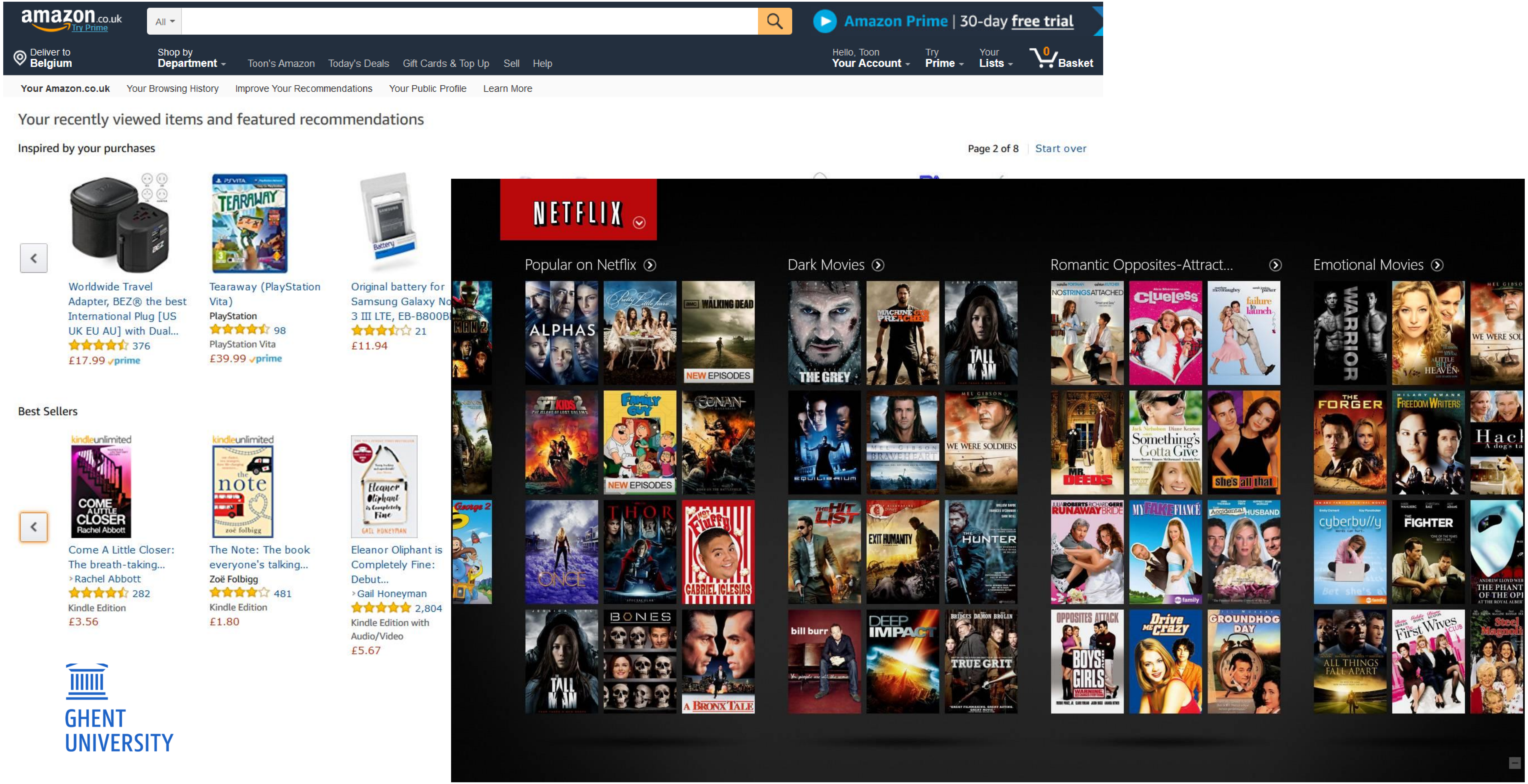
RECOMMENDER SYSTEMS,
THE MAGIC HIDDEN BEHIND THE INTERFACE OF ONLINE SERVICES
OR A CLEVER INTERPLAY BETWEEN SMART ALGORITHMS AND INNOVATIVE HUMAN-COMPUTER INTERACTION

Toon De Pessemier – May 23, 2018

WHAT ARE RECOMMENDER SYSTEMS?



GHENT UNIVERSITY

OVERVIEW

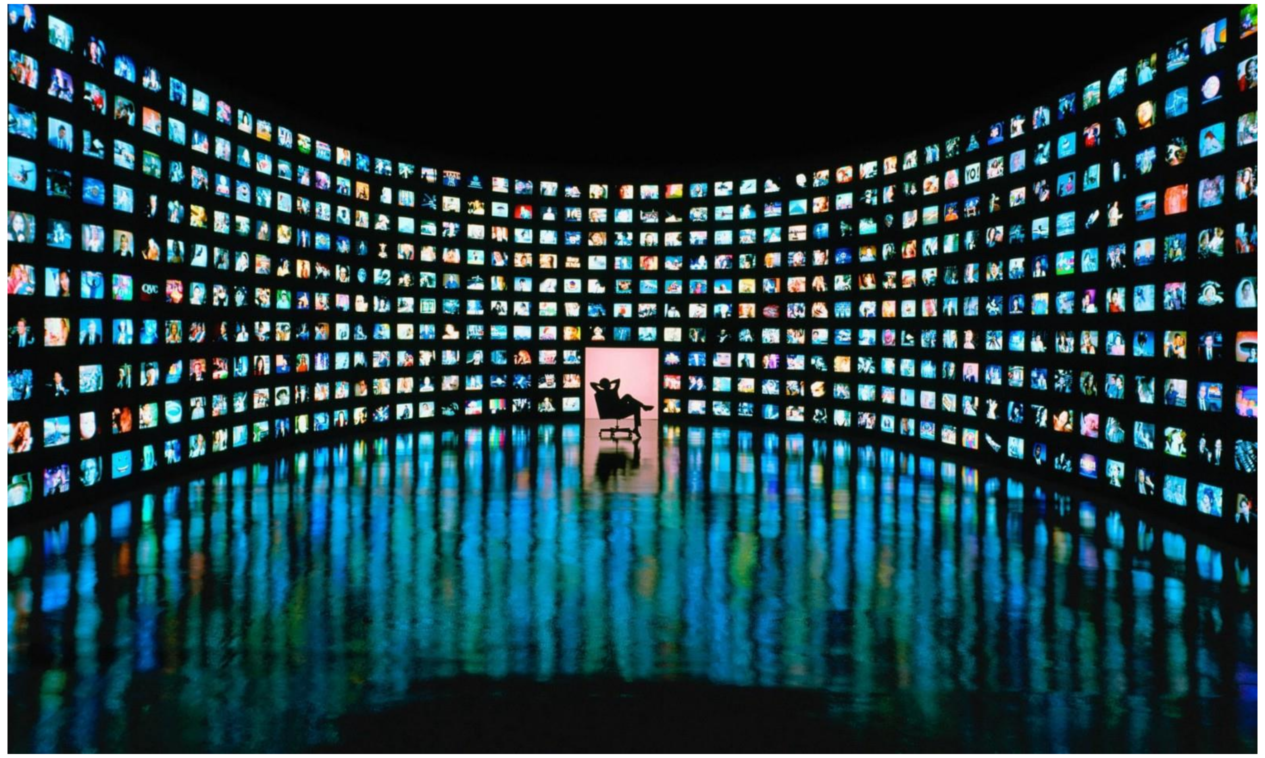
- Why do we need recommender systems?
- Preferences, Ratings, Gathering feedback
- Types of recommendations
  - Non-personalized recommenders
  - Content-based filtering
  - Collaborative filtering
  - Hybrid recommenders
- Evaluating recommendations
- Group recommendations

WHY RECOMMENDATIONS ? FROM SCARCITY TO ABUNDANCE

- Web enables near-zero-cost dissemination of information about products

Physical world

Online world

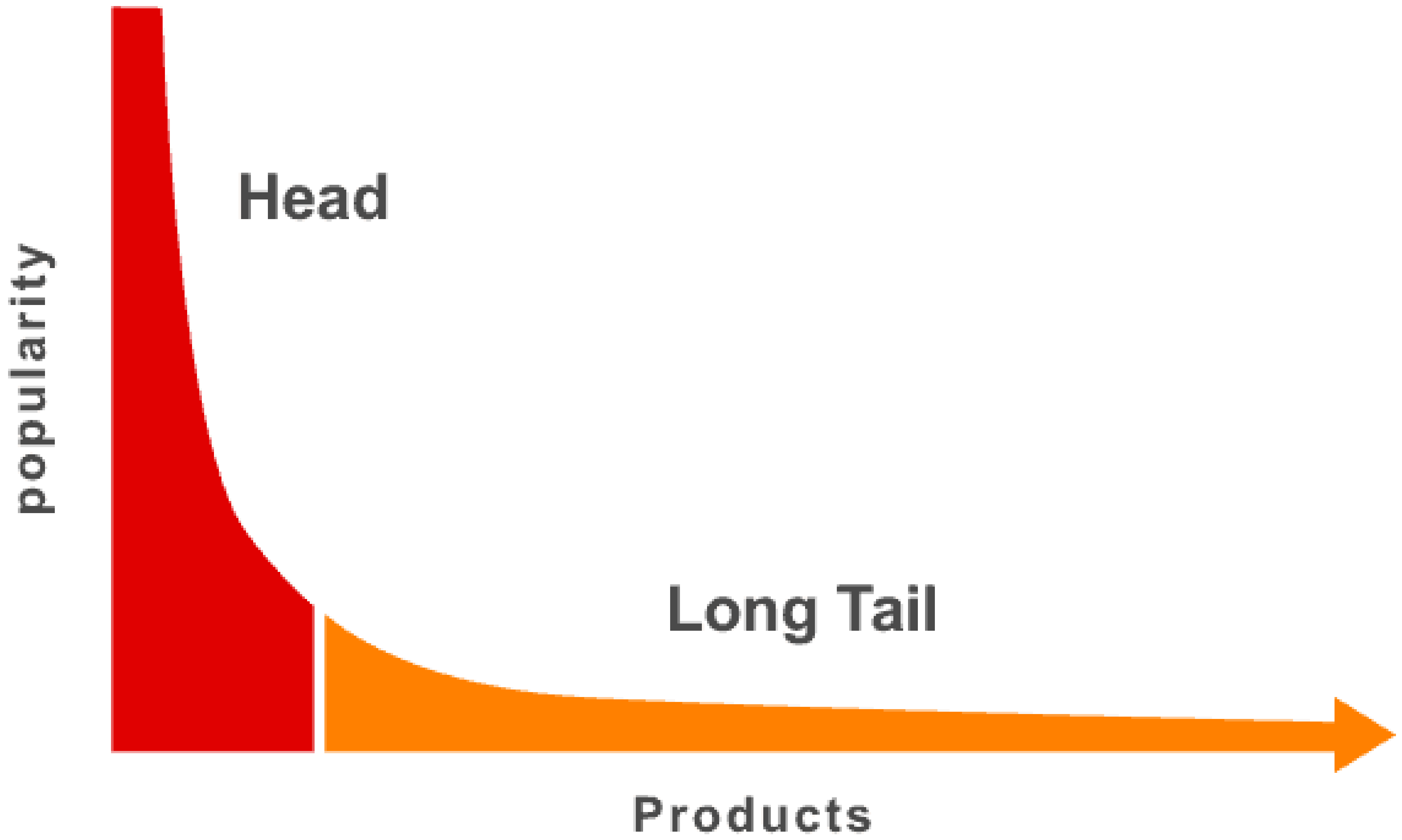


- A bigger catalog requires better filters
  – Personalized recommendations

GHENT UNIVERSITY

WHY RECOMMENDATIONS ? THE LONG TAIL

- Head: highly popular products. Typical in physical stores
- Long Tail: an infinite offer of niche products. Typically distributed online



