



INSTITUTE OF CRIMINOLOGY
at the Faculty of law Ljubljana



BIG DATA: NEW CHALLENGES FOR LAW AND ETHICS

International scientific conference, 22 - 23 May 2017, Faculty of Law University of Ljubljana

BIG DATA, PSYCHODIAGNOSTICS AND THREATS TO PERSONAL AUTONOMY

Friderik Klampfer, Bojan Musil, Nenad Čuš Babič, Domen Bajde



Univerza v Mariboru

BIG DATA IN SOCIAL SCIENCES AND PSYCHOLOGY

BIG DATA AND PSYCHOLOGY: PROS AND CONS

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

BIG DATA IN SOCIAL SCIENCES AND PSYCHOLOGY

Big data involves the storing, retrieval, and analysis of large amounts of information (Harlow and Oswald, 2016, p. 447).

Examples:

Work-Related Values (Hofstede, 1980, 2001)

Big Five Model – OCEAN (Costa, Terracciano, & McCrae, 2001; McCrae, Terracciano, et al., 2005)

WORLD VALUES SURVEY (WVS, 2017)

BIG DATA IN SOCIAL SCIENCES AND PSYCHOLOGY

Digital era...

Datafication of everyday life = a product of the extent to which people use digital and online technologies, as well as the extent to which digital technologies have replaced old ways of life (Chen & Wojcik, 2016, p. 458).

“The individual and (ideas of) individuality are being transformed into quantifiable, statistical, and objective data points that can (allegedly and supposedly) help reveal new possibilities, novel insights, and hidden facts about ourselves. A data-driven life, in other words, is datafying the individual.” (Kosciejew, 2013)

Three Vs of big data:

Volume: (large) size of dataset

Velocity: high speed of data acquisition and processing

Variety: diversity of data types (unstructured mixture of texts of speeches, photographs, videos and numbers)

+ fourth V – **Veracity** – the importance of truthfulness of the data

BIG DATA IN SOCIAL SCIENCES AND PSYCHOLOGY

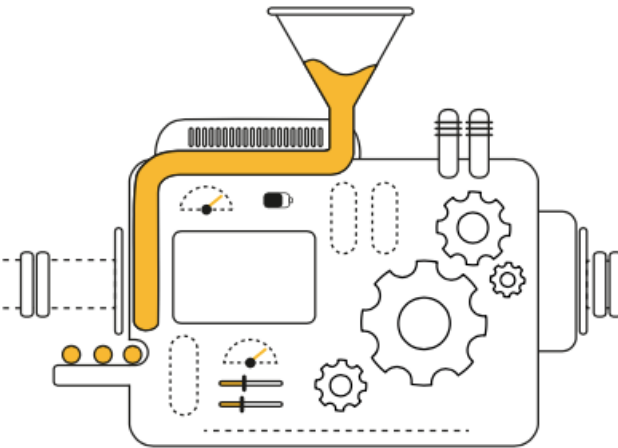
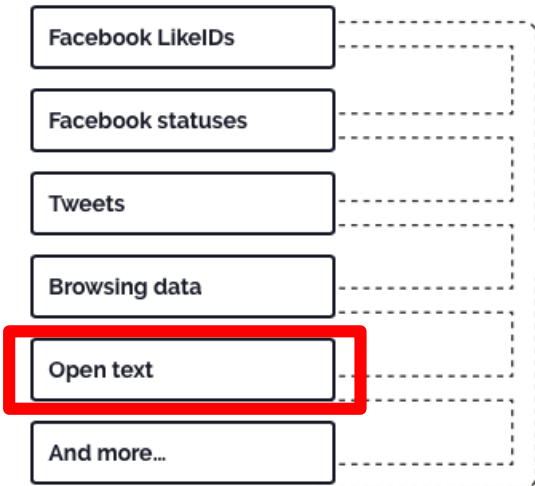
Studies predicting psychological variables from online social networks or communication channels (e.g., Facebook, Twitter).

