User-centric approaches

Lecture 2

Laura Boeschoten
Assistant professor
Department of Methodology and Statistics

Literature

- **Chapter 1**: “Introduction” in *Data collection with Wearables, Apps and Sensors* (2023), Florian Keusch, Bella Struminskaya, Stephanie Eckman & Heidi Guyer
  Sections 1.1 - 1.4
- Ohme, J. Araujo, T. Boeschoten, L., Freelon. D. Ram, N., Reeves, B.B. & Robinson, T.N. (2024) Digital Trace Data Collection for Social Media Effects Research: Apis, Data Donation and (Screen) Tracking, *Communication Methods and Measures, 18(2)* 124-141.
- Boeschoten, L. Ausloos, J. Möller, J.E., Araujo, T. & Oberski, D.L. (2022) A framework for privacy preserving digital trace data collection through data donation, *Computational communication research* 4(2), 388 – 423
  Pages 388 - 394

Goals of this lecture

1. Distinguish between platform- and user-centric approach to digital trace data collection.
2. Understand how commonly used user-centric approaches work and how they can be used for research.
3. Understand the advantages, challenges and ethical considerations of user-centric approaches.
4. Provide examples of studies and explain how the questions can be answered using a user-centric approach.

Digital Trace Data Collection

User-centric and platform-centric approaches

Digital traces

Digital footprint; Digital shadow

One's unique set of digital activities, actions, and communications that leave a data trace on the internet or on a computer or other digital device and can identify the particular user or device.

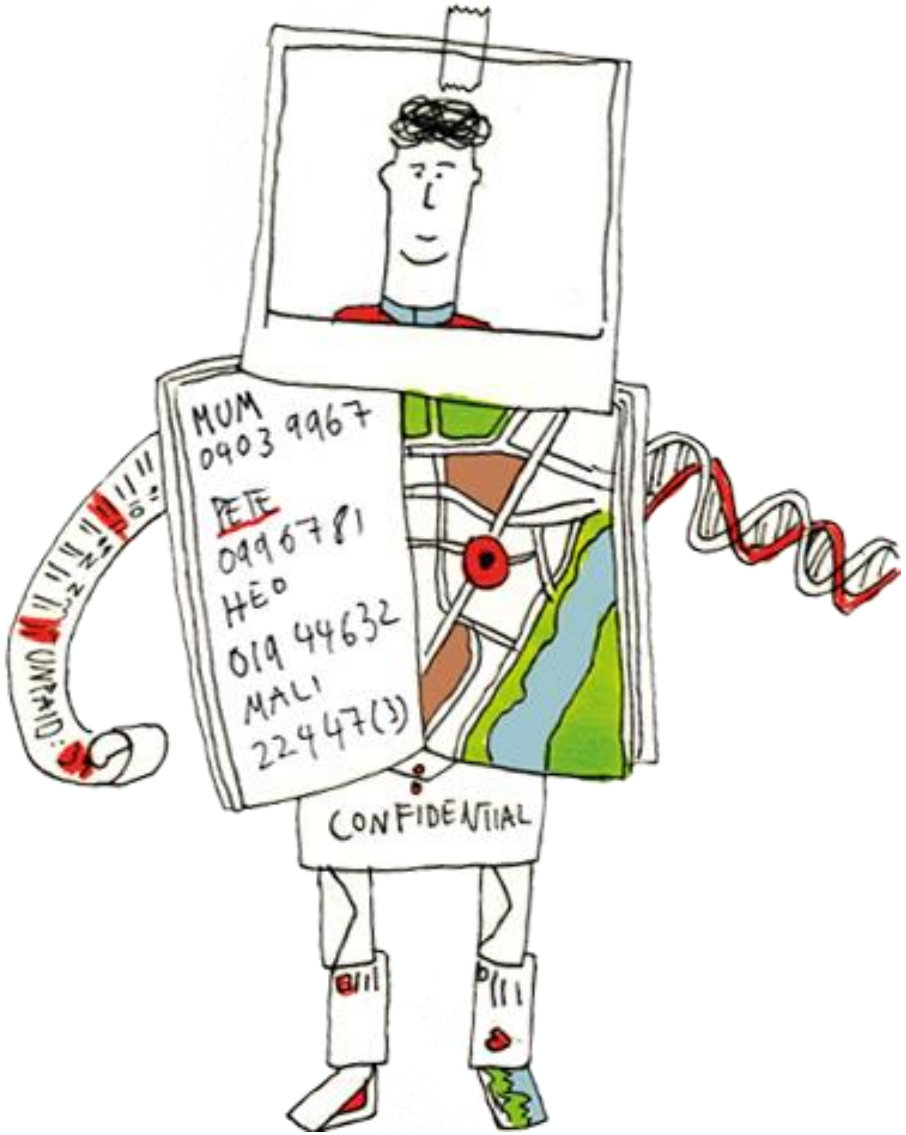
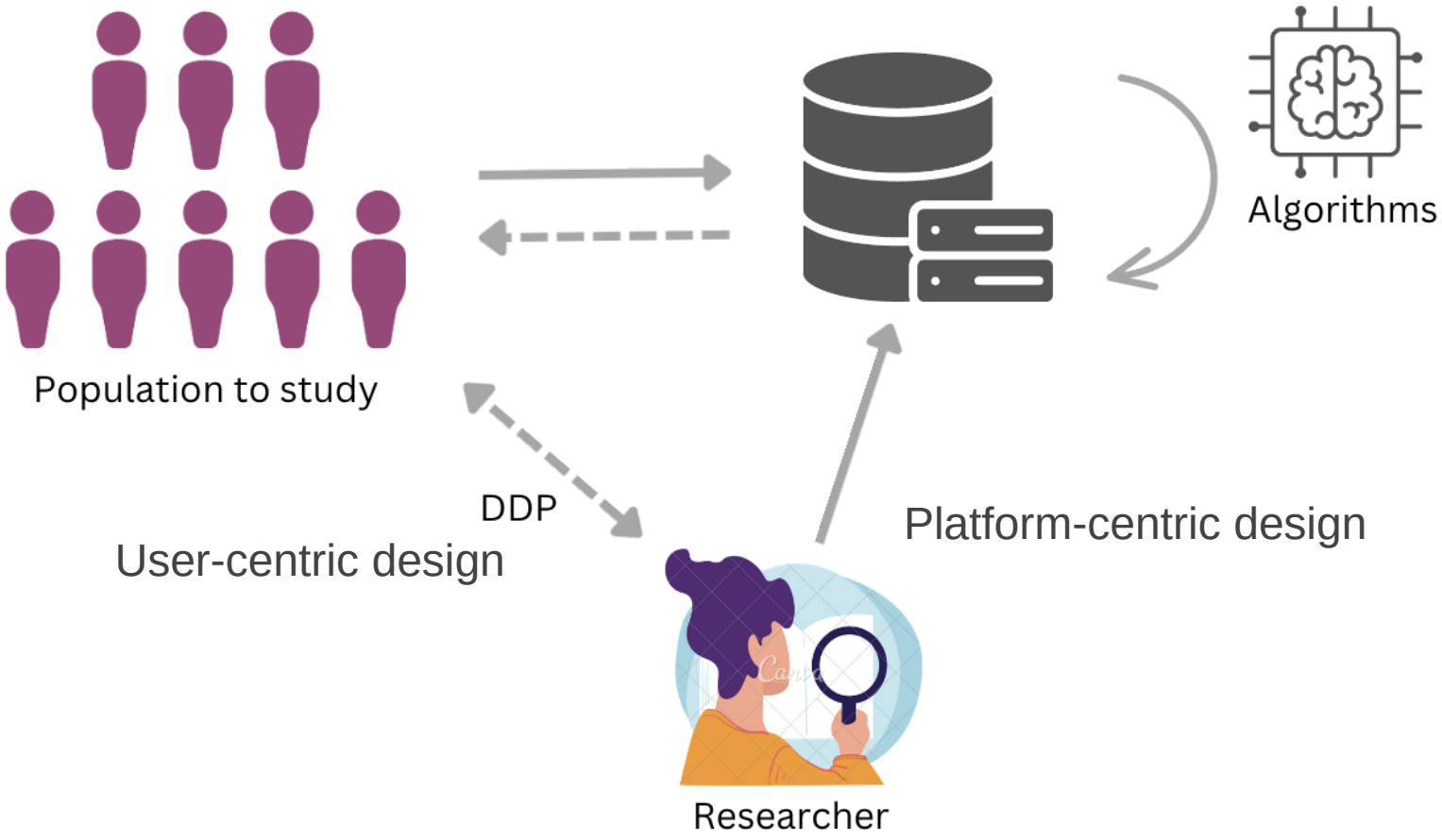


Image: https://myshadow.org/tracking-data-traces
Definition: https://www.dictionary.com/browse/digital-footprint

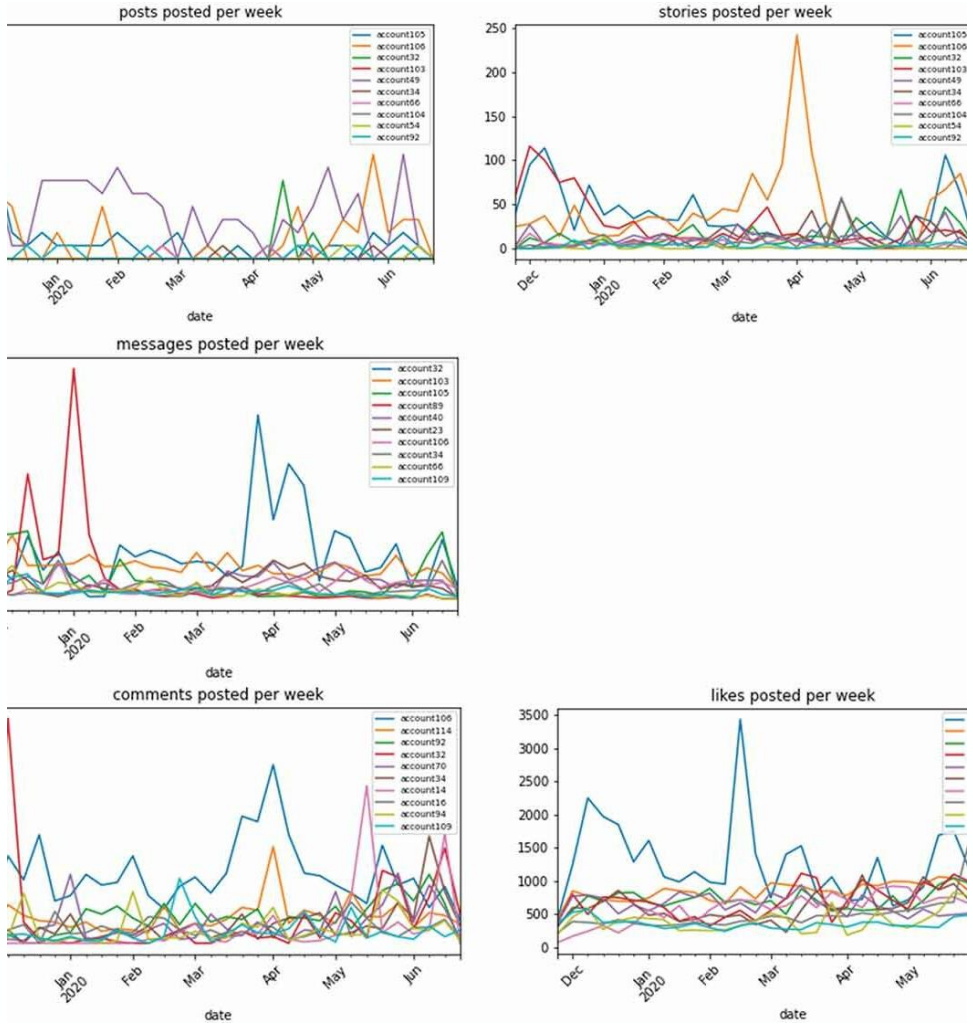
User-centric versus platform-centric



Why take a user-centric approach?