SITUATING RECOMMENDER SYSTEMS

Two sides of the same coin

Information Retrieval (IR)

- E.g. Google search
- Information need expressed through a **query**
- Goal: **Retrieve information** which might be useful
- Rather static content base ➔ **indexing** content
- **Dynamic information need**: real-time queries

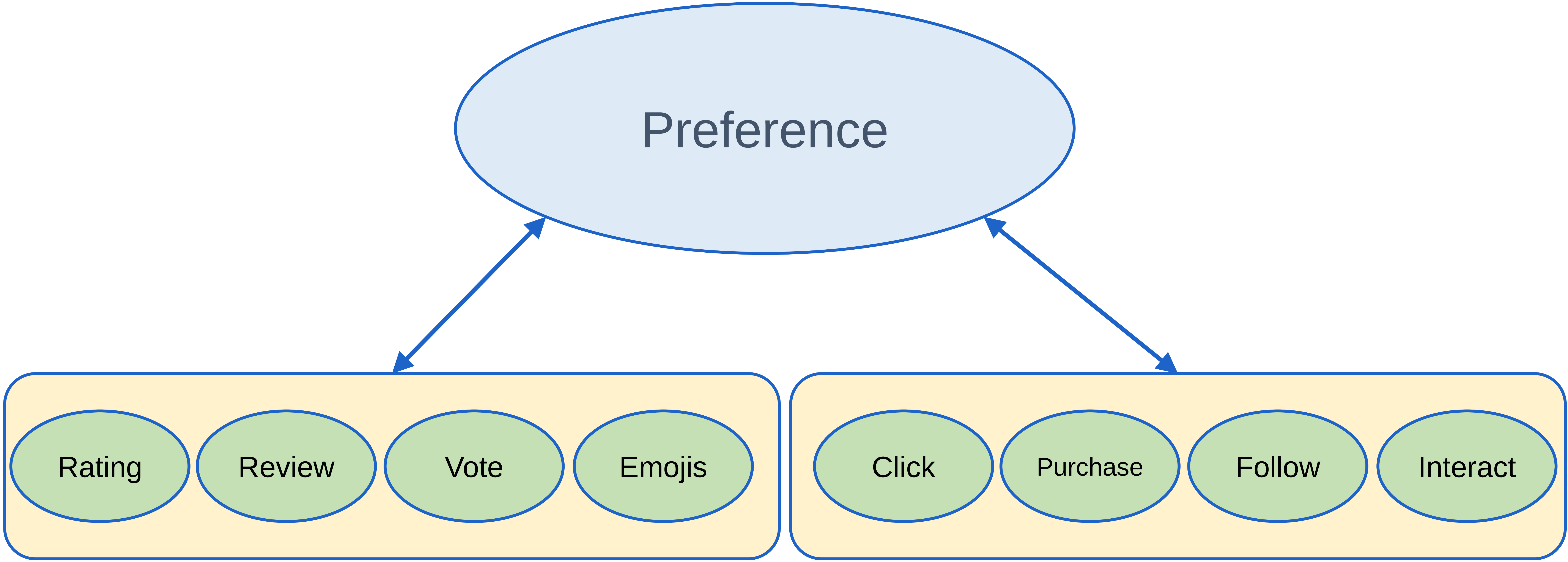
Information Filtering (IF)

- E.g. Filtering news
- Information need expressed through a **user profile**
- Goal: expose users to only the **information that is relevant** to them, according to their personal profile
- Reverse characteristics from IR
- Rather static information need
- **Dynamic content base**
- Invest effort in **modeling user** need
    - Hand-created "**profile**"
    - Machine learned profile
    - Feedback / updates
- Pass new content through **profile filter**

PREFERENCES, RATINGS, GATHERING FEEDBACK

GHENT UNIVERSITY

PREFERENCES: THE FUEL OF THE RECOMMENDER ENGINE



GHENT UNIVERSITY

Explicit

Users express what they think

Implicit

Users performs an action of the service

Not (necessarily) intending to communicate how much they like

DIFFICULTIES WITH RATINGS

- Meaning for the user
  - Scale: how much is “rather good” on a scale of 10?
  - Rating of content, (a/v) quality, service (e.g. delivery of goods)

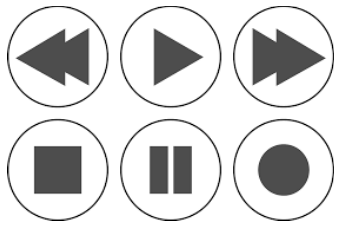
- Psychologic aspects
  - Noise in the psychological process of giving ratings
- Different personal intensions
  - Provide opinion, improve recommendations, influence top list
- Users often skip the rating process
  - Giving a rating is boring
  - Cognitive burden
  - No clear incentive for the user

IMPLICIT PREFERENCES FROM USER ACTIONS

- Data collection from **actions the user performs** (other than expressing preferences)
  - Interaction with the service, content, other users, …
- Often gathered on websites where
  - explicit feedback is **less common** (e.g. online advertisement)
  - **less desired** because of user experience (e.g. streaming music services)

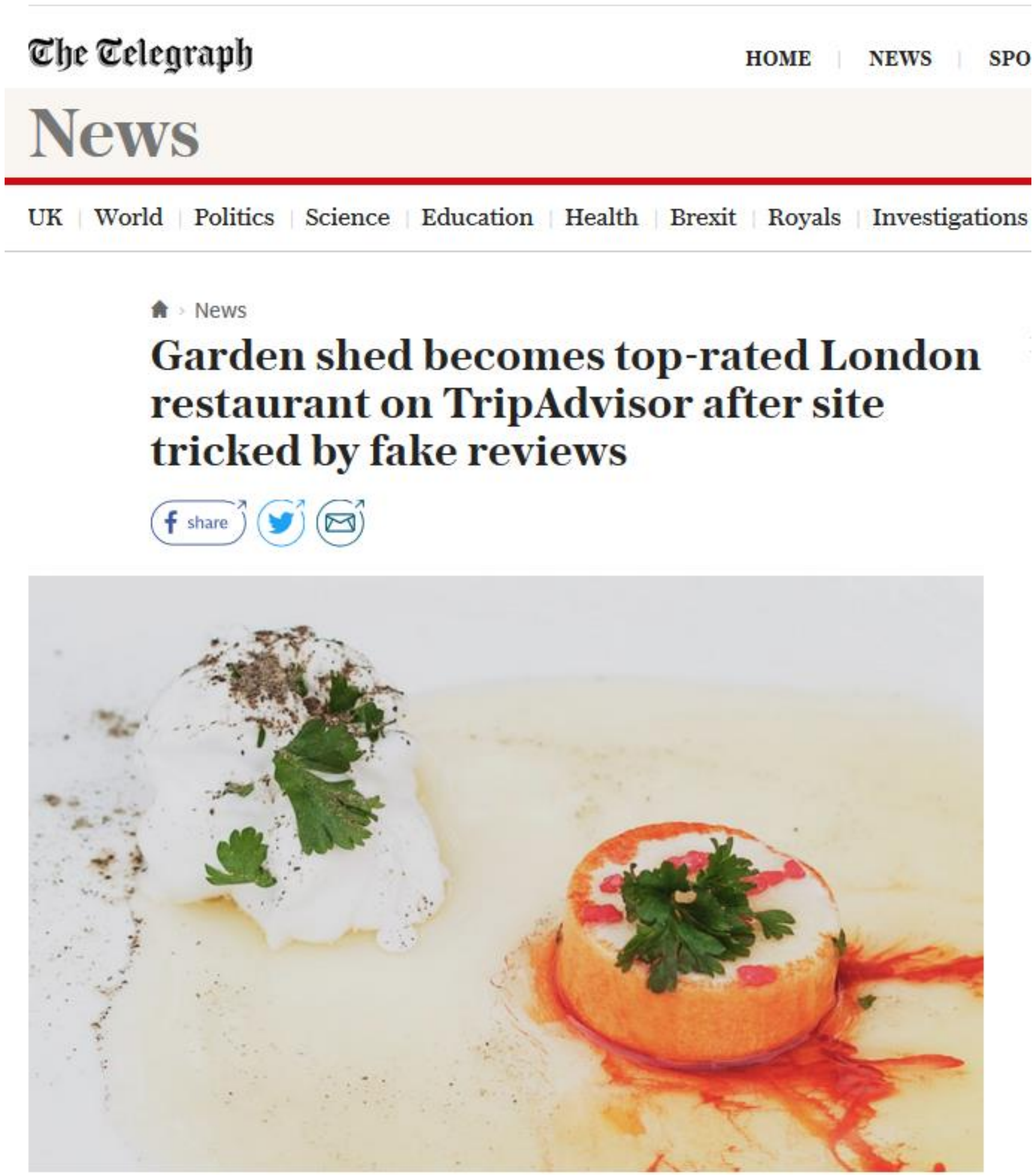
- **More implicit** feedback than explicit feedback
- Can be **complementary** to explicit feedback
- **Reading time**: How long does a user spend on a website?
  - Correlation with interest in the page

- Video **watching time**, music listening time
  - Skipping content, fast-forward, listening/watching twice,…