Studies of Kosinski et al.:

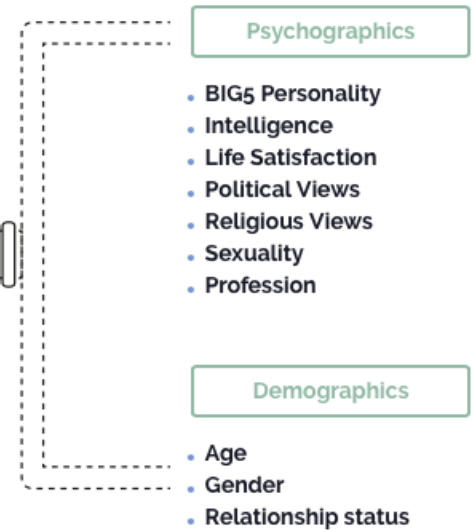
It is relatively reliably possible to predict some highly sensitive personal data like *“sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender”* (Kosinski, Stillwell & Greapel, 2013: 5802) using just Facebook likes.

Trait Prediction Engine

Digital footprints



Individual profiles



Psycholinguistic trait prediction

Write or paste some text in the box below to obtain a prediction. There is no minimum word count, but at least 200 words is preferable. Please also tag the source of the text you are submitting using the selection box provided below.

Text*

However, seriously, our interest in selfies was some sort of side effect and at the same time a next step from our previous research on the impact of videos or photos, depicted with webcams, on self-reflection. Since those results indicated a trend that webcam self-portrait photos had some (limited) effect on decrease in public states of self-consciousness, we thought about different a format of pictures, self-portrait photos, in different social situations and their relations to some psychological variables.

Source*

Website

Captcha*



I'm not a robot



reCAPTCHA
Privacy - Terms

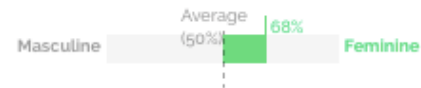
Predict

Age

34

Your digital footprint suggests that your online behaviour resembles that of a 30-39 years old

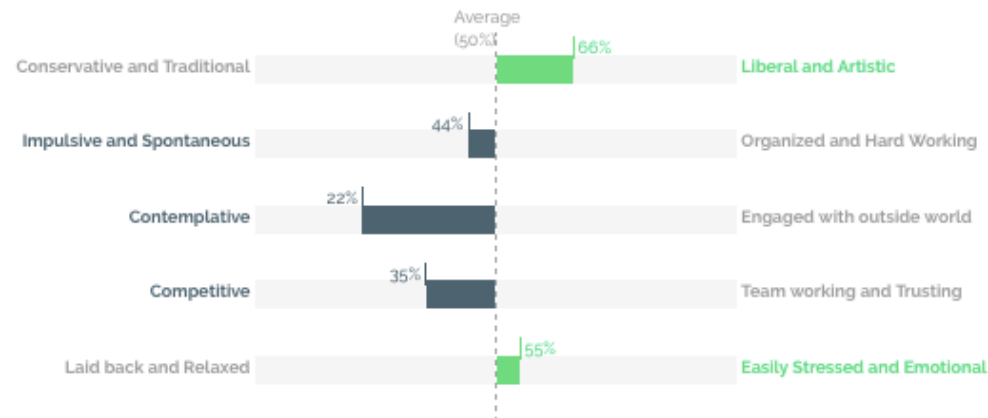
Psychological Gender



Your digital footprint suggests that you are Female

Big 5 Personality (Predictions are expressed as percentiles)

[Take personality test](#)



Openness to experience describes a dimension of personality that distinguishes imaginative, creative people from down-to-earth, conventional people.

Your digital footprint suggests that you are intellectually curious and appreciative of what you consider beautiful, no matter what others think. You might say that your imagination is vivid and makes you more creative than many others.

Conscientiousness concerns the way in which we control, regulate, and direct our impulses.

Your digital footprint suggests that you are random and fun to be around but can also plan and persist when life requires it. It appears that depending on the situation, you can make quick decisions or deliberate for longer if necessary.