Studying person specific effects

- A **user** is any individual who generates a digital trace on a platform by making use of its functionalities.
- Selection mechanisms and potential biases shape the social contexts on which interactions on **platforms** take place.
- Platform-centric research methods allow for the study of **aggregated** or **public** social media effects, while user-centric methods allow for studying **person-specific** effects.



Fluctuations in posting, updating stories, direct messaging, commenting, and liking over time from the top 10 accounts engaging in each activity.

Valkenburg et al. (2021) and Van Driel et al. (2022)

Supplement digital traces with more info

Measurement Burst Design (Nesselroade, 1991)

N = 388 (age 13-15)

T > 250

What user-centric approaches are out there?

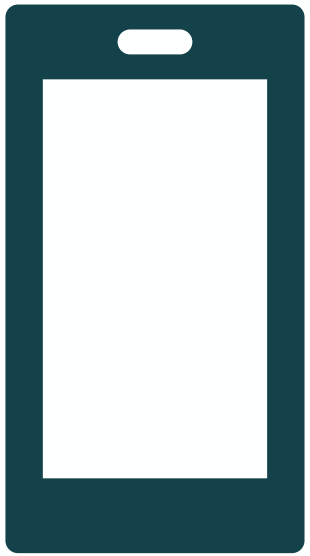
Sensor approaches

Data donation approaches

Sensor approaches

Image: Flaticon.com

What can you measure with these sensors?



Proximity

NFC

Bluetooth

Air humidity sensor

Thermometer

Location

Wi-Fi

GPS

Cellular network

Proximity sensor

Microphone

Light sensor

Ambience

Fingerprint sensor

Camera

Barometer

Compass

Accelerometer

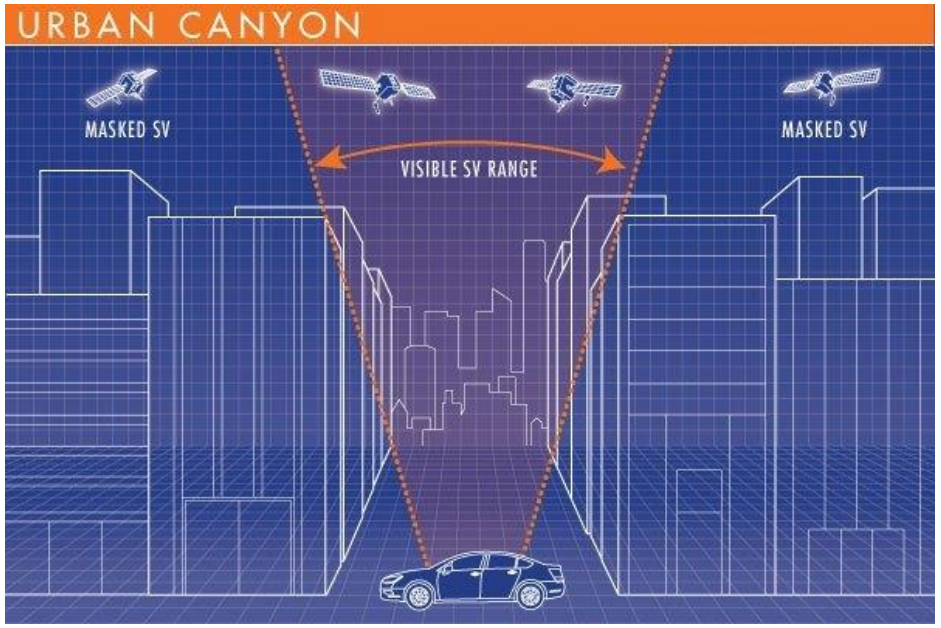
Pedometer

Gyroscope

Physical activity

Location sensors

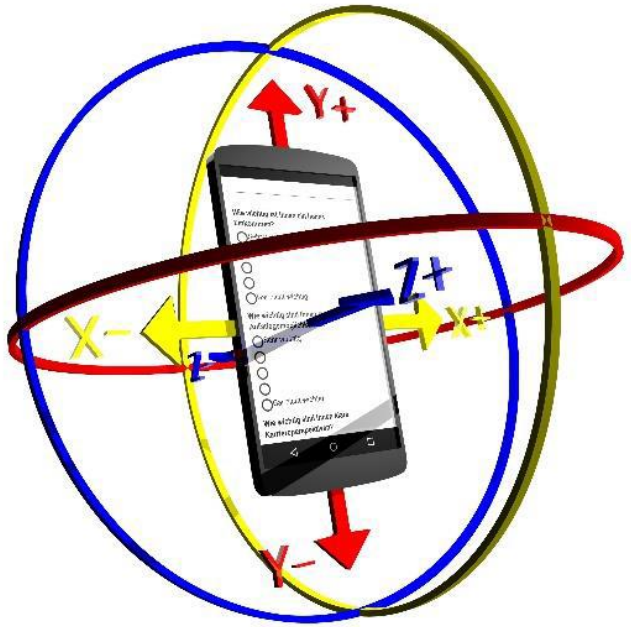
- GPS
  - Coordinates in longitude & latitude
  - High accuracy (newest generation 30 cm)
  - Works without cell/Internet connection
  - Performs worse in ‘urban canyons’, indoors, and underground (pseudo accuracy!)
  - Can be battery-draining
- Cellular network
  - Multilateration of radio signals between (several) cell towers
- WiFi
  - Inferring location from Wi-Fi access points (AP)
- Beacons
  - Bluetooth transmitters for indoors

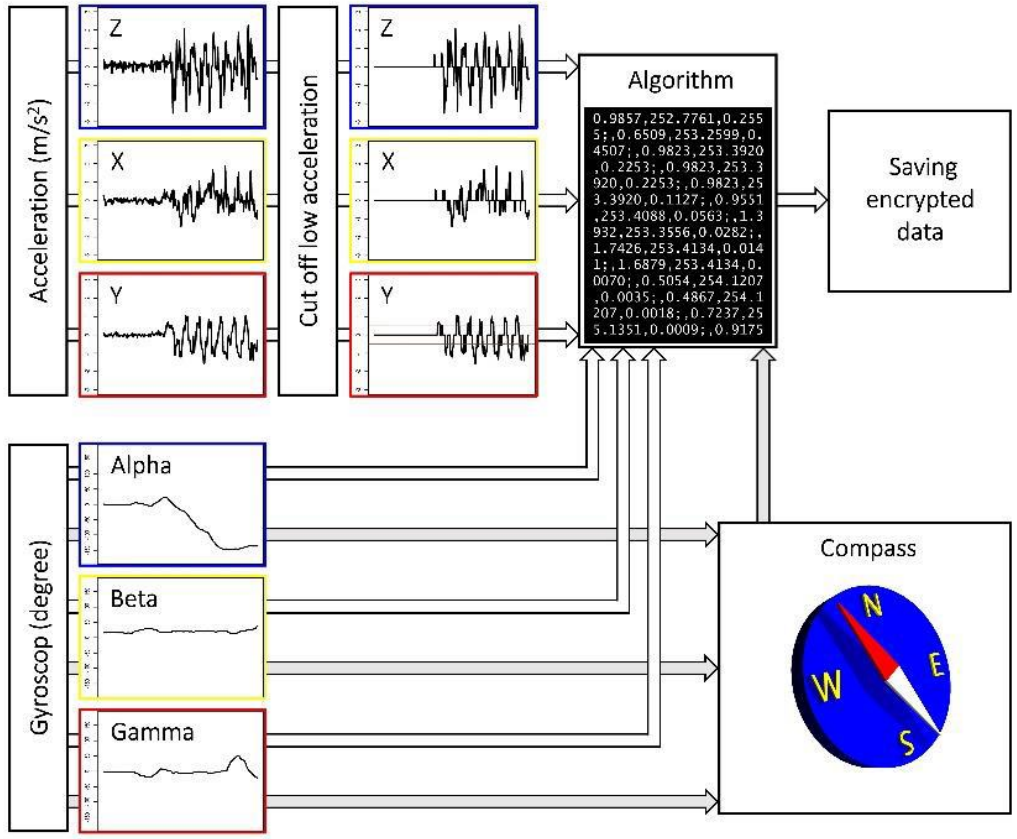


Images: https://i2.wp.com/geoawesomeness.com/wp-content/uploads/2014/01/urbancanyon.jpg?fit=600%2C400&ssl=1, https://locatify.com/wp-content/uploads/2015/03/beacon-wall-756x425.j

Physical activity sensors

- Accelerometer
- Gyroscope
- Barometer
- Pedometer





Images: Schlosser et al. (2019)

Ambience sensors, proximity sensors

- Camera
  - photos, videos, scanning of bar codes
  - linear distance
- Microphone
  - active and passive (ambient noise) recording
- Light sensor
  - e.g., identify idle state

- Bluetooth
- RFID (radio frequency)
- NFC

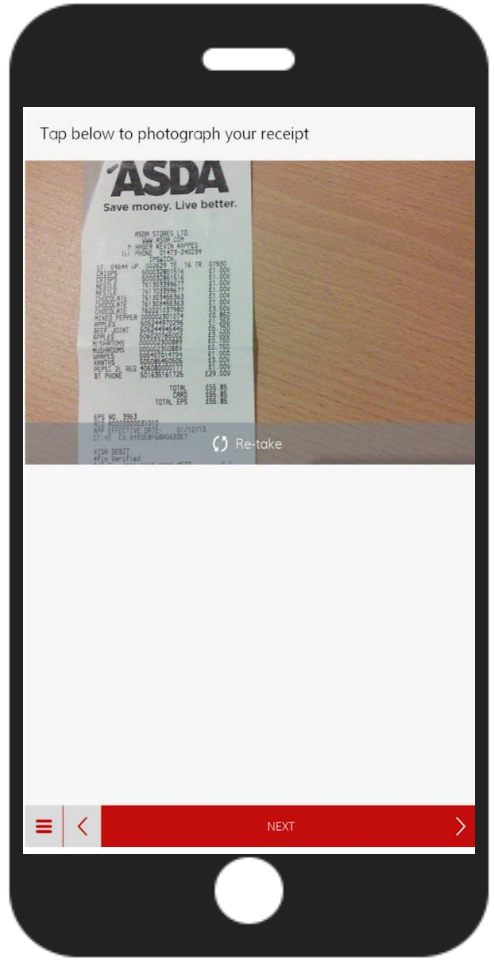


Image: Jäckle et al. (2018)

How does it work exactly?