TRIPADVISOR EXAMPLE

- Ratings/reviews reliable?
- Freelance writer created a **fake restaurant** on TripAdvisor
- Pushed his **own backyard** as a restaurant to the top
- Fake restaurant
  - Only telephone number
  - An appointment-only restaurant
- It became the **best restaurant of London**



The Telegraph — HOME | NEWS | SPO

News

UK | World | Politics | Science | Education | Health | Brexit | Royals | Investigations

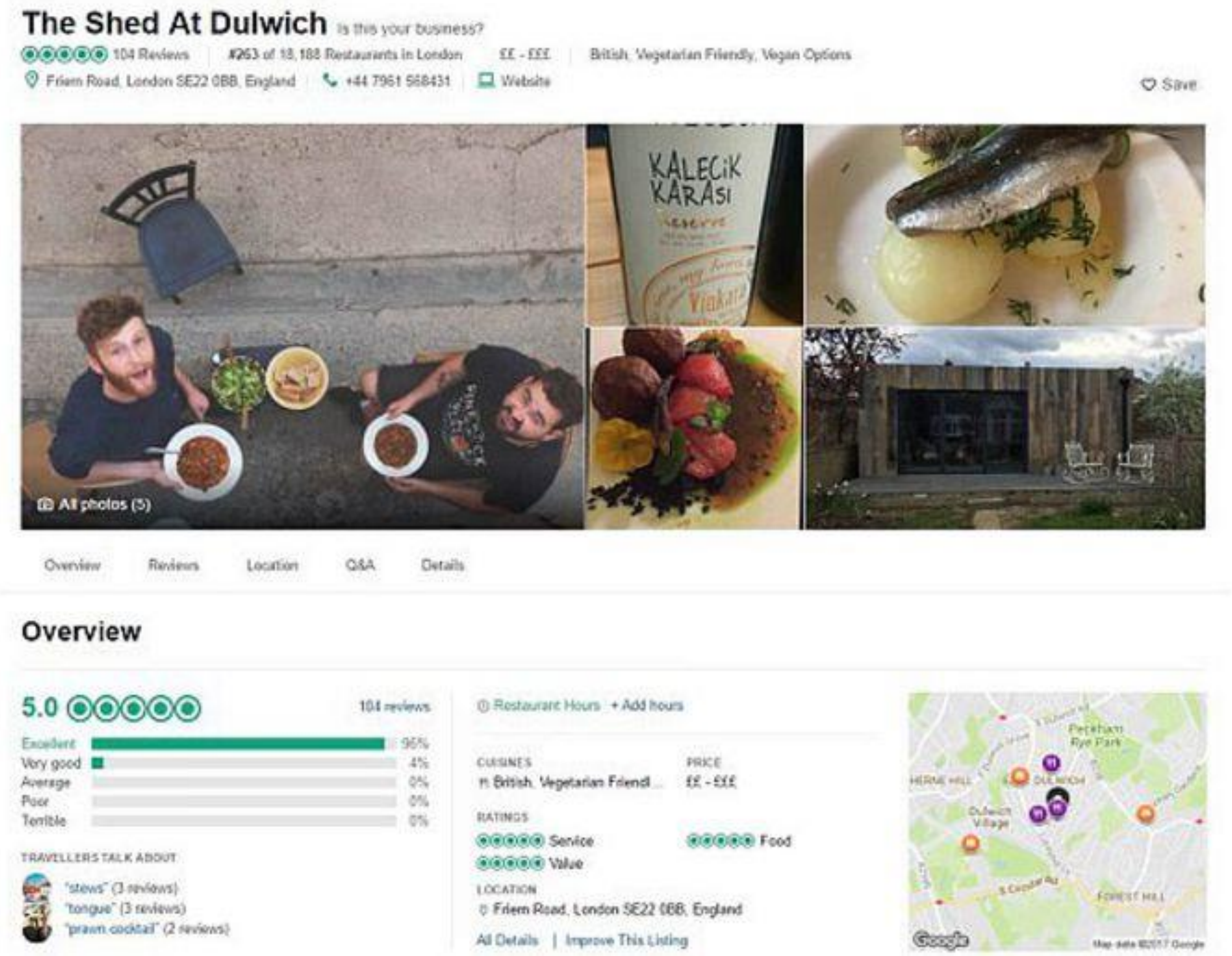
News

Garden shed becomes top-rated London restaurant on TripAdvisor after site tricked by fake reviews

share

GHENT UNIVERSITY

TRIPADVISOR EXAMPLE



GHENT UNIVERSITY

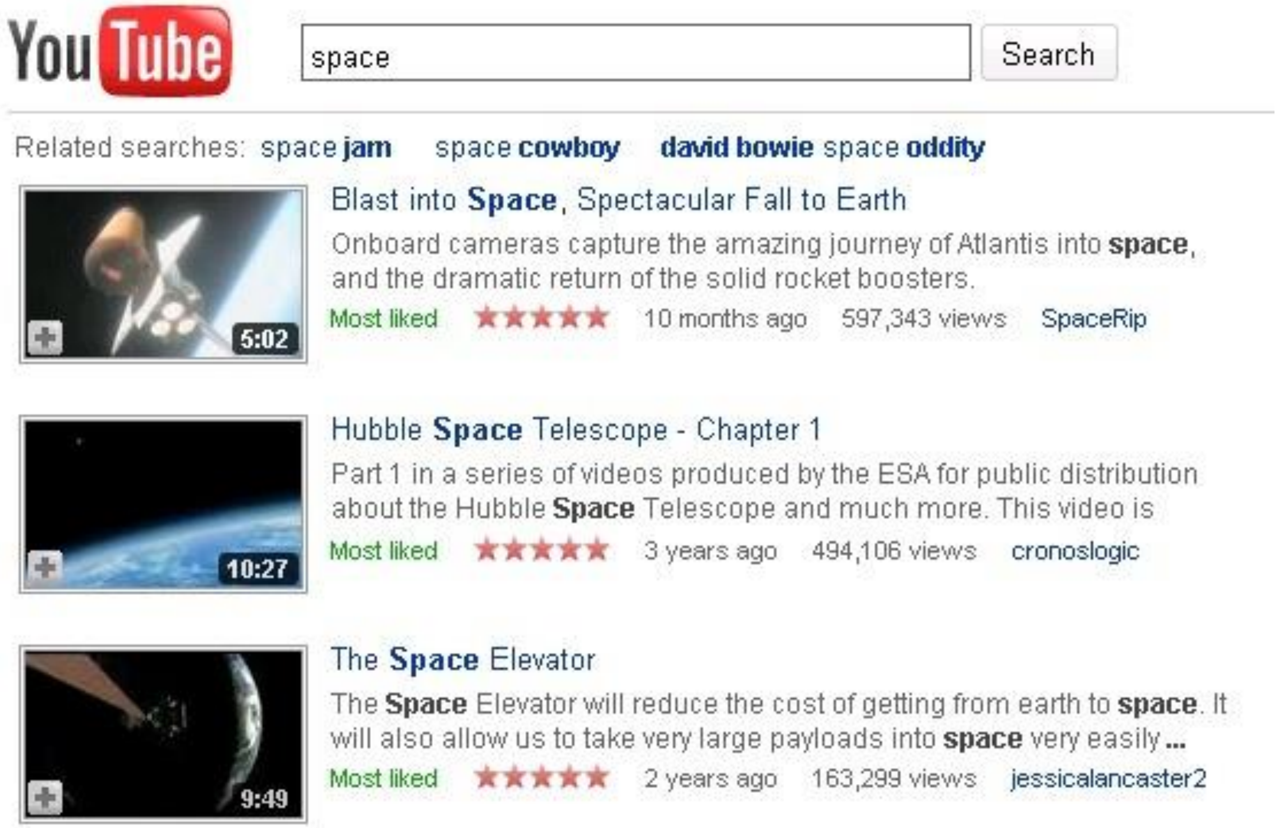
TRIPADVISOR EXAMPLE

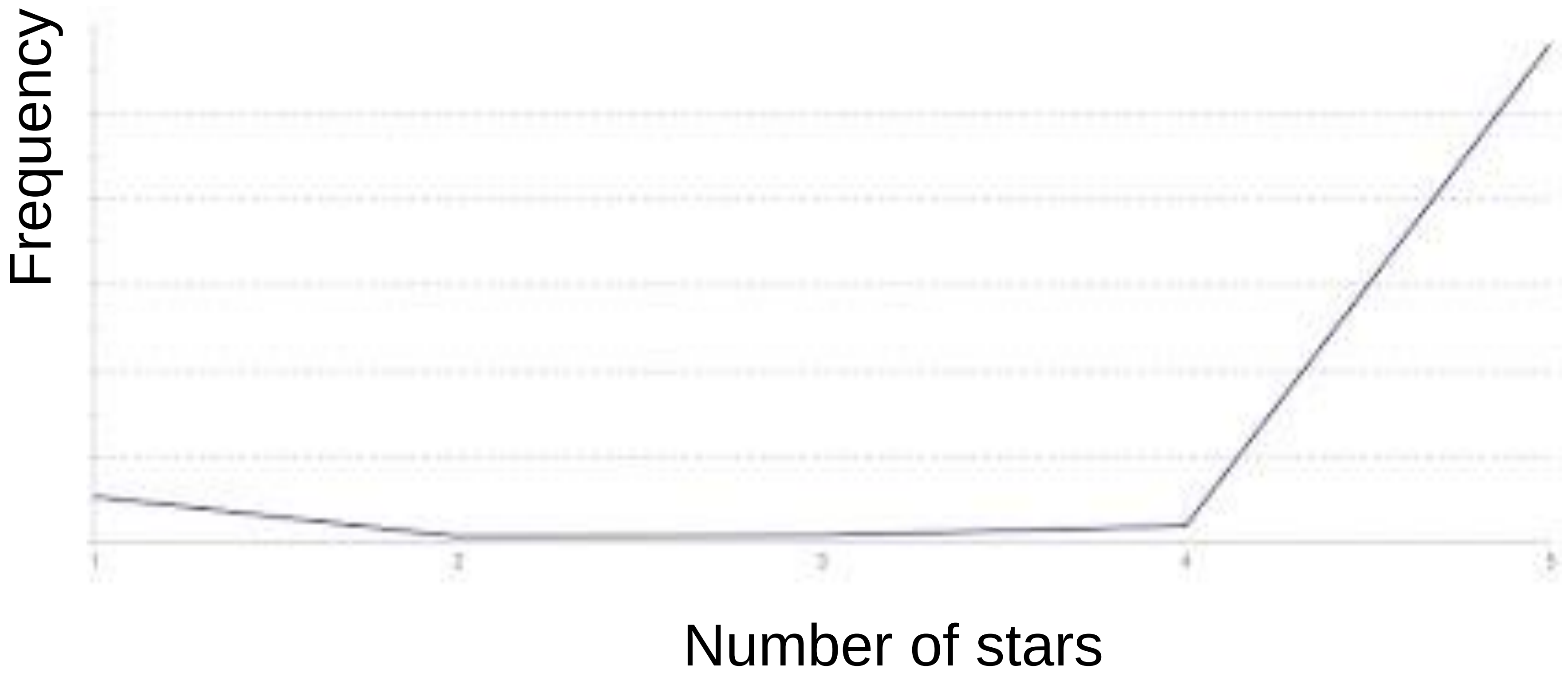
- Fake ratings & reviews
  - Illustrated with photos
    - Photographs of the "food" - close-ups of shaving foam and bleach
  - Different accounts and devices
- **Limited credibility check** of TripAdvisor
- **Attacks** on recommender systems
  - Fake ratings, fake reviews
    - Boost your own business
    - Counteract a competitor
- Nowadays: Many services try to **detect** fake ratings

GHENT UNIVERSITY

YOUTUBE EXAMPLE (2009)

- Five Stars Dominate Ratings





- Are all YouTube videos so good ?
- Reason: **Great videos** prompt **action**; anything less prompts indifference
- Rating system = seal of approval

  ≠ an editorial indicator of what the community thinks
- How useful is this system really?