Extraversion is marked by pronounced engagement with the external world, versus being comfortable with your own company.

Your digital footprint suggests that you are similar to people who prefer low-key social occasions, with a few close friends. You might say that it's not that you are afraid of large parties; they're just not that fun for you.

Agreeableness reflects individual differences in concern with cooperation and social harmony.

Your digital footprint suggests that you can find it difficult to get along with others when you first meet them. You might be suspicious of others' motives in this situation. It also looks like people warm to you over time, and you to them, although that doesn't stop you telling them "how it is".

Neuroticism refers to the tendency to experience negative emotions.

Your digital footprint suggests that you are generally calm. You come across as someone who can feel emotional or stressed out by some experiences, but your feelings tend to be warranted by the situation.

Age

Psychological Gender

34

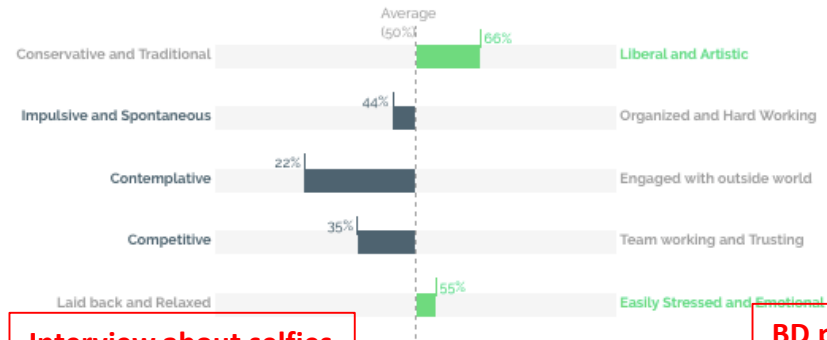


Your digital footprint suggests that you are Female

Your digital footprint suggests that your online behaviour resembles that of a 30-39 years old

Big 5 Personality (Predictions are expressed as percentiles)

Take personality test



Interview about selfies

BD paper – New definition of BD

Age

Psychological Gender

60

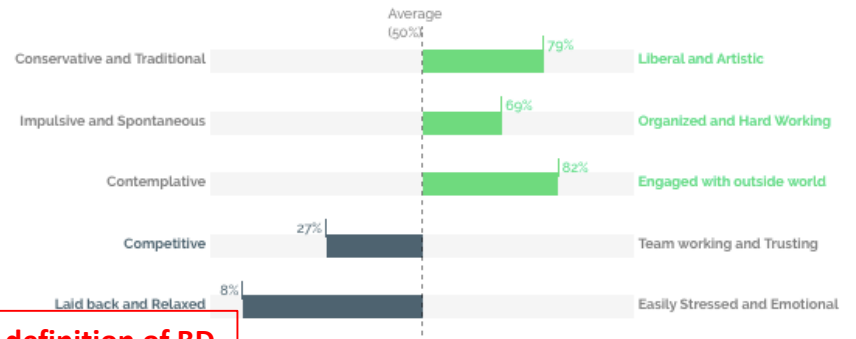


Your digital footprint suggests that you are the epitome of masculinity

Your digital footprint suggests that your online behaviour resembles that of someone aged 40+

Big 5 Personality (Predictions are expressed as percentiles)

Take personality test



Age

Psychological Gender

47

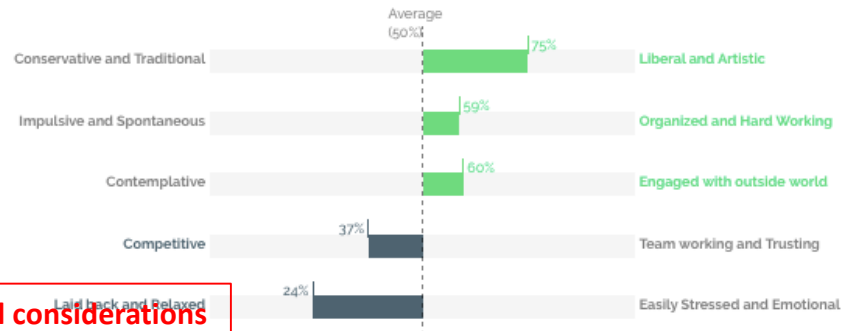


Your digital footprint suggests that you are Male

Your digital footprint suggests that your online behaviour resembles that of someone aged 40+