- As a researcher, you utilize the existing sensors on participants' devices.
- Three options:

1. You ask participants to use the sensor directly.
2. You develop a piece of software (such as an app or plug-in) that captures the traces you are interested in.
3. You make use of an existing app or plug-in.

- You invite participants.
- Data collection is prospective, it collects the traces as they are being produced.

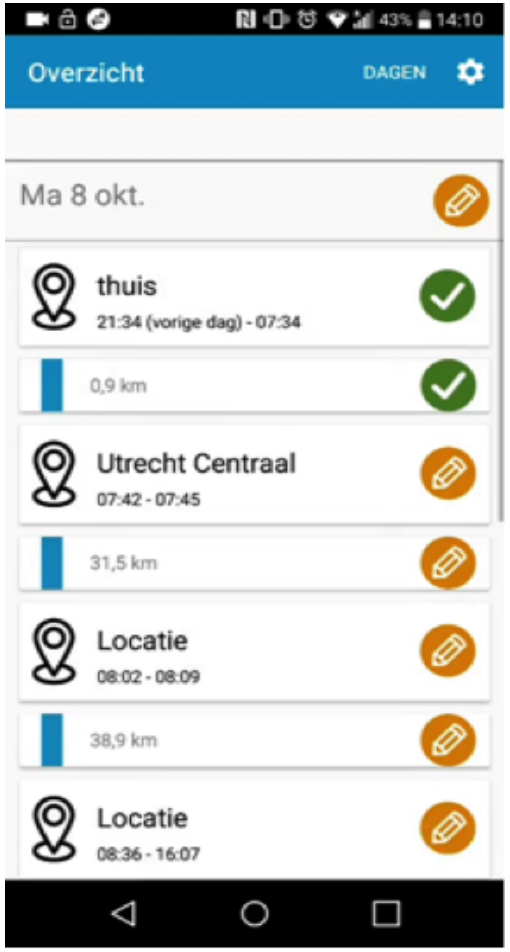
Example 1

Measuring everyday mobility
using an app

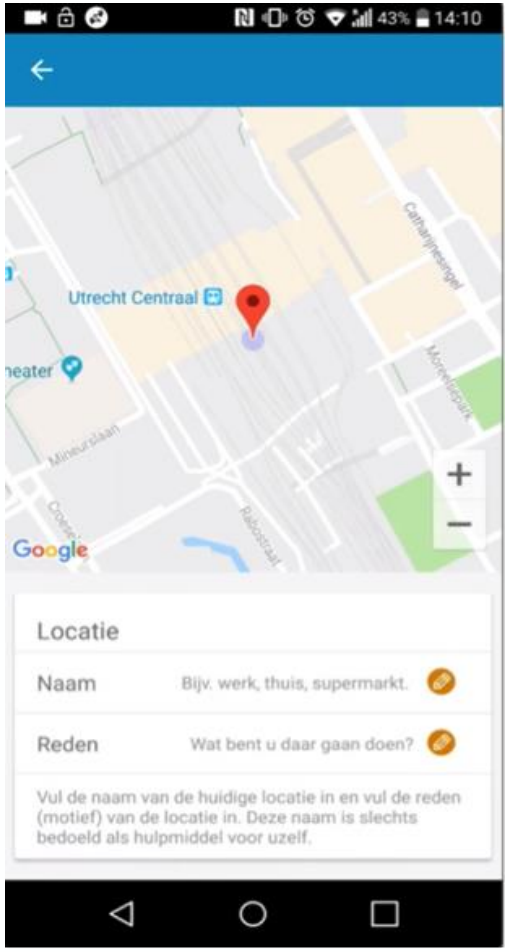
How do people move in everyday life?

- Field test in the Dutch general population (Nov-Dec 2018)
  - Travel app of Statistics Netherlands ("TABI APP" Android & iPhone)
  - Data collection for 7 days
  - N = 1,902
- Sensing location per second (when moving) & per minute (when still):
  - GPS
  - Wi-Fi
- Respondents eagerly provide additional information that helps understand travel behavior (label stops and motives for travel).

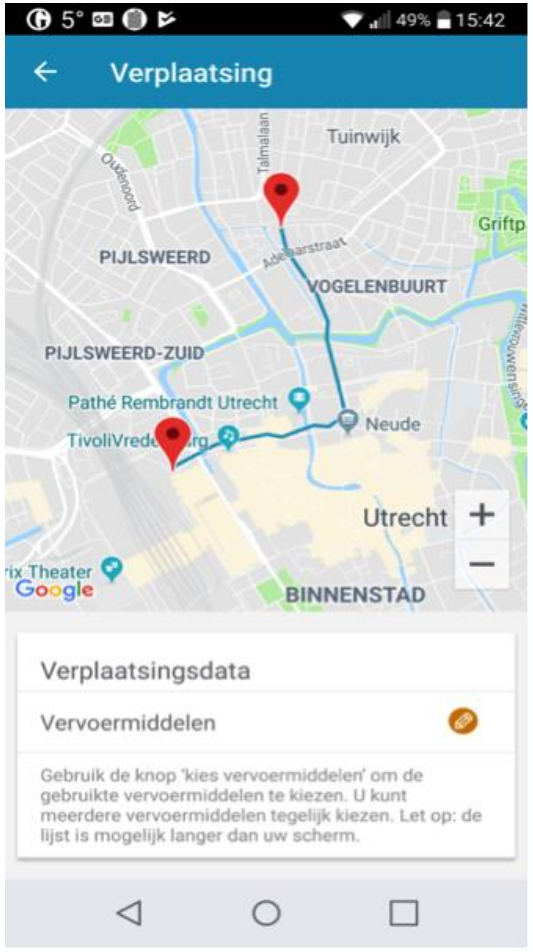
See McCool et al. (2021)