GHENT UNIVERSITY

https://youtube.googleblog.com/2009/09/five-stars-dominate-ratings.html

EXPLICIT FEEDBACK: AS EASY AS POSSIBLE

- Counters in public spaces
- E.g. Evaluation of the infrastructure/services at airports
- Easy to use
- Fast & simple
- Basic feedback
- High participation rate
- Disadvantage:
  - No user identification or demographics

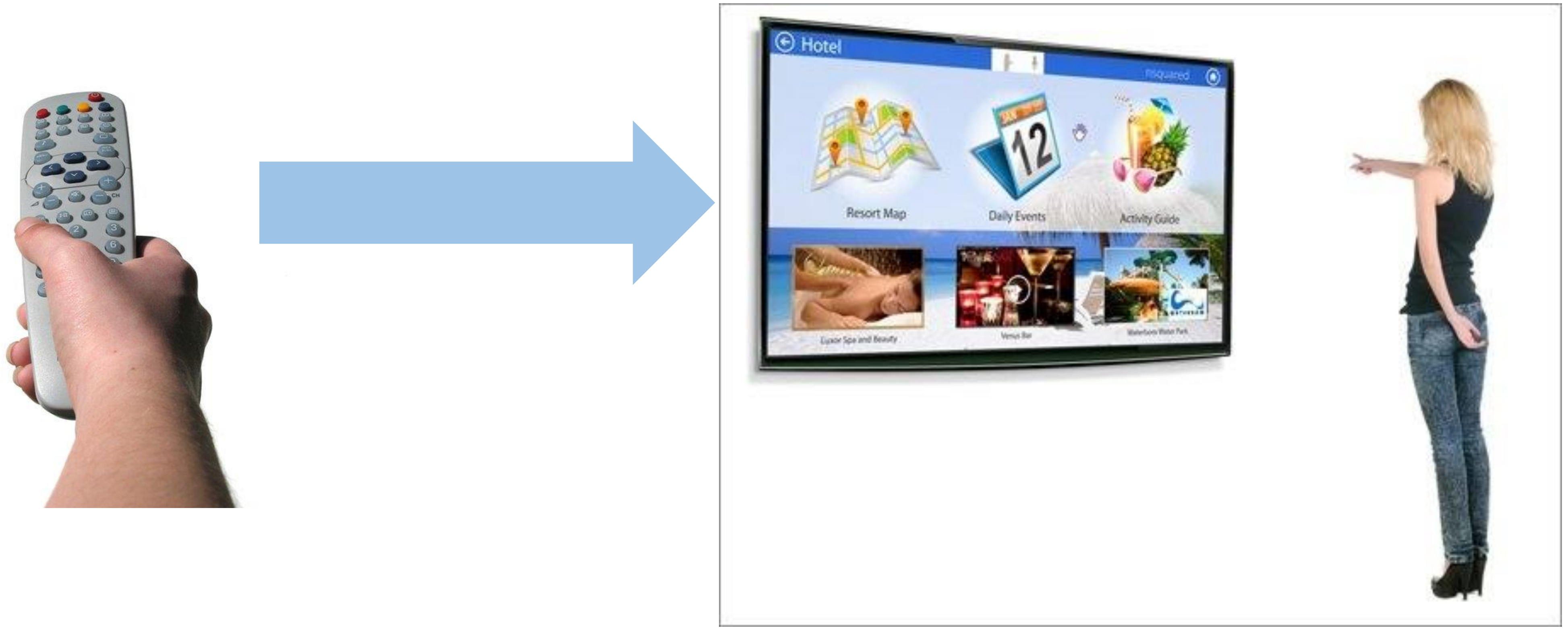
GHENT UNIVERSITY

INNOVATIVE METHODS TO GATHER FEEDBACK

- Not all users of (online) systems are behind a computer
- TV environment
  - Ratings with remote control?
  - Camera as a solution for intuitive human-device interaction

Also for:

- Video delivering systems: video control
- Content library: browsing & selection



GHENT UNIVERSITY

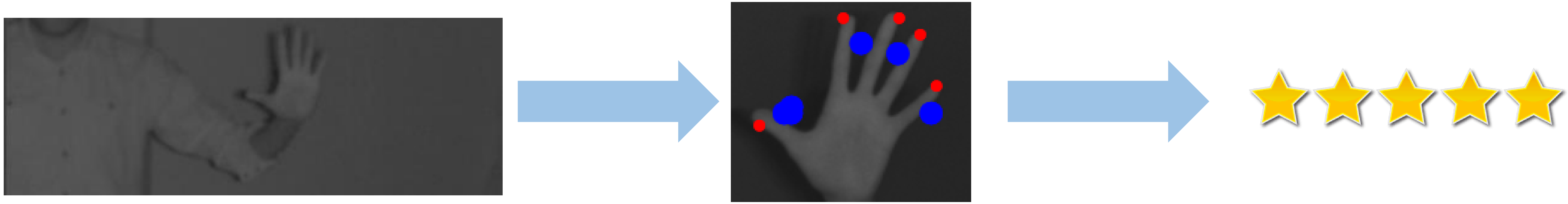
INTUITIVE HUMAN-DEVICE INTERACTIVITY

Example: Microsoft Kinect as motion sensing input device

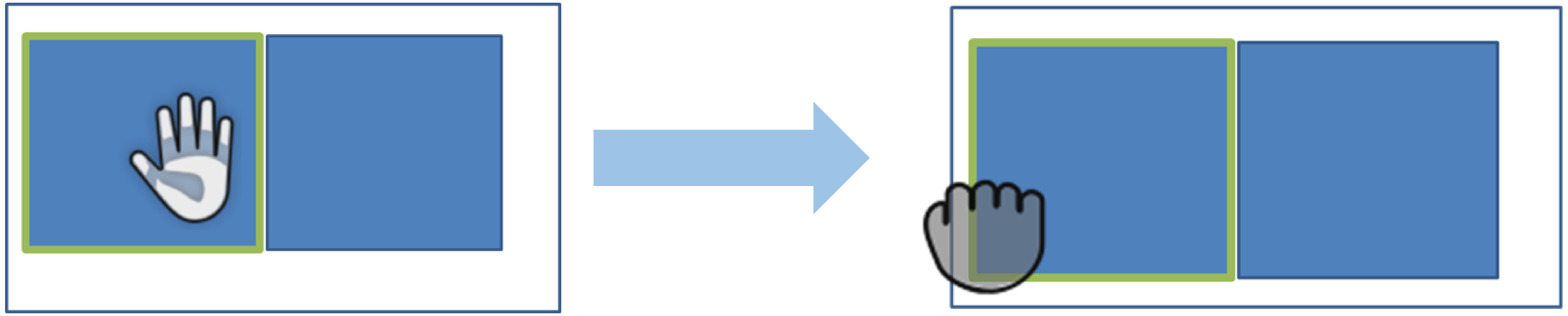
- **Speech recognition**: video control
- **Text-to-speech**: feedback to the user
- **Tracking movements**: browsing, content selection, explicit feedback
- **Facial recognition**: user identification (authentication)
- **Localization of a sound source**: context detection
- **Emotion recognition**: implicit feedback for the content
- **Body position**: implicit feedback (engagement)

INTUITIVE INTERACTION WITH A CAMERA: EXPLICIT FEEDBACK

Finger tracking: Raise a number of fingers to give a specific score



Drag and drop: Drag an object to a specific area to give a score



GHENT UNIVERSITY

INTUITIVE INTERACTION WITH A CAMERA: EXPLICIT FEEDBACK

Hand recognition: Raise the hand until a specific counter is reached to give a score