Big 5 Personality (Predictions are expressed as percentiles)

Take personality test



BD paper – Ethical considerations

BIG DATA IN SOCIAL SCIENCES AND PSYCHOLOGY

BIG DATA AND PSYCHOLOGY: PROS AND CONS

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

BIG DATA AND PSYCHOLOGY: PROS AND CONS

Common themes in the research of big data and psychology

(from Harlow and Oswald, 2016) :

1. There are mutual **benefits of collaboration of diversity of disciplines**, such as those from social sciences, applied statistics, and computer science.
2. Availability of large data sets from different OSNs provide a psychological window into the **attitudes and behaviors of a broad spectrum of the population**, and, in methodological sense, there is an opportunity and also necessity of **testing diversity of predictive models** in big data.
3. In process of acquiring and processing large data sets from public or private sources there are **important ethical considerations**.

BIG DATA AND PSYCHOLOGY: PROS AND CONS

Common themes in the research of big data and psychology

(from Harlow and Oswald, 2016) :

1. **Benefits of collaboration of diversity of disciplines...**

Psychologists are in majority suspicious about big data movement, mostly because of widespread subjective perception about their impairments in the area of computer programming and related IT skills and lack of access to big data (e.g., Cheung & Jak, 2016), but there is much more of unrestricted optimism in the field of computer science.

However, psychologists with knowledge, skills and capacities of psychological and behavioral theories, psychometrics and statistics are valuable in understanding and processing big data (Cheung and Jak (2016)).

BIG DATA AND PSYCHOLOGY: PROS AND CONS

Common themes in the research of big data and psychology

(from Harlow and Oswald, 2016) :

2. Testing diversity of predictive models of attitudes and behaviors of a broad spectrum of the population...



- classical psychological research techniques (questionnaires and experiments) online + opportunity for the third fundamental technique – observation of human-computer interaction on a very large scale,
- data can be divided in large spectrum of samples (e.g., distinct groups, socio-structural groups, subcultures and cultures) + historical perspective,
- focus on directly recorded behavior(s) -> no biases such as tendency to produce socially desirable answers on self-report measures or in spite of tracking real behavior tracking behavioral intention or perception of behavior from past experiences.

BIG DATA AND PSYCHOLOGY: PROS AND CONS

Common themes in the research of big data and psychology

(from Harlow and Oswald, 2016) :

2. Testing diversity of predictive models of attitudes and behaviors of a broad spectrum of the population...



- focus on directly recorded behavior(s) -> do digital or virtual behaviors (clicking or tapping) truly or authentically represent individual? Can we adequately infer internal processes from these behaviors? Or vice versa?

- social interaction thru computer-mediated communication (CMC) represents significant part of our everyday lives, but is still specific kind of interaction.

- online social network sites represent the complexity of the social interaction -> important place(s) of social influence, thus having significant effect on someone's thoughts, feelings, behaviors, consequently hers/his self-definition and identity -> informative (i.e., what is right) or/and normative (i.e., what is consensual) impact on individual, and consequently influence hers/his future decisions and behaviors.

BIG DATA AND PSYCHOLOGY: PROS AND CONS

Common themes in the research of big data and psychology

(from Harlow and Oswald, 2016) :

3. Important ethical considerations...

Ambiguous situations, such as...:

Researcher or research team can deduce personality features of a user from his/her online behavior and hence have the potential to deny him/her a particular contribution (Montag et al., 2016).

In the analysis of big data characteristics of people, not even being in the scope of the particular big data sample can be processed (e.g., see article Webb, 2013).