Daily overview



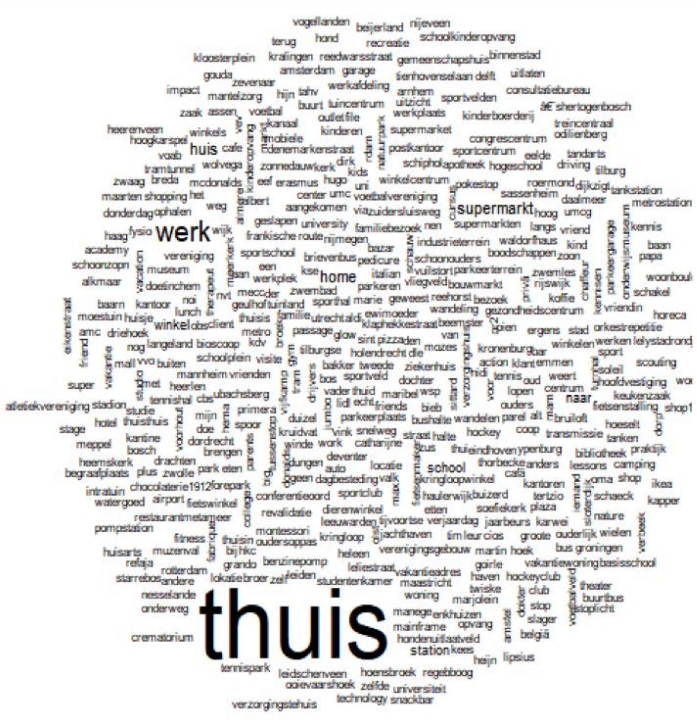
Questions about stops

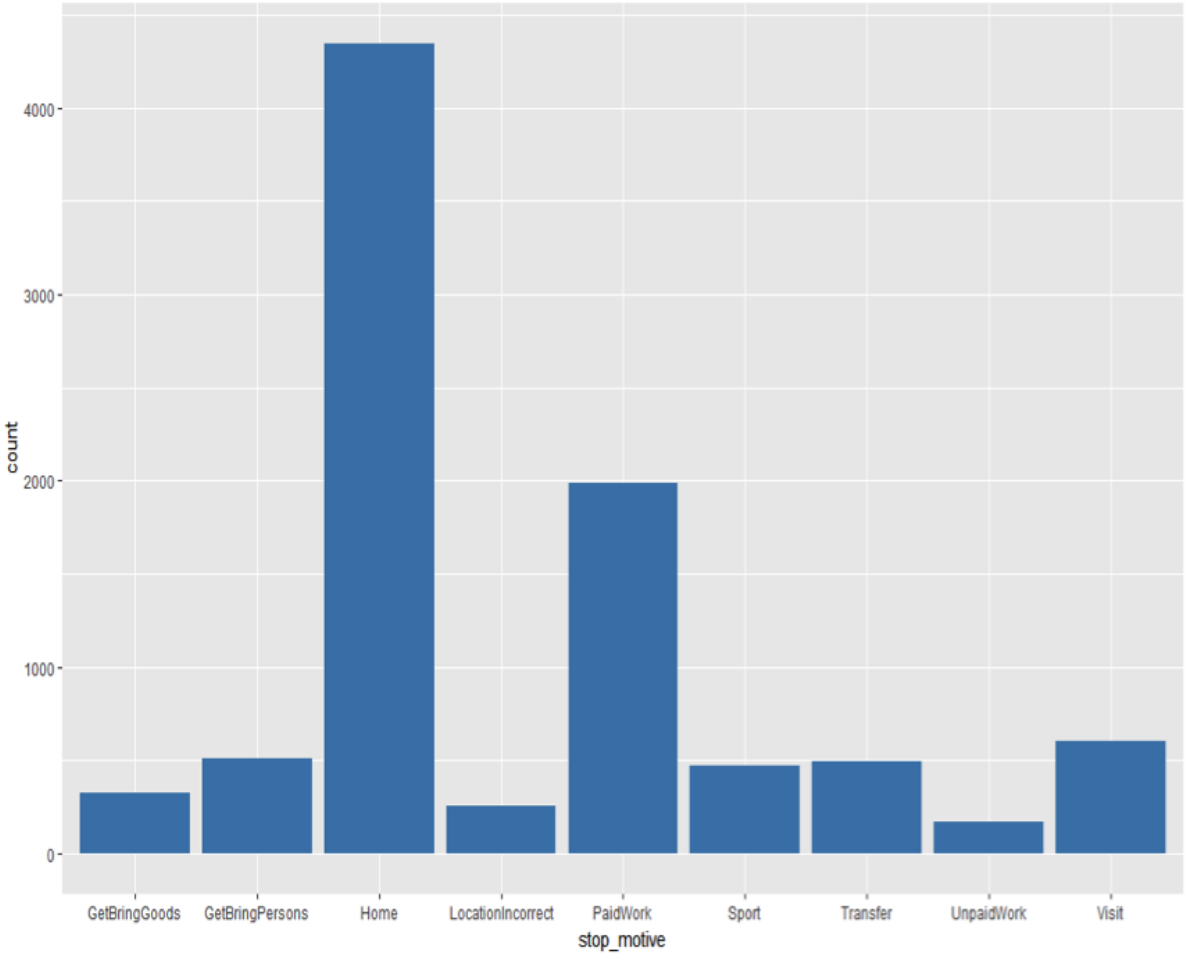


Questions about trips

Motives and means of transportation

- 22,000 stops
  - 13,000 (60%) labeled
  - Overall 50% of respondents give complete/almost complete details





Example 2

Mobility and employment using an app

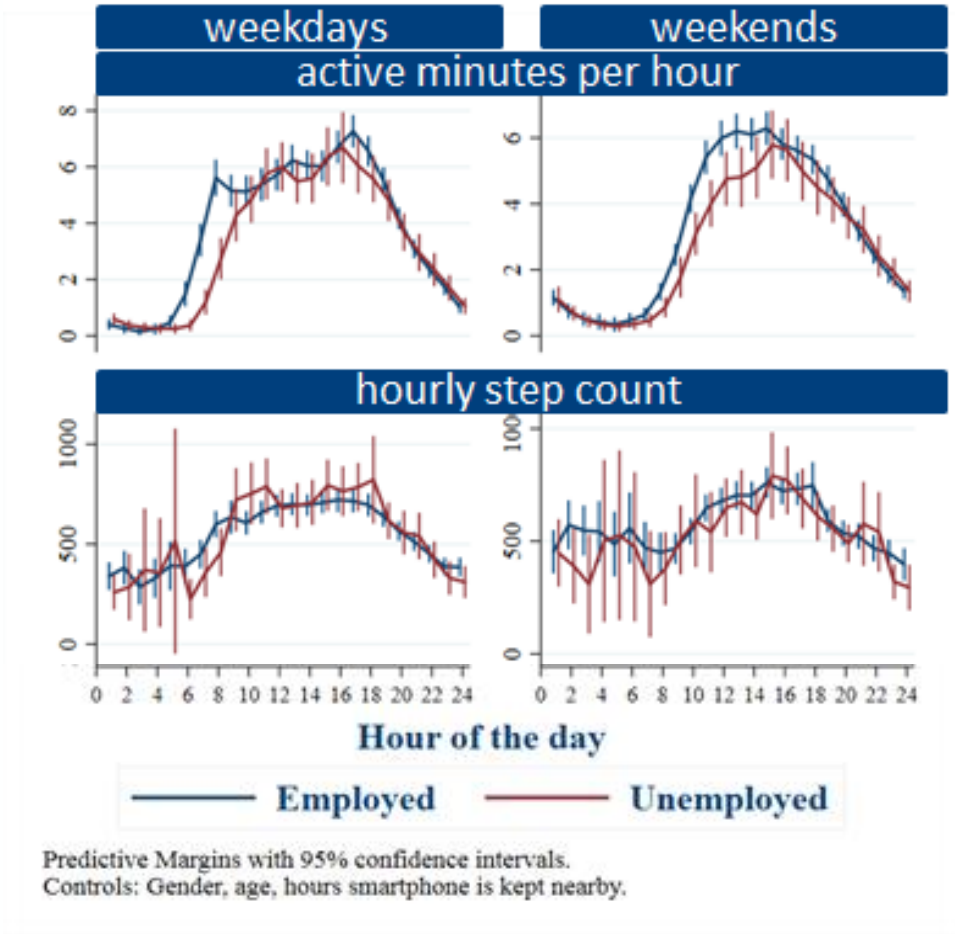
What are the effects of unemployment?

(Kreuter et al. 2018)

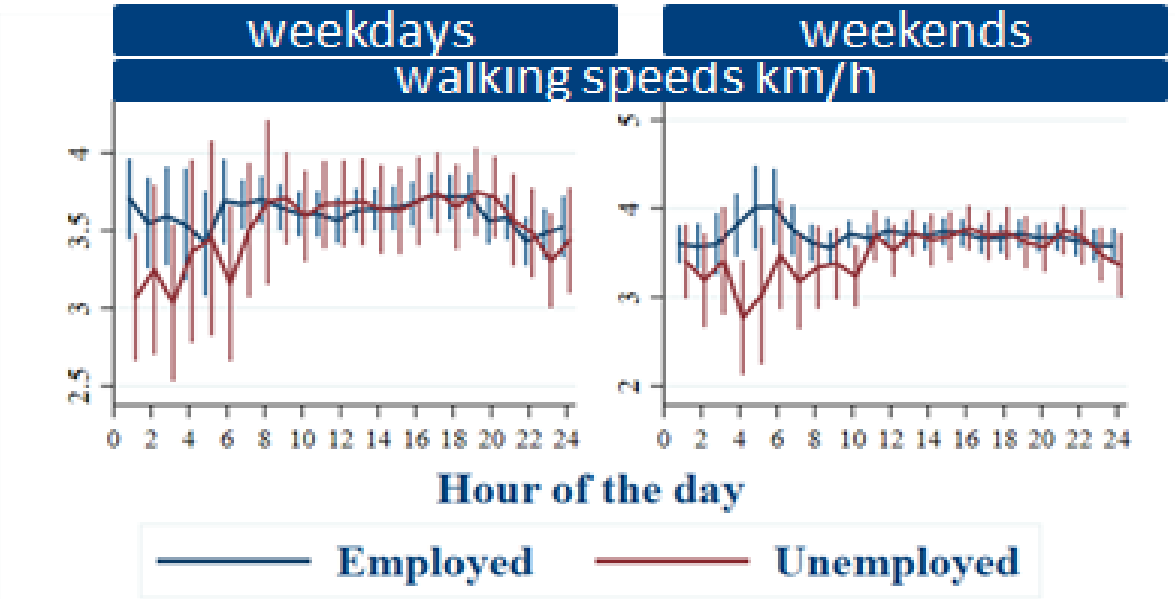
- 650 Android owners from German panel study "Labour Market and Social Security" (PASS) downloaded *IAB-SMART* app for 6 months
- EMA questions concerning both subjective (e.g., affective impact of daily smartphone use, Big 5 personality) and objective phenomena (e.g., employment and job search activities, use of smartphones in everyday life, memberships in professional and voluntary organizations)
- Five sensing modules:
  - Location using GPS, Wi-Fi, and cellular sensors every 30 minutes
  - Activity and means of transportation (e.g., walking, biking, riding in/on a motorized vehicle) using accelerometer and pedometer
  - Call and texting behavior using phone and SMS logs
  - Use of apps installed on smartphone
  - Social network characteristics from contact lists.

What are the effects of unemployment?

(Kreuter et al. 2018)



Predictive Margins with 95% confidence intervals.
Controls: Gender, age, hours smartphone is kept nearby.



Predictive Margins with 95% confidence intervals.
Controls: Gender, age, hours smartphone is kept nearby.