+ 1 2 3 4 5

Speech recognition: Users can “say” the number of stars

INTUITIVE INTERACTION WITH A CAMERA: EXPLICIT FEEDBACK

Hand Writing: Writing the number in the air

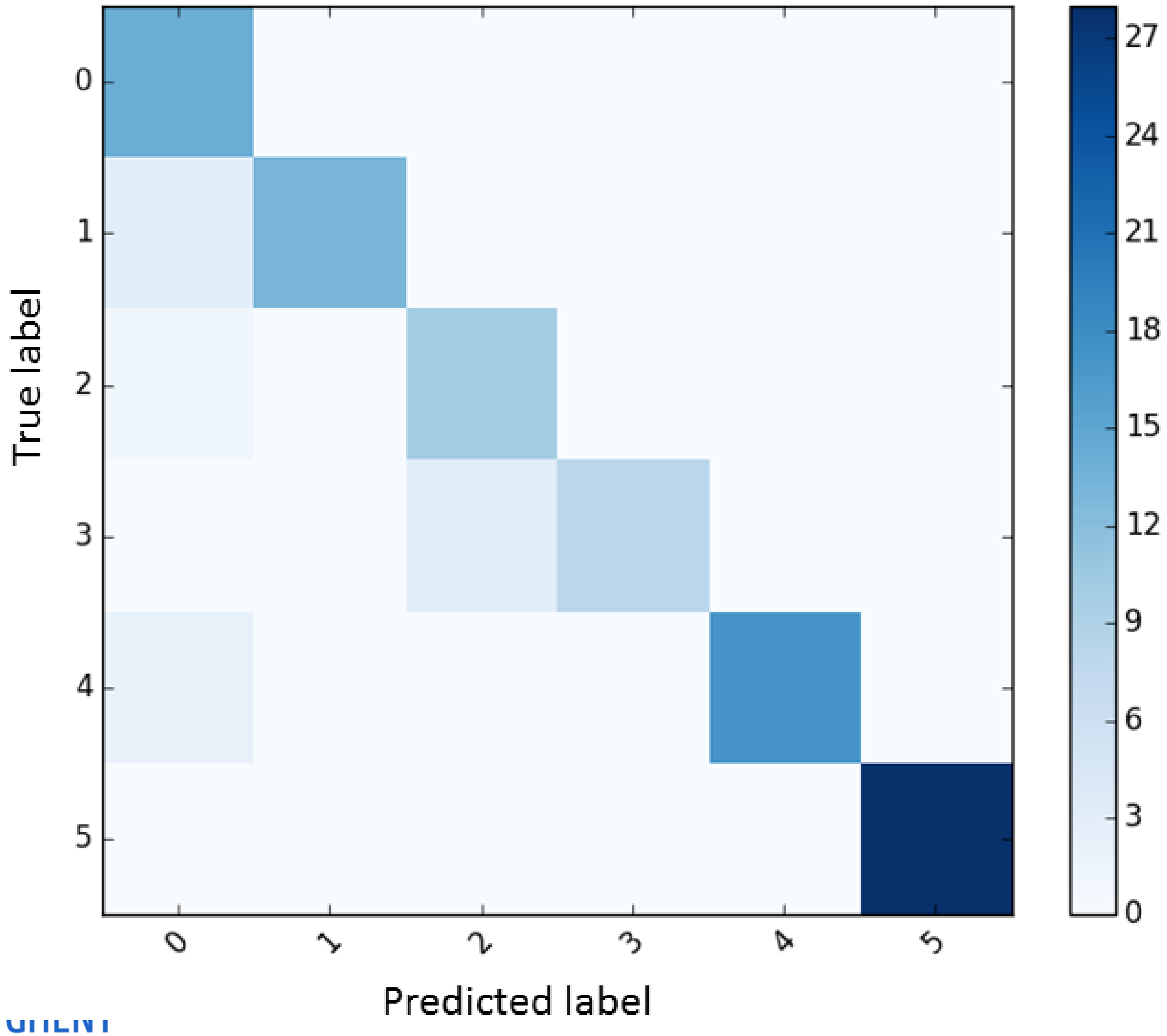
PROBLEM OF INTUITIVE INTERACTION

- Accuracy of the detection method
- Most accurate results in optimal conditions:
  - Frontal view
  - Uniform background
  - No noise (speech recognition)
  - Sufficient light
    - No shadows
    - No reflections
  - Not too close, not too far from the camera
    - Depending on the camera type
    - Microsoft Kinect experiment: between 1 and 2.5 meters