Major ethical concerns of big data processing – **questions of privacy, anonymity, and autonomy**. How researchers deal with private information in phase of collecting, processing and disseminating the data? Is the anonymity of research subjects preserved? Are participants informed about research and do they have a right to opt out of research? To sum it, how is a **respect for persons** (i.e. ICT users) addressed?

Researchers are obliged to seek and obtain **informed consent** of the research subjects.

BIG DATA IN SOCIAL SCIENCES AND PSYCHOLOGY

BIG DATA AND PSYCHOLOGY: PROS AND CONS

**BIG DATA AND ETHICS: THE MANIPULATIVE
POTENTIAL OF PSYCHODIAGNOSTIC TOOLS**

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

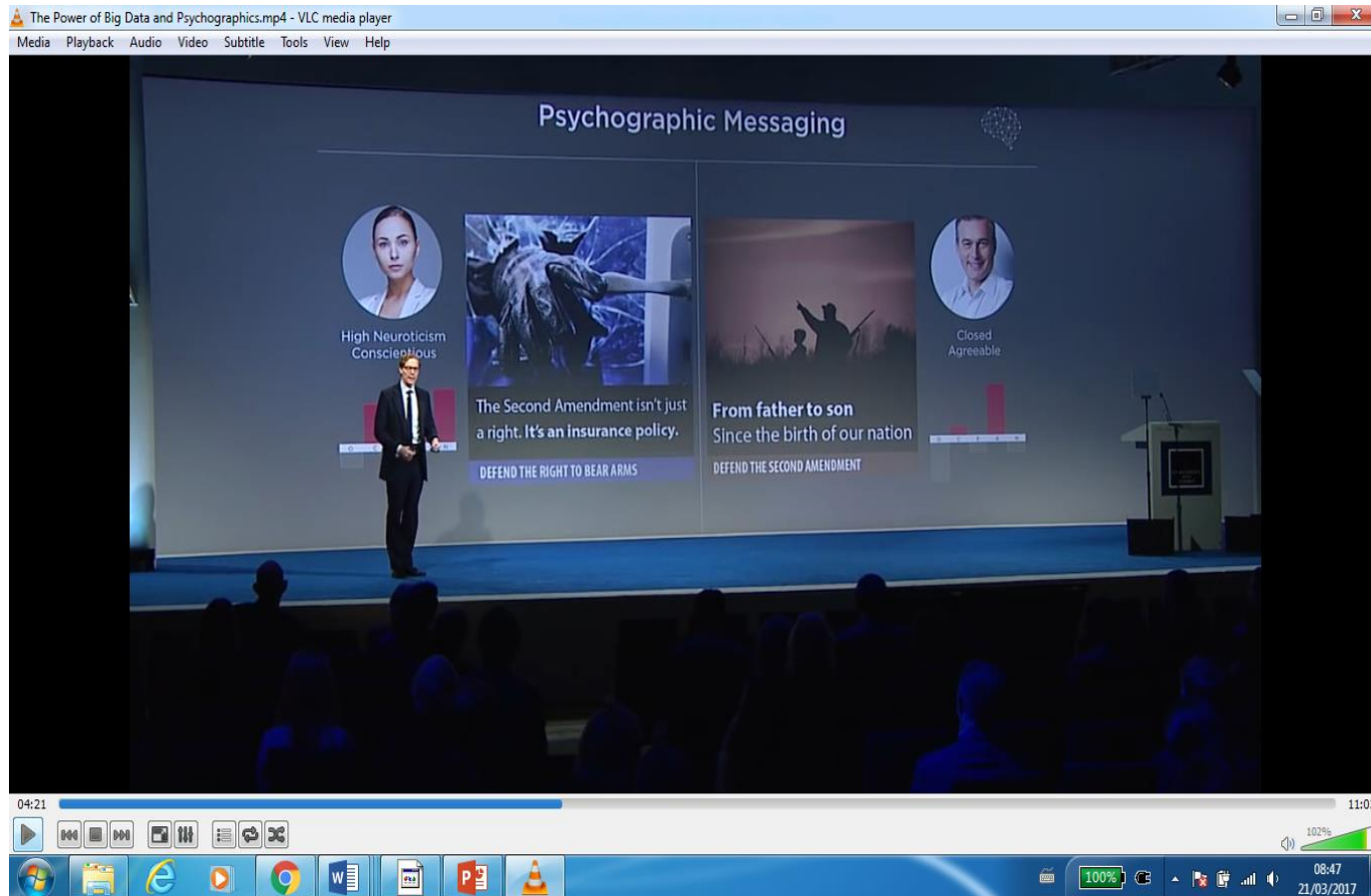
The story of Alexander Nix and Cambridge Analytica