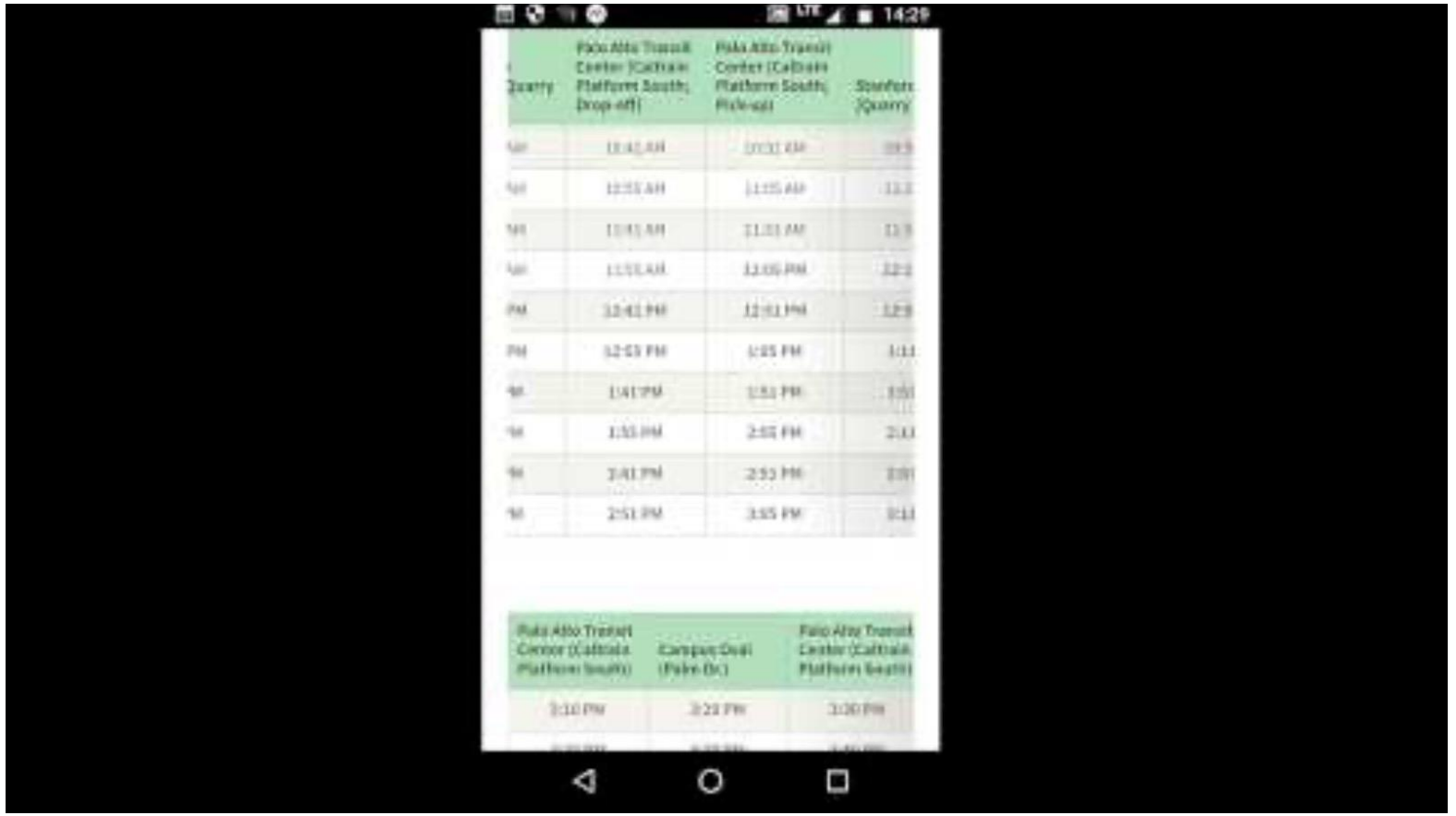
Example 3

The human screenome project

The human screenome project

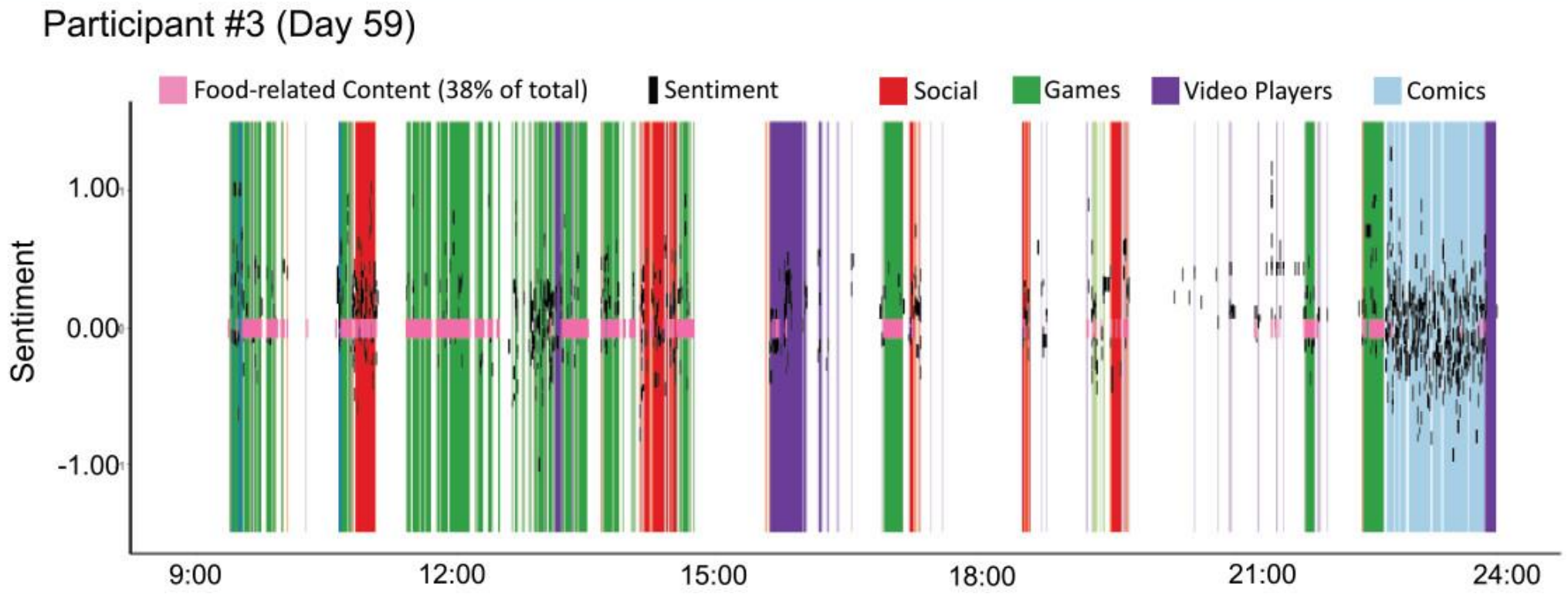
- The **Screenomics** software records, encrypts and transmits screenshots automatically and unobtrusively **every 5 seconds**, whenever a device is turned on.

- When it is deployed on **multiple devices** at once, the screenshots from each one are synced in time.



See Reeves et al. (2020)

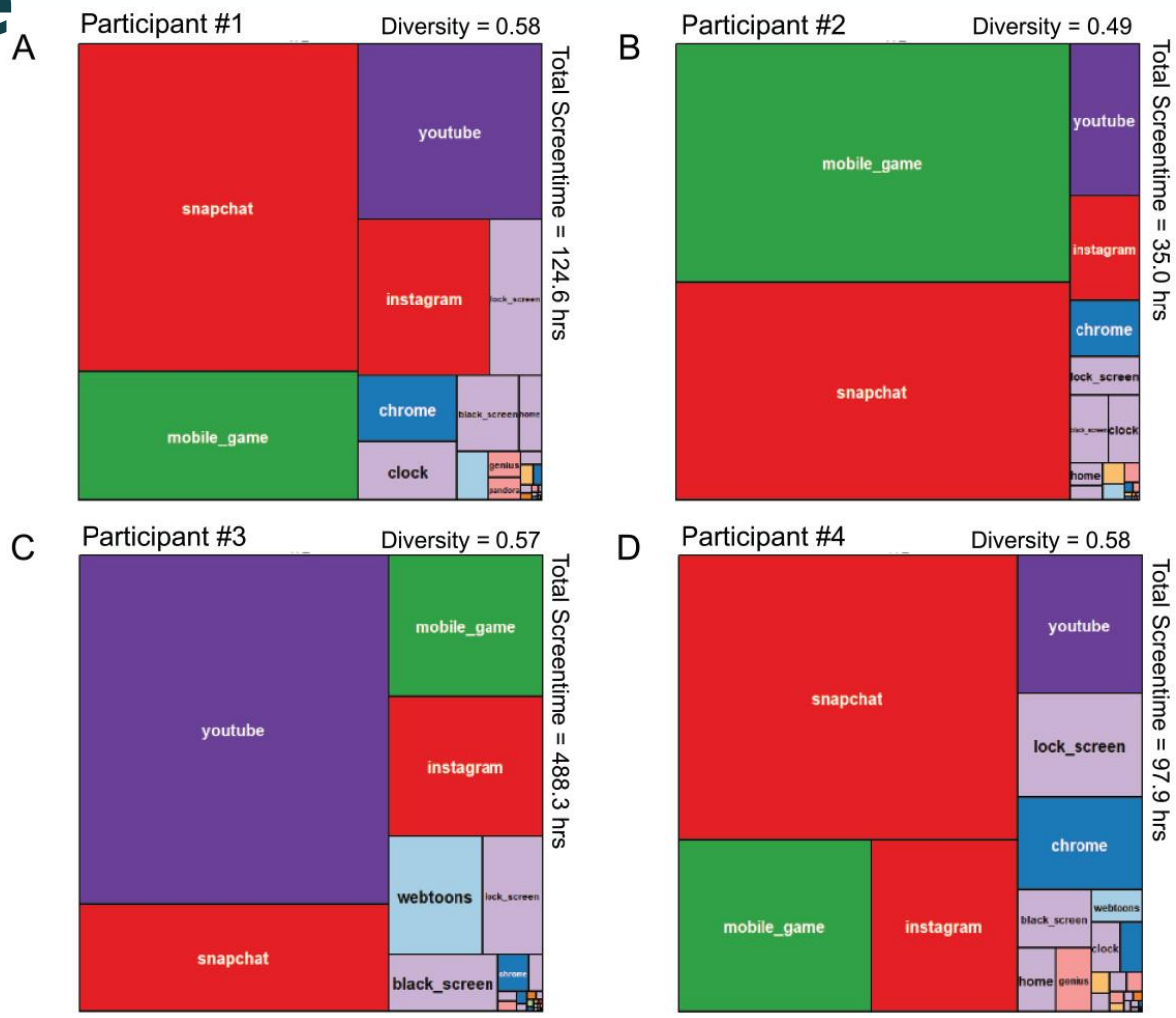
Example data from the Screenomics app



See Ram et al. (2019)