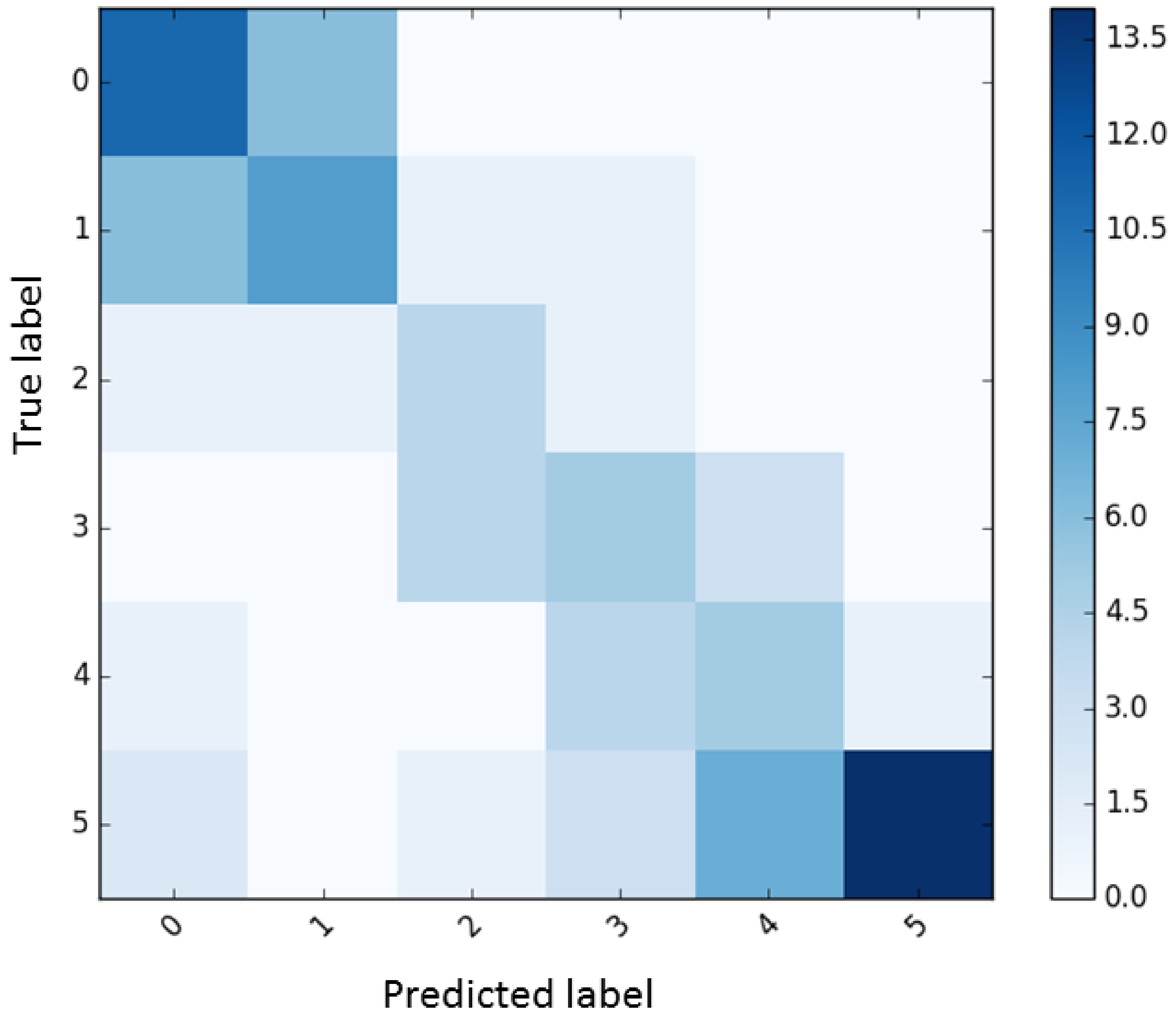
GHENT UNIVERSITY

FINGER TRACKING

Distance: 1 to 2.5 meters

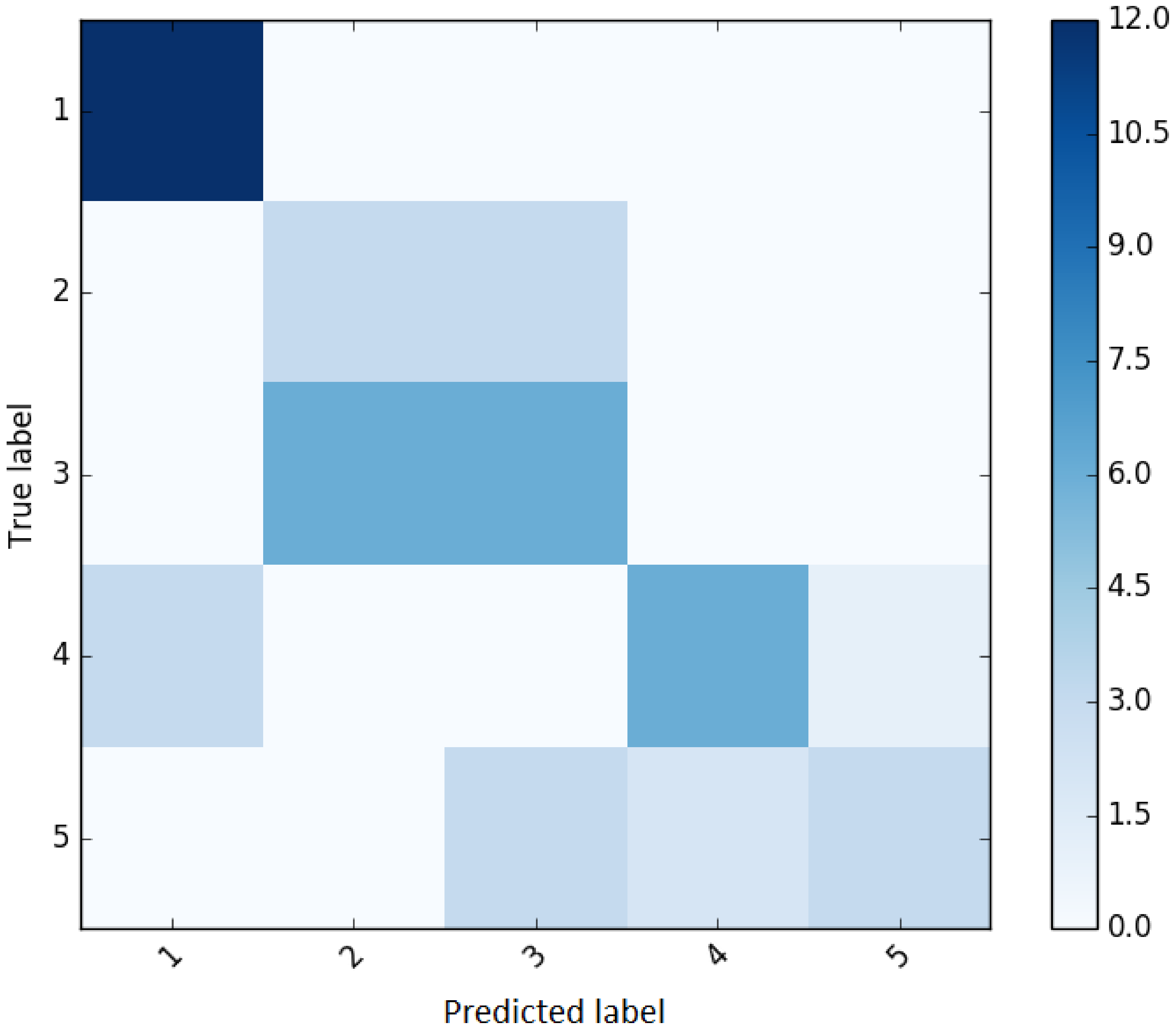


Distance: further than 2.5 meters



HAND WRITING

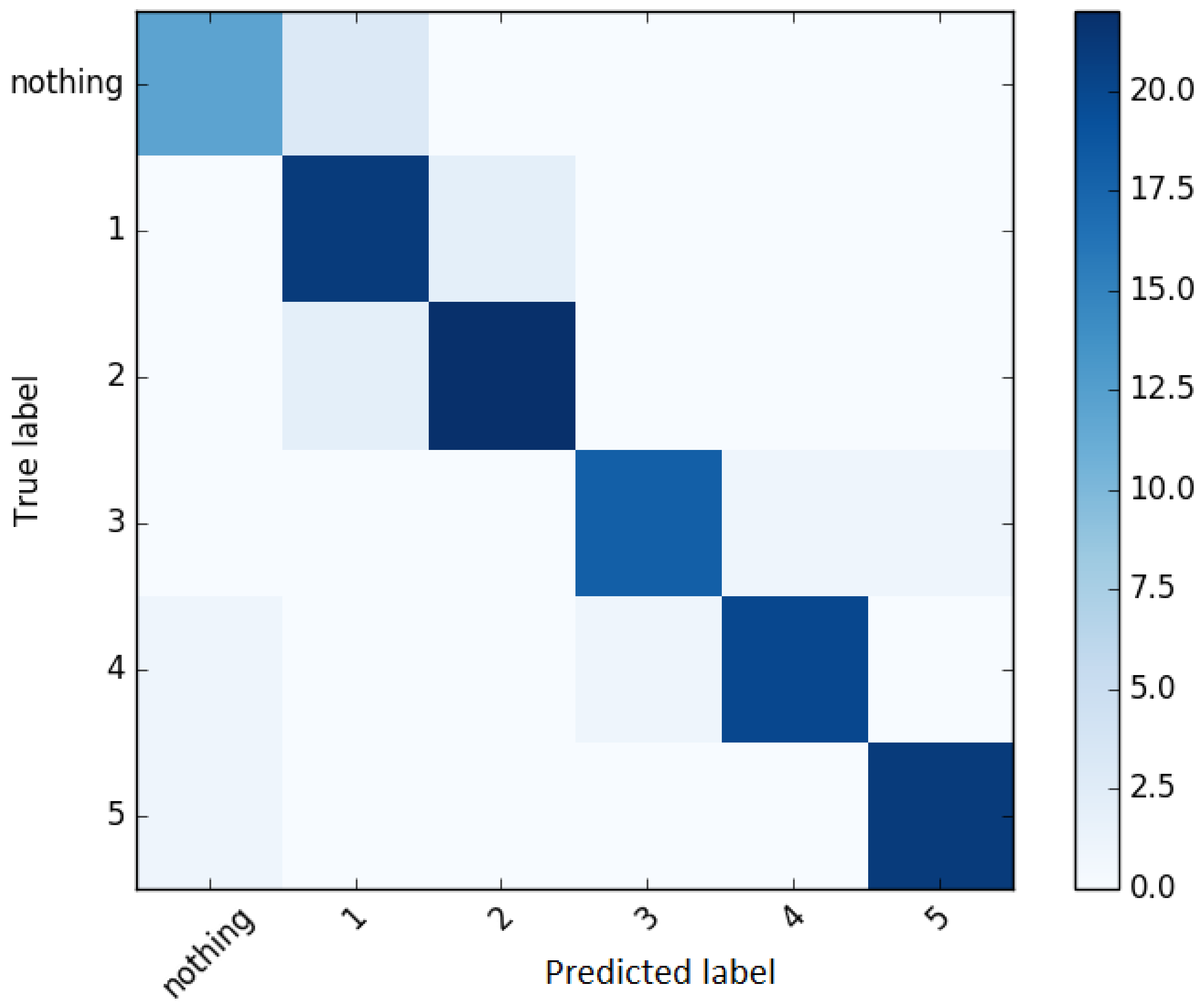
Distance: 1 to 2.5 meters



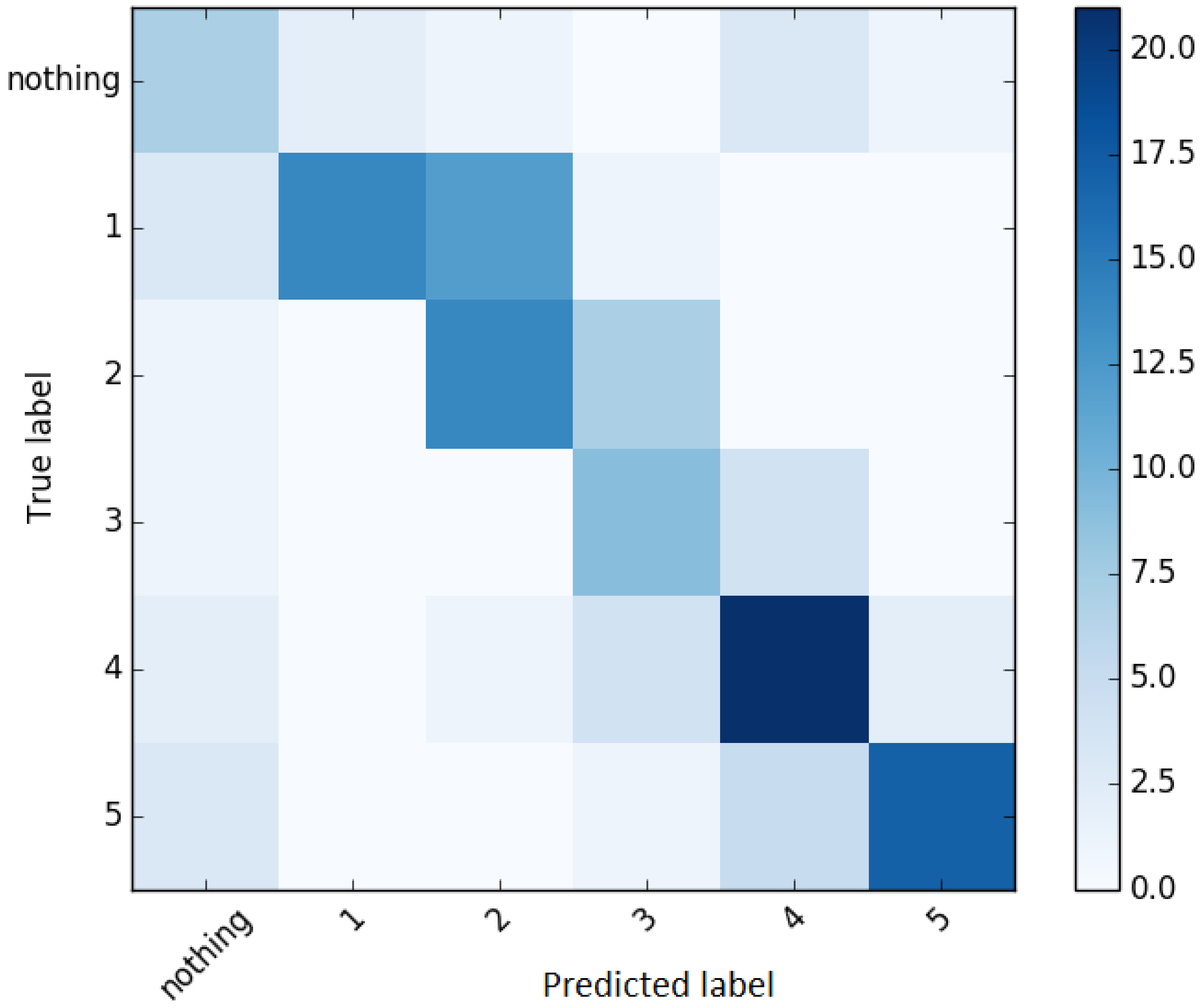
- Rather low accuracy: 30/48 correct
- Difficulties to recognize the begin and end of a hand writing gesture

SPEECH RECOGNITION

Distance: within 3.5 meters



Distance: further than 3.5 meters



USER IDENTIFICATION

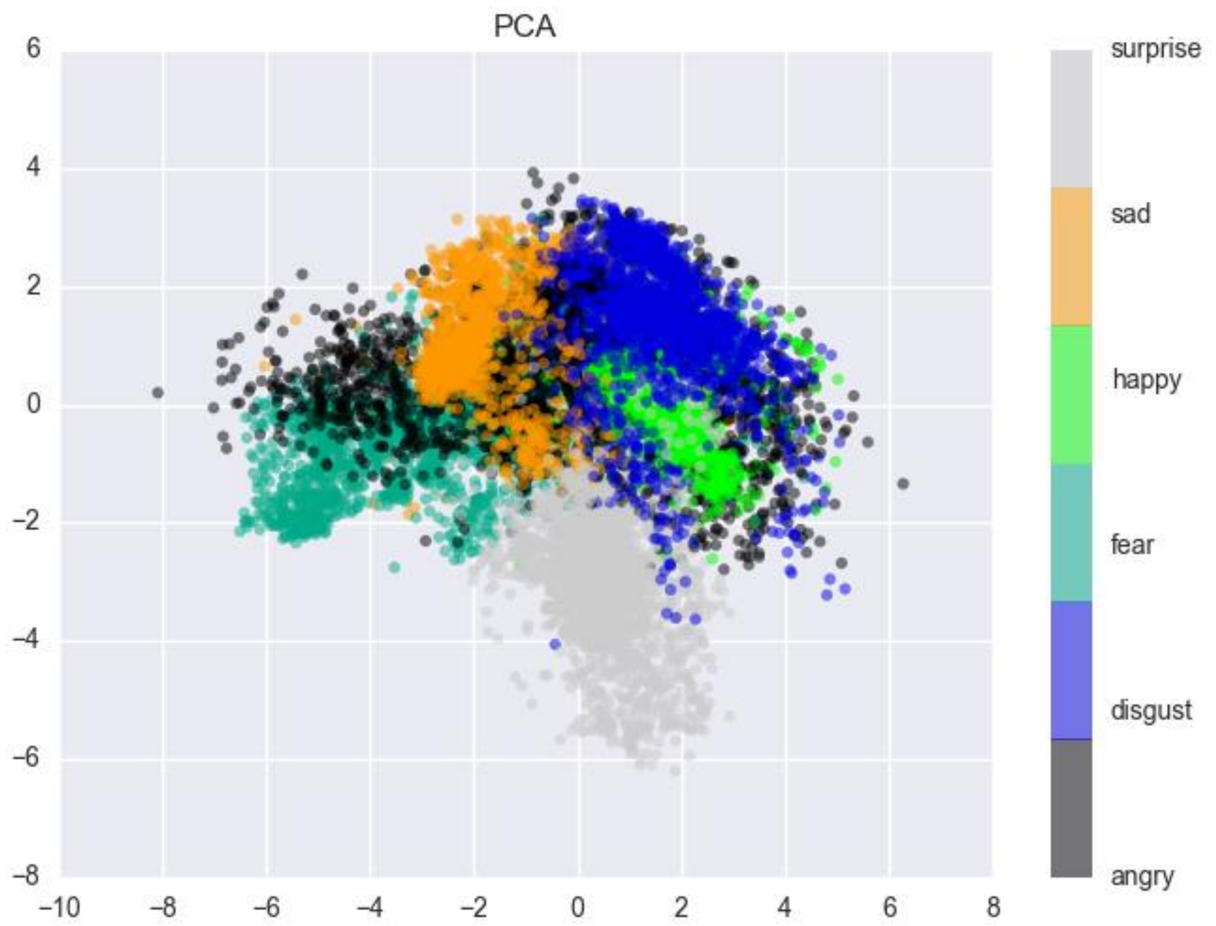
- Classic solution: username + password
- Nobody wants to input username/password on TV
  - Especially not with a remote control
- But TV is a shared device
- Solution: automatic user identification
- Disadvantage: privacy aspects

