/0013667	A1 *	1/2013	Serena	709/203
2013/0174055	A1 *	7/2013	Johnson et al.	715/753

FOREIGN PATENT DOCUMENTS

WO WO 2011000046 A1 * 1/2011
OTHER PUBLICATIONS

Mislove, A. et al., "You Are Who You Know: Inferring User Profiles
in Online Social Networks," *Proceedings of the Third ACM Interna-
tional Conference on Web Search and Data Mining*, Feb. 4-6, 2010,
pp. 251-260, New York, New York, USA.

* cited by examiner

Primary Examiner — Kostas Katsikis
(74) *Attorney, Agent, or Firm* — Fenwick & West LLP

(57) **ABSTRACT**

A social networking system obtains linguistic data from a
user's text communications on the social networking system.
For example, occurrences of words in various types of com-
munications by the user in the social networking system are
determined. The linguistic data and non-linguistic data asso-
ciated with the user are used in a trained model to predict one
or more personality characteristics for the user. The inferred
personality characteristics are stored in connection with the
user's profile, and may be used for targeting, ranking, select-
ing versions of products, and various other purposes.

14 Claims, 2 Drawing Sheets

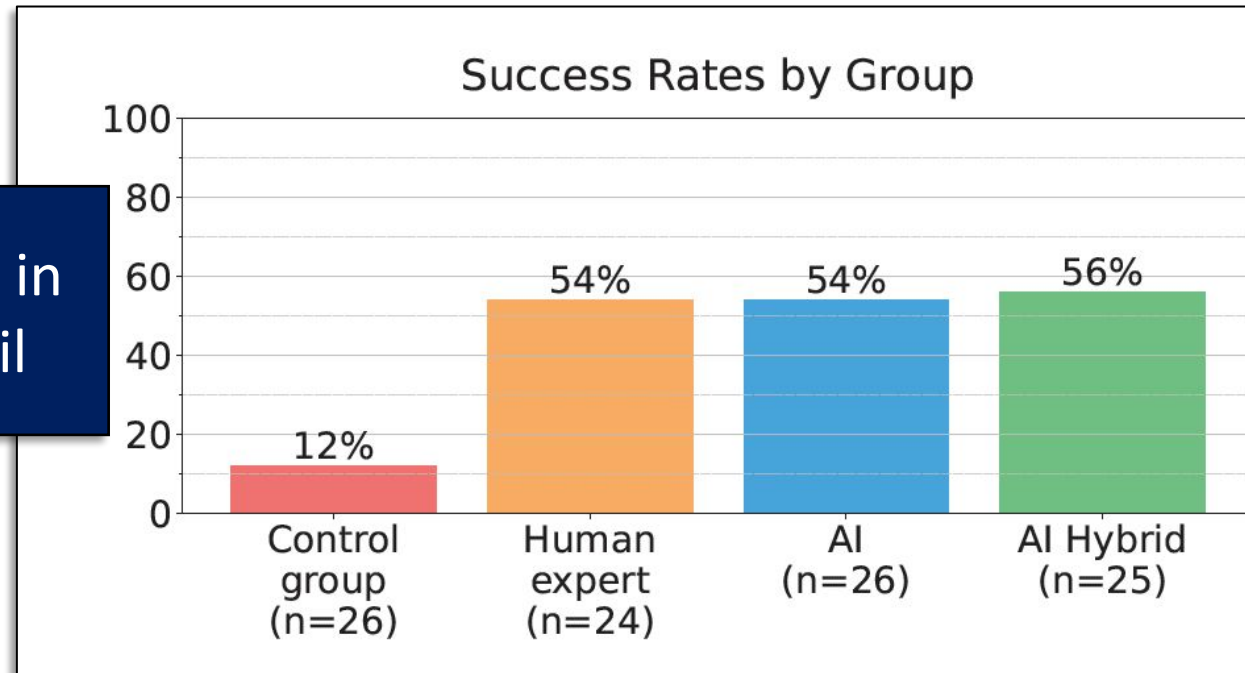
A person is shown from the chest up, wearing a dark, futuristic head-mounted device. The device has several glowing lights: a green one on the left side, a red one on the right side, and a purple one in the center. A small yellow screen on the front of the device displays a blue silhouette of a person. Numerous colorful wires (red, blue, green, yellow) extend from the top and sides of the device. The person is holding a handgun in their right hand, pointing it downwards. The background is a plain, light-colored wall.

Yes, persuasiveness increases with
psychological targeting of
AI-tailored ads

AI-tailored ads can be rolled out
at scale

Not An Isolated Finding: LLM-Driven Spear Phishing

% clicks on link in phishing email



- AI performed OSINT (web crawl) to create profile of target
- AI crafted customized emails on basis of each target's profile

Regulatory Context



Part of Chapter IV: Transparency Obligations for Providers and Deployers of Certain AI Systems and GPAI Models

Article 50: Transparency Obligations for Providers and Users of Certain AI Systems and GPAI Models

Date of entry into force: July 2026
According to: Article 113

Note: This date is an estimate based on the expected date of entry into force. See [here](#) for a full implementation timeline.