Example data from the Screenomics app

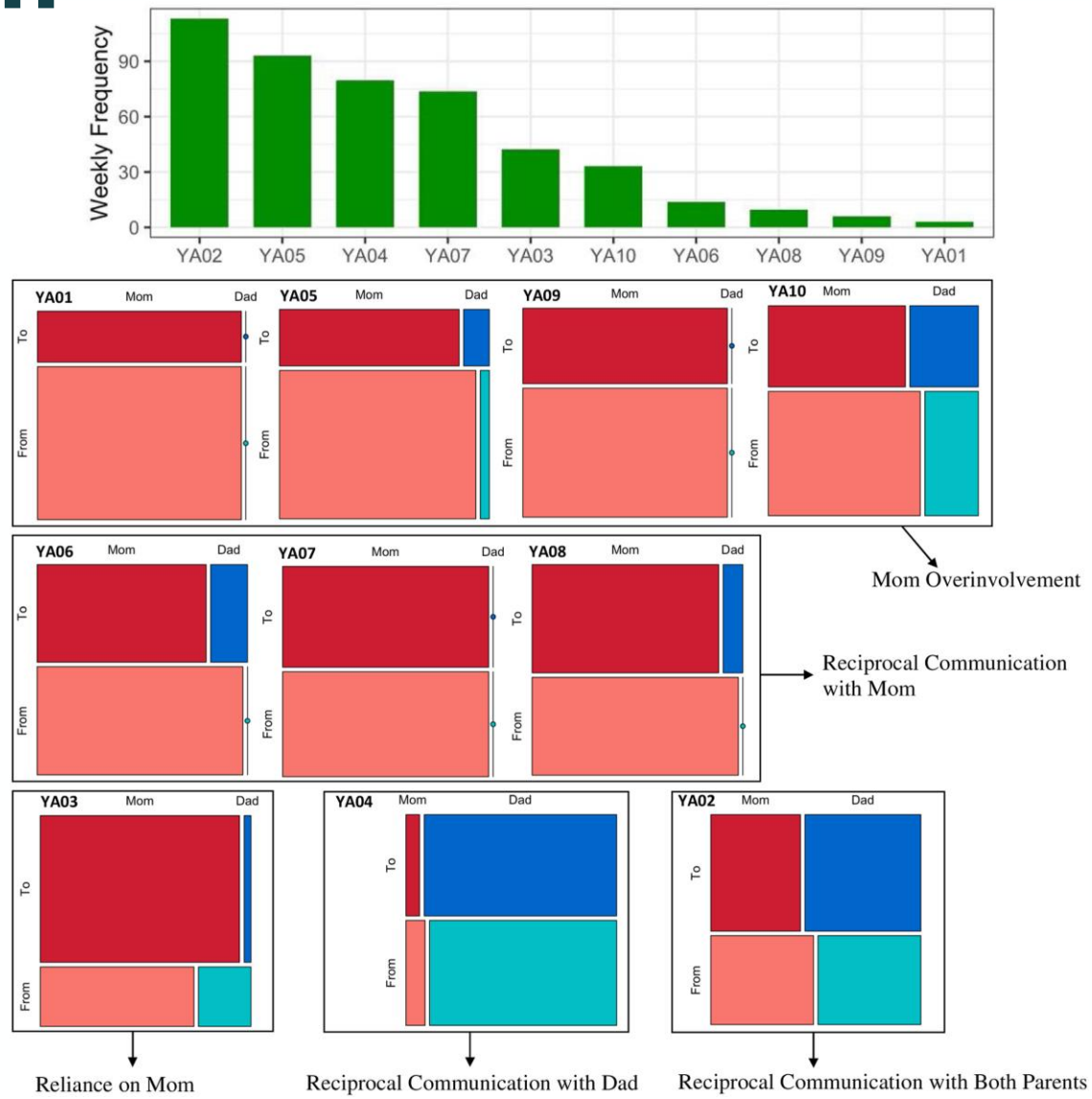
- Activities on smartphones of four different participants
- Blocks represent different apps/activities
- Notice the difference in screentime!



Ram et al. (2019) https://doi.org/10.1177/0743558419883362

Screenomics application

- Connectedness and independence of young adults and parents in the digital world:



See Sun et al. (2023)

Example 4

Capture social media website visits using a plug-in

Browser plug-in Zeeschuimer

- You install a browser plugin developed for a specific platform (for example TikTok).
- When you use the platform of interest (like TikTok) it captures the video's that you see and stores these in a list.
- You share that file with the researcher.

Application: Migration of academic Twitter

- After Twitter was taken over by Elon Musk and many of the changes that followed in content moderation policies, many researchers looked to move to other platforms.
- Among other things, these researchers collected tweets on specific queries ("Mastodon" and "Researcher") using Zeeschuimer.

Fig. 3. Dynamic frequency of Twitter conversations related to academics migrating to Mastodon.

See Wang et al. (2024)

Some definitions

Wearables, apps and sensors

- **Wearable:** A lightweight device including one or multiple sensors that people can wear close to or on their skin, e.g., a fitness bracelet or smartwatch.

- **App:** A software program that runs on a smartphone or other smart device. Apps allow users to interact with a device. For researchers, apps provide the opportunity to collect, aggregate, and process data collected on the device.

- **Sensors:** A piece of hardware that converts a physical phenomenon into an electrical signal. For example, the accelerometer sensor measures the rate of change of velocity of an object.

See Keusch et al. (2024)

Passive and active data collection

- **Passive data collection:** Data are collected, for example, by sensors, as a byproduct of everyday activity (e.g., accelerometer collecting data about walking).

- **Active data collection:** A participant deliberately creates data as a reaction to a stimulus designed by the researcher (e.g., responding to a survey question, taking a picture and sharing it with the researcher).

See Keusch et al. (2024)

Designed and found data

From the perspective of researchers, big data sources are “**found**”. However, they are **“designed” by someone.**

See Salganik (2017)

Exercise:

Of the four examples discussed:

- Was this data collection active or passive?
- Who designed the data?

Exercise

Example 1: Measuring everyday mobility using an app

- Passive / active?
- Who designed the data?

Example 2: Mobility and employment using an app

- Passive / active?
- Who designed the data?

Exercise

Example 3: Connectedness of young adults and parents through Screenomics

- Passive / active?
- Who designed the data?

Example 4: Migration of academic twitter using Zeeschuimer

- Passive / active?
- Who designed the data?


Coffee Break

The sensor approach

Some reflections

What are advantages of this approach?

- Researchers can study human behaviour and social interaction on a much **larger scale** and in a more **naturalistic context** than was ever possible before.
- Phenomena can be measured that were **difficult**, or even **impossible**, to measure using self-report.
- More **detailed data** without increasing participant burden.

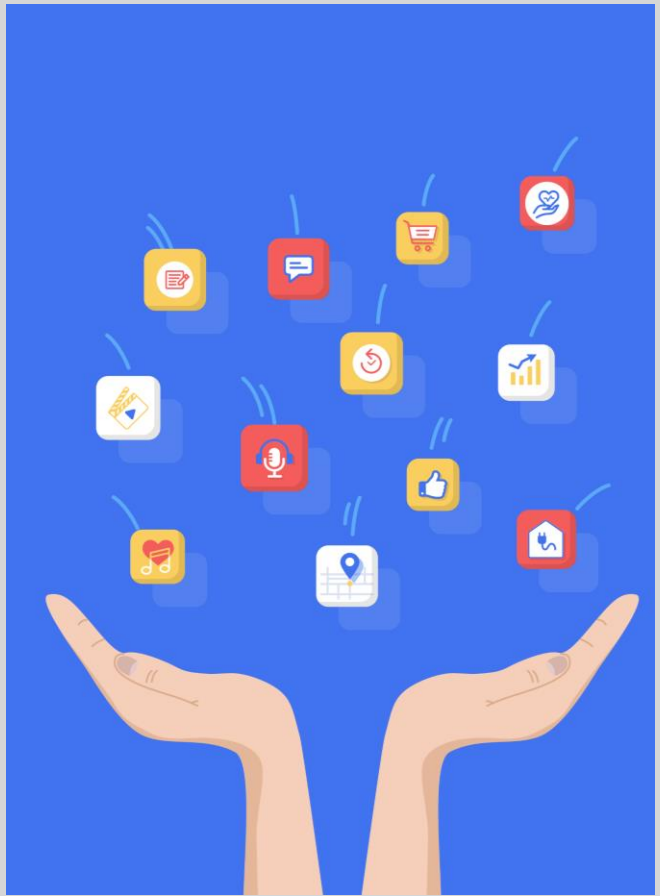
And what are challenges?

- **Coverage:** Those who do not own the device can systematically differ from owners on many variables (potentially even the phenomena that are being measured for a study).
- Sharing devices, or creating apps, is expensive.
- Nonparticipation due to privacy concerns.
- WAS data can contain measurement error:
    - External factors influencing the measurements.
    - Participants interact with the devices.
    - Sensor data needs to be processed to make sense.

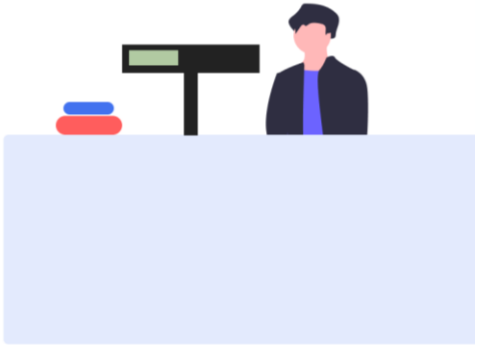
What are ethical considerations?

- Participants might forget they are being observed.
- Information from the contacts of the participant might be revealed. They did not consent.
- Sometimes, consent is given prior to sharing (Screenomics), with others it is given after (Zeeschuimer).

Data donation approaches



With almost everything we do, we leave digital traces behind



GDPR

Article 15 – Right of data access

Right to obtain from a data controller:

- *Confirmation whether personal data are being processed*
- *Access to the personal data*
- *Access to information regarding data recipients and sources and data derived from your personal data*

Image: https://edps.europa.eu/sites/default/files/picture/gdpr2.png

GDPR

Article 20 – Right of data portability

Grants data subjects the right to

- *receive the personal data in a structured, commonly used and machine-readable format (“Data Download Package”)*
- *transmit those data to another data controller*

Image: https://freeicons.io/

Data donation

Donation of digital traces collected as Data

Download Packages (DDPs) for research purposes.
See: https://datadonation.eu/

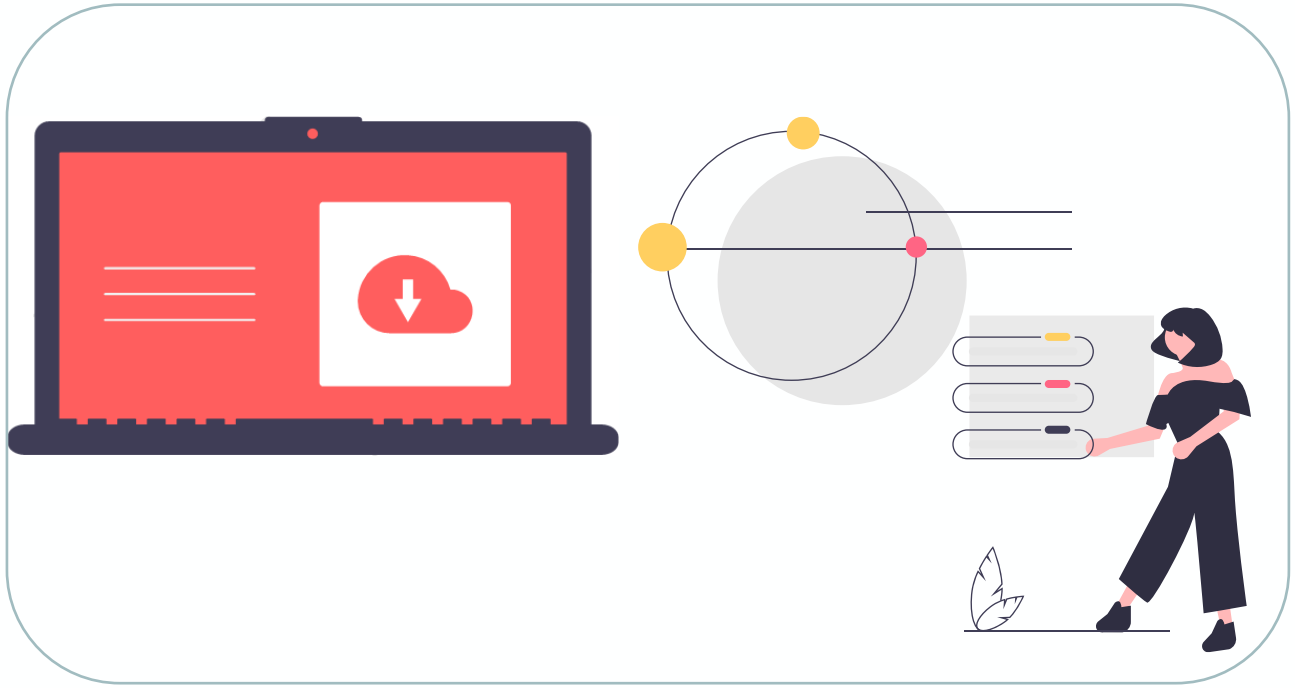
Image: https://www.webaccessglobal.com/sites/default/files/home-page-slideshow/images/13.png