GHENT
UNIVERSITY

INTUITIVE INTERACTION WITH A CAMERA: IMPLICIT FEEDBACK

Emotion detection

- 6 emotions: anger, disgust, fear, happiness, sadness, and surprise
- Based on 17 action units
  - Contractions or relaxations of muscles in the face
- During content playback



GHENT UNIVERSITY

OUR EXPERIENCES

- Users like intuitive interaction methods for video control and browsing (good user experience)
- Explicit feedback can be provided using gestures and speech recognition
- Negative impact on recognition accuracy: background, noise, large distance, light
- Emotion recognition is a potential implicit feedback mechanism ➔ Personalization

THE FUTURE OF HUMAN COMPUTER INTERACTION

- **Privacy** concerns of camera solutions
- Alternatives for motion recognition using **wearables**
- Use **accelerometer** and gyroscope to detect specific movements
- **Localization** techniques using sensors to determine if the user is in front of TV
- Personal devices allow user **identification**
- Devices often have built-in **microphone** (speech recognition)
- **Heart rate** sensor to derive **user engagement**? (ongoing research)
- **Sound** detection to identity **key scenes** in the content

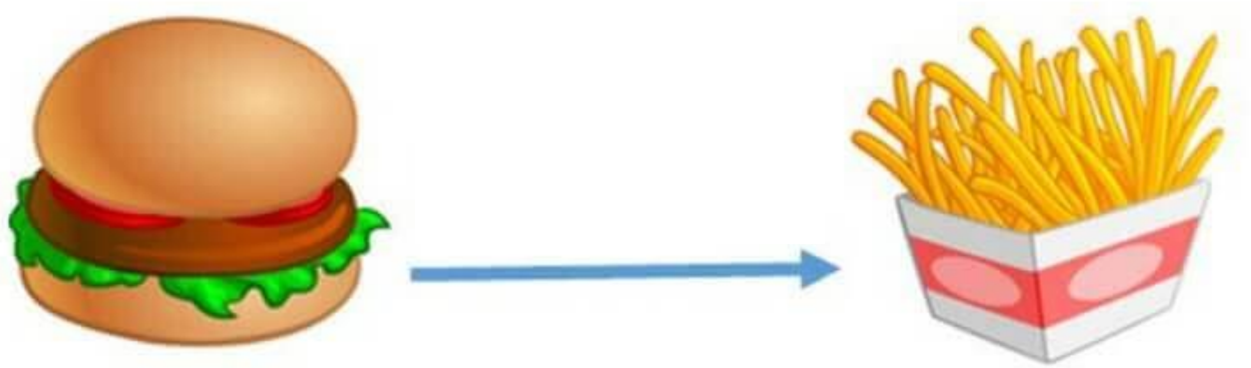
GHENT UNIVERSITY

NON-PERSONALIZED RECOMMENDERS

GHENT UNIVERSITY

TYPE OF NON-PERSONALIZED RECOMMENDATIONS

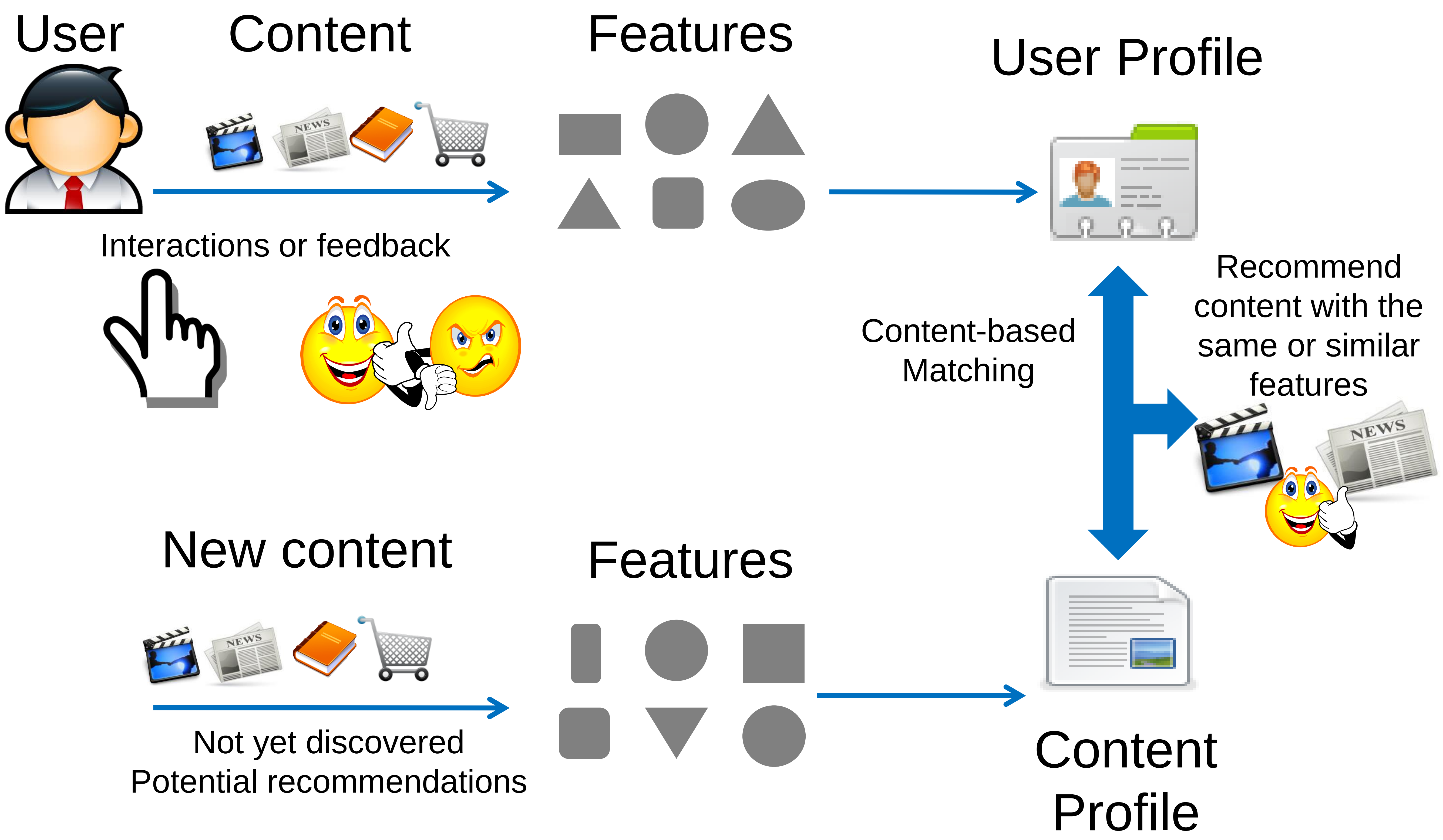
- Simple **aggregates**
  - Automatic unpersonalized lists
  - Most popular, most viewed, best rated, recently uploaded
- Association rules
  - Co-occurrence of items
    - E.g. items bought together
  - Often ephemeral
    - Matching the current activity (e.g., current browsing activity)