- **background check:** military origins (originally SLC), an important protagonist in the right-wing online propaganda war against liberal, left-biased mainstream media, owned and financed by a US billionaire Robert Mercer
- **service:** offers targeted, individualized, personalized political ads (what they cryptically call 'psychographic messaging')
- **references:** provided services for the Ted Cruz and (after his defeat in the Republican preliminaries) the Donald Trump campaign, possibly also implicated in the campaign for Brexit in the UK
- **value:** boasts of harvesting 5000 pieces of (legally obtained) information on every single one of 150 million US voters

Disclaimer: a lot of this is still fairly speculative

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

‘Psychographic messaging’



The screenshot shows a VLC media player window titled "The Power of Big Data and Psychographics.mp4 - VLC media player". The video content displays a presentation slide titled "Psychographic Messaging". The slide is divided into two columns. The left column features a circular portrait of a woman, the text "High Neuroticism Conscientious", an image of a soldier in a trench, and the text "The Second Amendment isn't just a right. It's an insurance policy." Below this is a blue button labeled "DEFEND THE RIGHT TO BEAR ARMS". The right column features a circular portrait of a man, the text "Closed Agreeable", an image of a soldier in a trench, and the text "From father to son Since the birth of our nation". Below this is a blue button labeled "DEFEND THE SECOND AMENDMENT". A presenter in a suit is visible on the stage in front of the screen. The VLC player interface includes a progress bar at 04:21, a volume control at 102%, and a Windows taskbar at the bottom showing the time 08:47 on 21/03/2017.

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Psychographic messaging: How is it supposed to work?
And, more importantly, does it work?

- the tactics CA uses is a combination of three elements: **behavioral science, big data analysis and targeted ads**
- demographics is substituted (or, better, supplemented) by **psychometrics**
- psychological profiles enable **personalized messages**, with tone and content tailored to every voter's personality, ("Pretty much every message Trump put out was data-driven." A. Nix)

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Manipulation: What is it?

Three conditions seem necessary for most if not all instances of manipulation.

First, like coercion (and deception and lying and blackmail and psychological compulsion) it is an ***intentional attempt to influence another's choices via "non-rational" means*** rather than "rational persuasion" or a plea for a favor.

Second, unlike coercion (and psychological compulsion), **manipulation does not limit one's options** or reduce one's negative liberty.

Third, **the means used exploit some deficiency** in the recipient and would be wholly ineffective on someone who lacks the deficiency—that is, someone who is well informed, in perfect self-control, and in possession of full economic and epistemic rationality.

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Manipulation: Ethical assessment

What makes manipulation **pro tanto** wrong?

The fact that **it compromises (undermines, thwarts, subverts, cancels out) our rational autonomy:**

“When a person has been manipulated or coerced her rational autonomy is compromised. One is rationally autonomous to the extent that one’s choices accord with and are guided by reasons that one does or would count as good, justifying ones. When I influence your choices via open and honest rational persuasion, they reflect your rational autonomy. **When I deceive you, withhold relevant information from you, blackmail you, or take advantage of your weak will, gullibility, or emotional insecurities, I get you to make choices that you yourself would not see as having good reason to make, if you were well informed, free from duress, and fully rational; thus, I hinder or reduce your rational autonomy.** Since rational autonomy is a good, manipulation is *pro tanto* wrong.” (Scoccia 2014)

Whenever we manipulate someone into doing something, we **degrade him_her as a person, i.e. treat him_her as less than a person.** (Noggle)

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Manipulation and manipulative advertising: Definition

In typical cases of non-coercive manipulation the manipulator seeks to lure his or her victim into some decision by exploiting her ignorance, irrationality or some other vulnerability.