Data donation workflow

Online platform

Device of participant

Researcher environment



Images: https://vecteezy.com

Three approaches to acquire the donations

1. Donation of complete DDPs.
   - Often, de-identification protocols are in place as soon as the data comes in.
   - Particularly adopted in early data donation studies.
2. Donation of digital traces after local processing.
   - Currently most used approach.
3. Facilitating data access through creating a “data lake” of data donations.
   - Examples are currently in the making, but no output yet (see e.g. Data4Science).

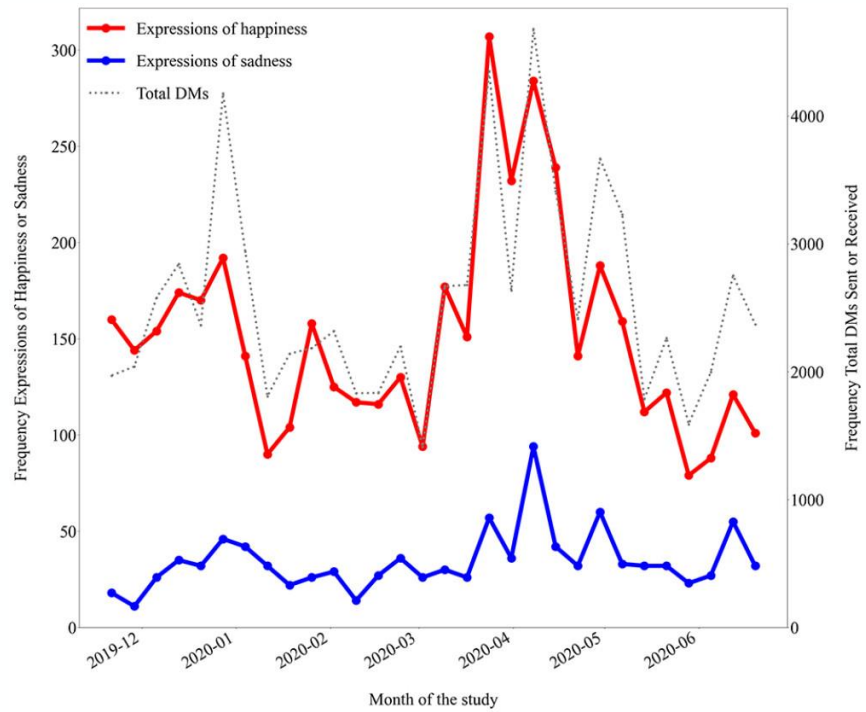
Example 1

Happiness and sadness in adolescent's Instagram direct messages

Measurement Burst Design (Nesselroade, 1991)

N = 388 (age 13-15)

T > 250



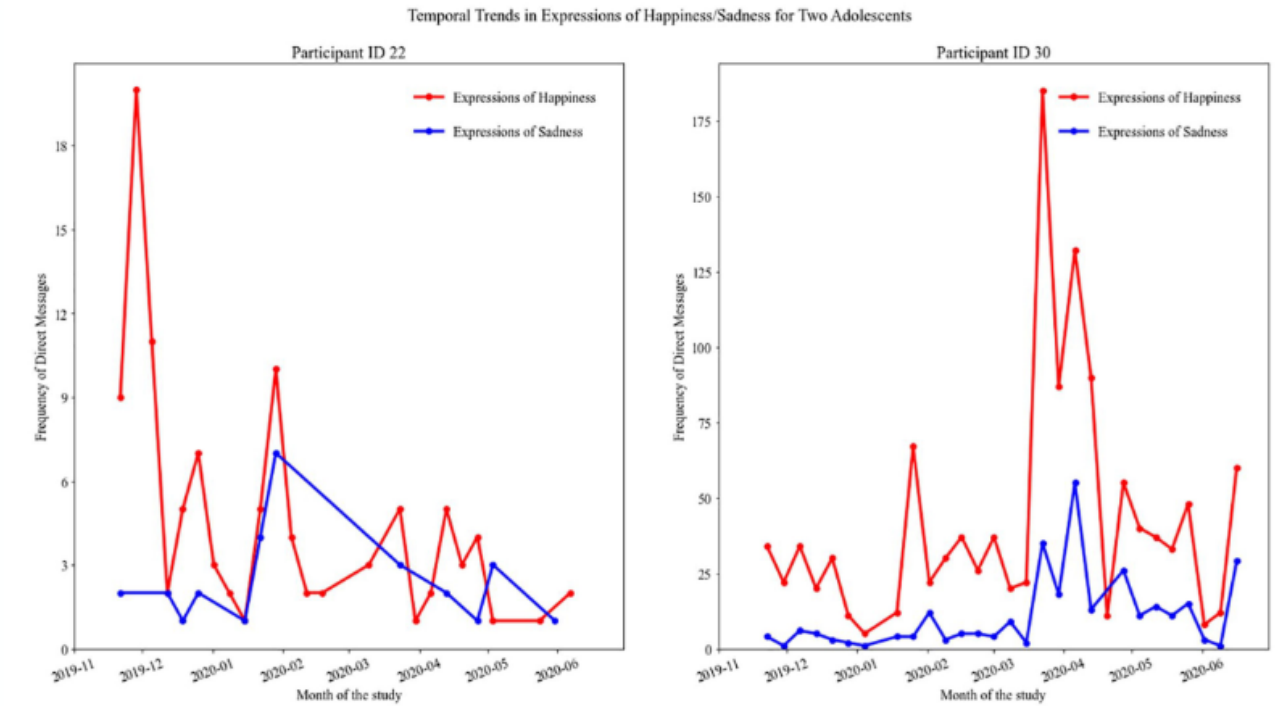


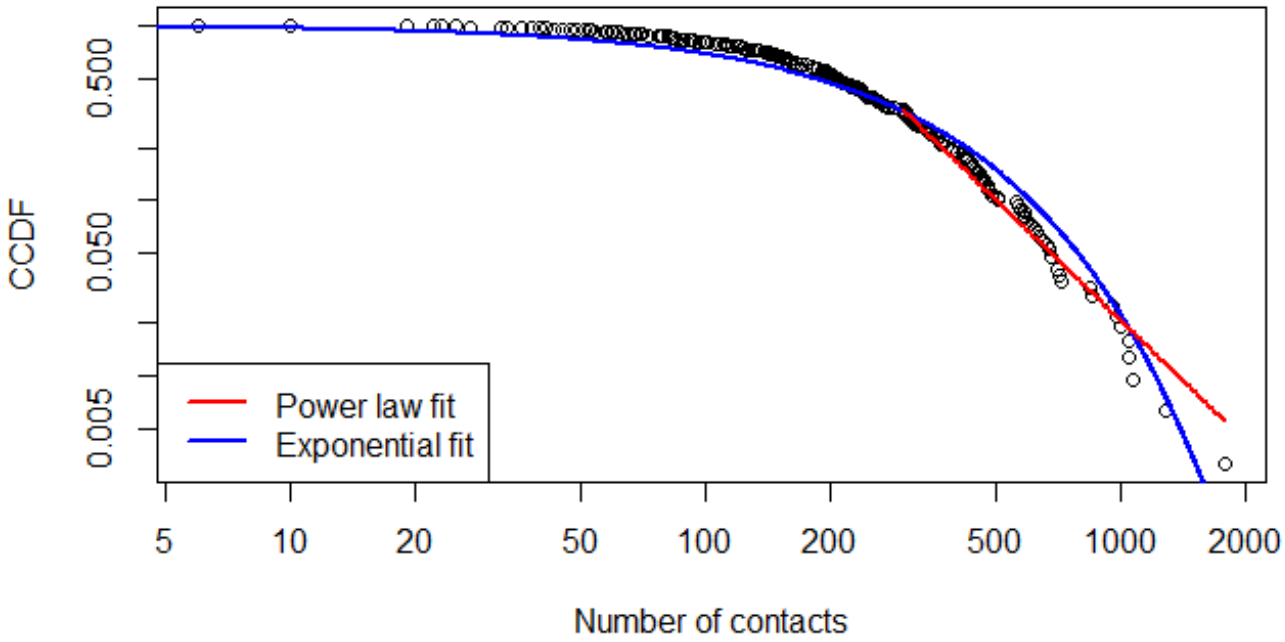
Figure 4. Temporal trends in the frequency of direct messages containing expressions of happiness (red line) and sadness (blue line) for two different participants.