SUMMARY +

1. Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate or prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, unless those systems are available for the public to report a criminal offence.

3. Can We Defang Microtargeting?

The “Popup” Intervention: Studies 4-6

3. Can We Defang Microtargeting?

The “Popup” Intervention: Studies 4-6

- Real Facebook ads
- Persuasiveness ratings (dependent variable)
- Type of ad (independent variable)
 - Non-targeted
 - Targeted no popup
 - Targeted with popup

} Critical contrast

The “Popup” Intervention



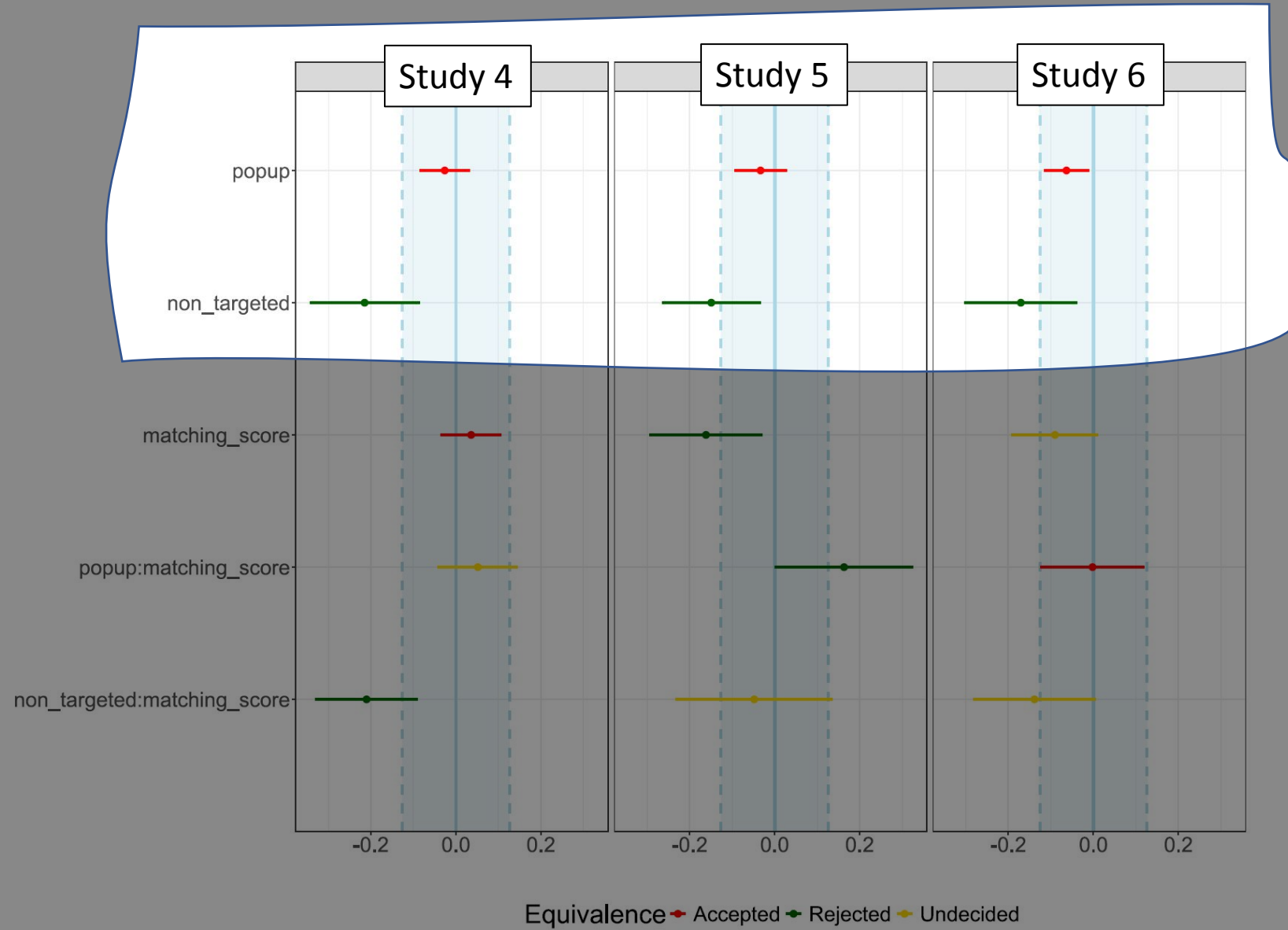
Warning: This advertisement may appear highly appealing or tailored specifically to you. If you come across such content, it's probably due to psychological microtargeting.

Microtargeting customizes content based on your personal data, preferences, and online behavior. It becomes 'psychological' when it covertly infers psychological features from your data to tailor manipulative messages, often for political purposes or to spread misinformation.

Close



Results



(Carrella, Simchon, Edwards, & Lewandowsky, 2025)

Conclusions

- Personality predicts what text people prefer to consume
- Machine-learning model can identify match between text and person
- Personality-based microtargeting “works”
- Popup warnings do not cancel targeting advantage
- Transparency may not be enough

Latest publication