The definition of **manipulative behaviour**:

A person, A manipulates another person, B, iff in an attempt to make her do X (i.e. act according to A's will), deliberately **bypasses B's rational and autonomous will**, i.e. if A deliberately profits from her irrationality (some deeply engrained fear, weakness of will, fallacious reasoning, especially causal and statistical, or ignorance).

And, by extension, of **manipulative advertising**:

Advertising is manipulative whenever it creates in the addressee a mental association at an unconscious level between the product and something else that's desired but is in reality causally unrelated (or even inversely related, as is the case with gun possession and physical safety).

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Manipulative ads



Hello Shopping Shirt
goodbye plastic bags.

www.shoppingshirt.com.au

Why be environ-mental when you can be environ-normal?

The Golf BlueMotion.
So fuel-efficient you can travel over 1400km on one tank.*
Visit environ-normal.com.au

BLUEMOTION


Das Auto.

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Psychographic messaging: So (how) is it manipulative?

And what, if anything, makes it more dangerously so than the traditional political advertising?

- doesn't exploit just universally shared cognitive shortcomings, but also every **individual's particular emotional vulnerabilities and insecurities** (if you happen to worry about crime, then the right to bear arms will be presented to you as a reasonable insurance policy; if, on the other hand, you cherish traditional values, it will be presented to you as a way of honouring and preserving a valuable tradition)
- every advertisement seeks **an emotional trigger**, but only psychographic messaging **knows where to find** it and so need not depend on sheer luck
- the difference is between knowing there are so and so many neurotics in the electorate and sending out a specially crafted message in the hope that it'll reach at least some of them, and **knowing who they are and how to reach them**

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Psychographic messaging: Other issues

- Subverts democratic process
- Undermines the legitimacy of democratic governance

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

Psychographic messaging: What's the fuss?

After all, the outcome of a particular popular/election can be manipulated in a myriad of ways: by

- manipulating ballots (destroying some, supplementing others)
- gerrymandering
- making it hard or practically impossible for specific groups of voters to register and vote
- outlawing rival parties and candidates
- creating favourable media coverage while at the same time suppressing public criticism, including the shutting down of the critical media
- manipulating statistics in order to present an incorrect picture of the state of a country's economy or finances
- exploiting stereotypes, prejudices, fears and intolerance against particular social groups (immigrants, homosexuals, Jews, Roma people, unemployed, homeless, Southern Europeans,...) prevalent in the given society for the purpose of scapegoating

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

A glimpse into the dark future:

The use of psychograms based on Big Data seriously threatens our autonomy as both consumers and citizens, in the context where:

- there is **no regulation and/or efficient control over the collection of, and trading in, huge amounts of personal data** (seemingly useless traces of our online/digital activities, which, however, can be easily converted into relatively precise psychological profiles);
- the **legal rules governing political campaigns are outdated** and inadequate to protect citizens' autonomous political choices and the democratic process against systematic subversion
- there is a concerted effort to win over, or sway towards a particular candidate, a significant portion of population/electorate
- there are **powerful players** in- and outside of politics wealthy enough to hire data-mining and data-analysing companies, with ordinary **moral constraints** against motive and choice manipulations pretty much suspended (**Machiavellism reigns**)
- and the public was led to believe that privacy (personal information) is sufficiently protected as long as they've gone through privacy settings on FB and made sure no unwelcome outsiders can access and share their contents, thereby foregoing **the Silent Listeners**

BIG DATA AND ETHICS: THE MANIPULATIVE POTENTIAL OF PSYCHODIAGNOSTIC TOOLS

(A somewhat overdramatic) Conclusion:

The dark future is now.