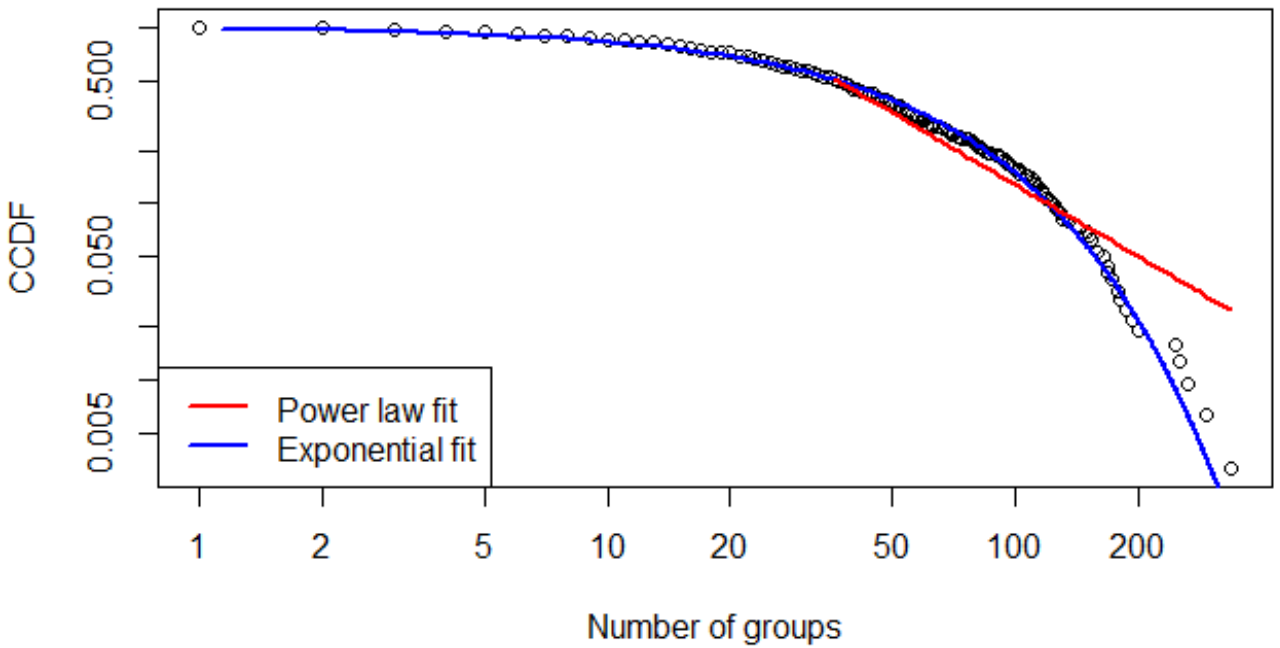
Note. The *y*-axis differs per plot.

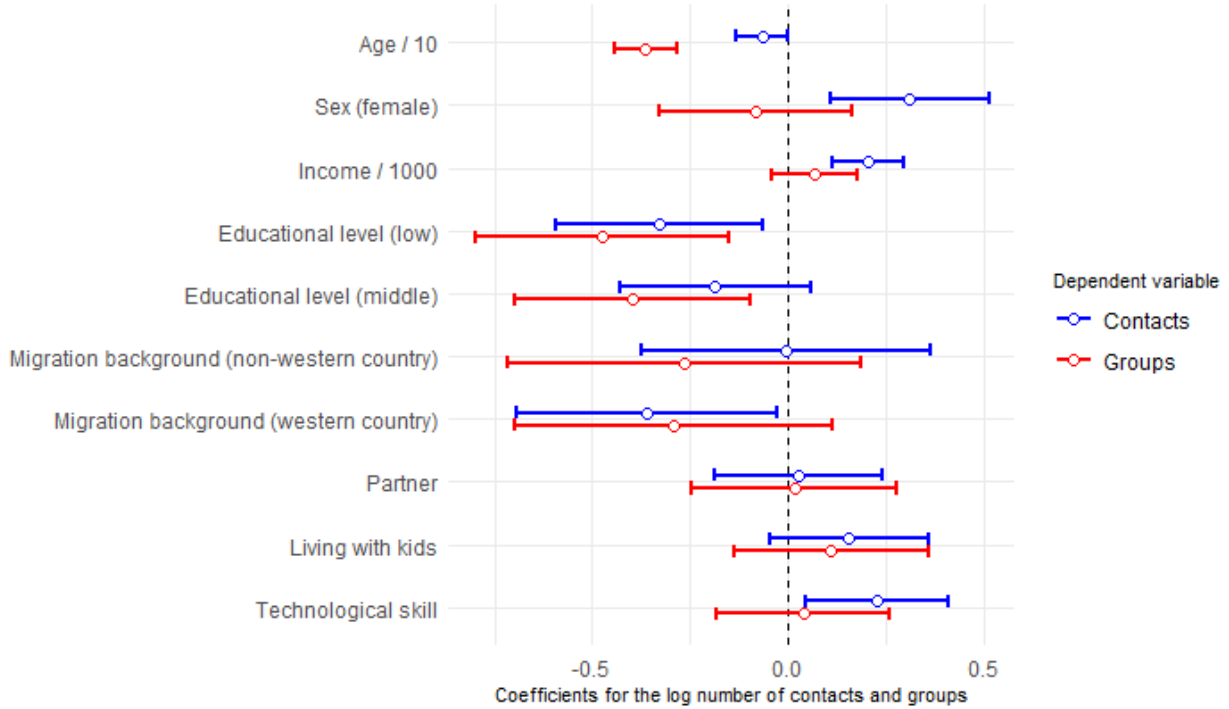
See Verbeij et al. (2024)

Example 2

Composition and structure of an Instant Messaging network







See Corten et al (2024)

Example 3

Sensitivity and Intimacy with Google Assistant Users

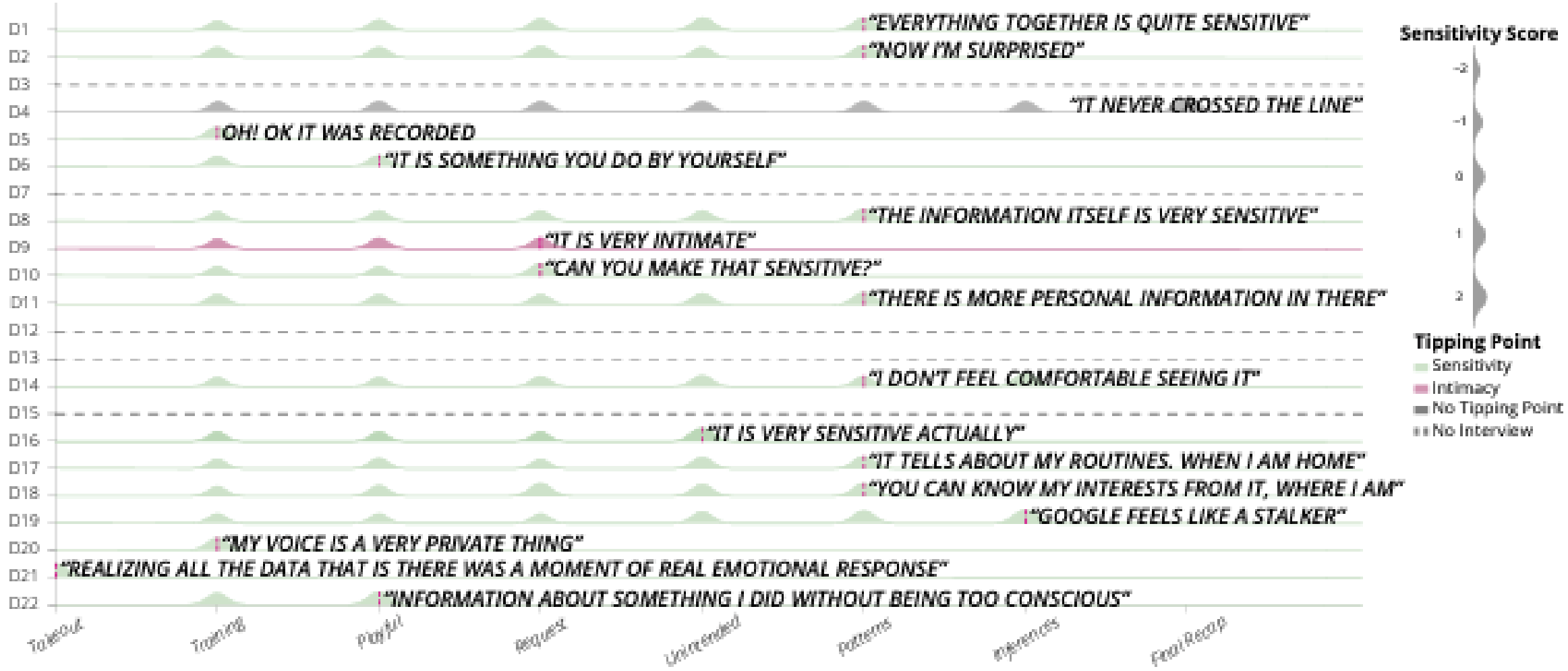


Figure 7: Tipping point in the course of data donation process and data-centric interviews.

Exercise:

- Of the examples discussed:
  - Was this data collection active or passive?
  - Who designed the data?

Exercise

Example 1: Happiness and sadness in Instagram DMs

- Passive / active?
- Who designed the data?

Example 2: Composition and structure of an Instant Messaging network

- Passive / active?
- Who designed the data?

Example 3: Sensitivity and intimacy with Google Assistant Users

- Passive / active?
- Who designed the data?

Definitions

GDPR Definitions

- **Personal data:** Information relating to an identified or identifiable natural person.
- **Data subject:** The natural person that the personal data refer to.
- **Data controller:** The person or organization responsible for processing personal data. The controller decides which data will be processed, how and why.

Data donation definitions

- **Data Download Package (DDP):** Because of the right of data access, data subjects are always allowed to retrieve their personal data from data controllers. Data controllers are obliged to comply with such requests and because of the right of data portability, provide the requested data in a machine-readable format. To comply with these regulations, large data controllers such as social media platforms typically provide data subjects with a .zip file containing the personal data requested.

- **Local processing:** Locally at the device of a research participant, data is extracted from a DDP that is relevant for a particular research question under investigation.

What are advantages of data donation?

- Researchers can study human behaviour and social interaction on a much larger scale and in a more naturalistic context than was ever possible before.
- Phenomena can be measured that were difficult, or even impossible, to measure using self-report.
- All personal data collected by any data controller can be obtained for research purposes.
- Data can be collected in retrospective.
- Data is not influenced by participants “feeling seen”.
- Digital traces can be supplemented with other sources of data.
- Participants can look at and potentially even interact with their data.

And what are challenges?

- Each study is different and needs a custom script (this is often still a lot cheaper compared to creating a completely new app).
- Substantive effort from participants is required.
- Platforms can make it difficult to facilitate data donation.
- Not all participants have the digital skills required to participate.
- Participants can be concerned about their privacy.
- The population of interest does not always all use the platform chosen.

What are ethical considerations?

- Participants can inspect their data; participants can give their “true” informed consent.
    - → Do participants always have the capacity to oversee large amounts of data?
- Local processing can prevent data from the contacts of the participant to be shared.
    - → This does require substantive knowledge from the individual researcher.


Questions?

After lunch:

Labmeeting:

- Learn how to work with Zeeschuimer and understand how data collected using Zeeschuimer looks like.
- Explore your own digital traces using the “Digital Footprints Explorer”.
- Questions about literature.

Next week:
Platform-centric approaches