GHENT UNIVERSITY

CONTENT-BASED FILTERING

CONTENT-BASED FILTERING



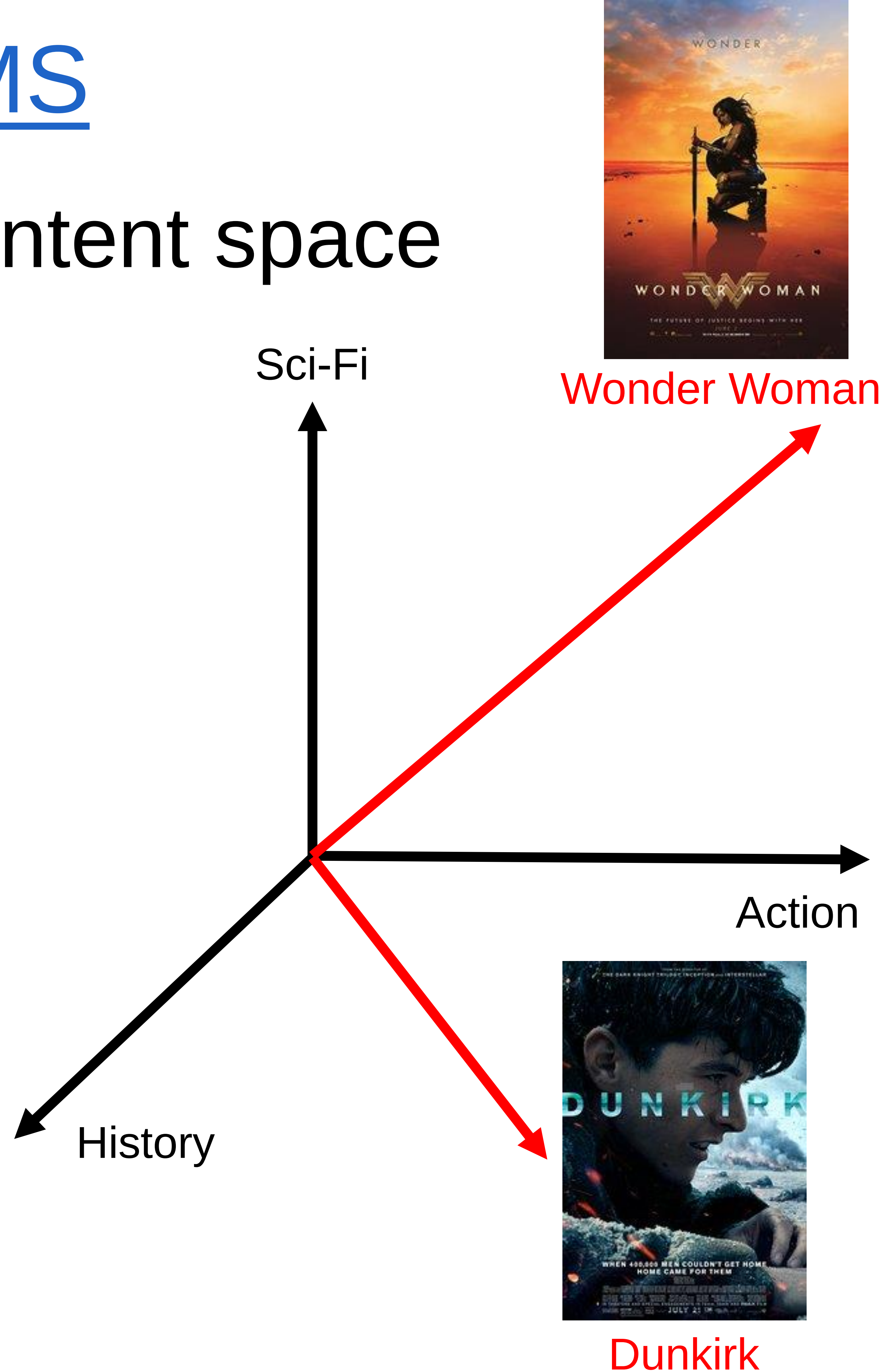
GHENT
UNIVERSITY

VECTOR SPACE MODEL: ITEMS

- Universe of attributes defines a content space
- Each attribute is one dimension
- Item = set of attributes

  =position in the content space

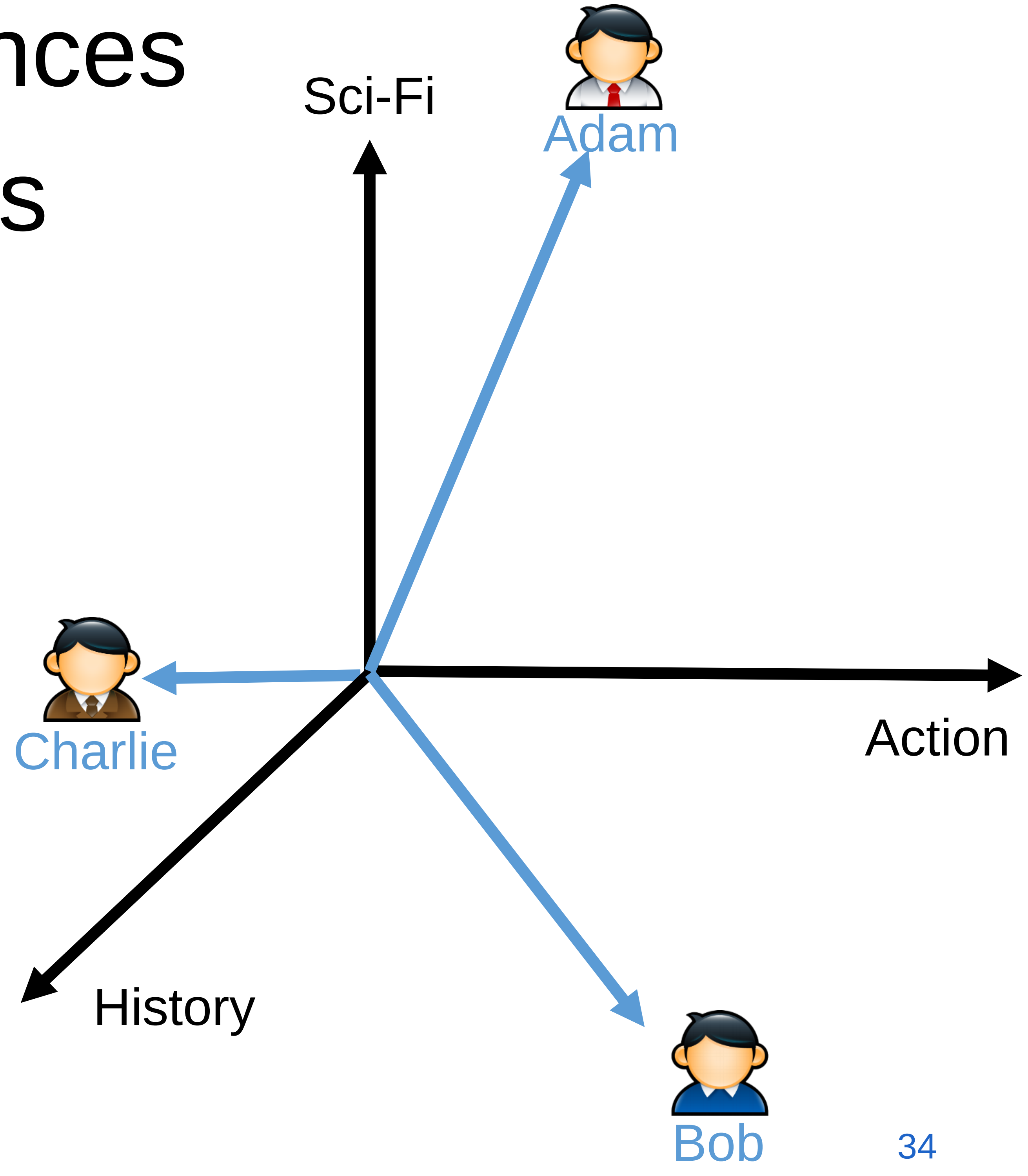
  ➔ position defines vector



GHENT
UNIVERSITY

VECTOR SPACE MODEL: USERS

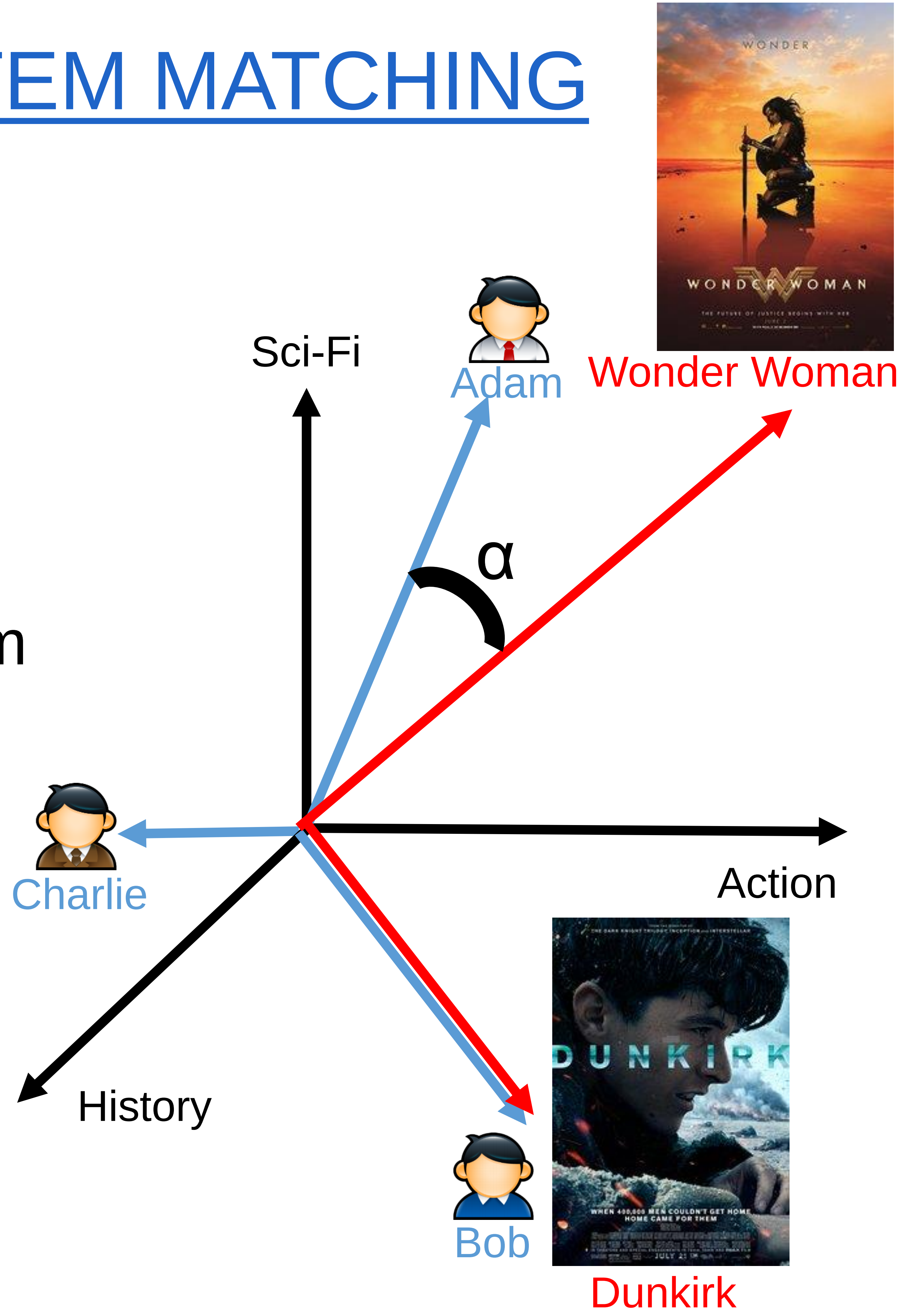
- Users can also be presented in the content space
- Based on user profile with preferences
- Preferences in terms of dimensions



GHENT UNIVERSITY

VECTOR SPACE MODEL: USER-ITEM MATCHING

- Matching users and items based on their vector
- How closely do the vectors align?
- Calculation often based on angle between vectors
  - E.g. Dunkirk is a perfect match for Bob
  - Wonder Woman is a good match for Adam
  - No good match for Charlie

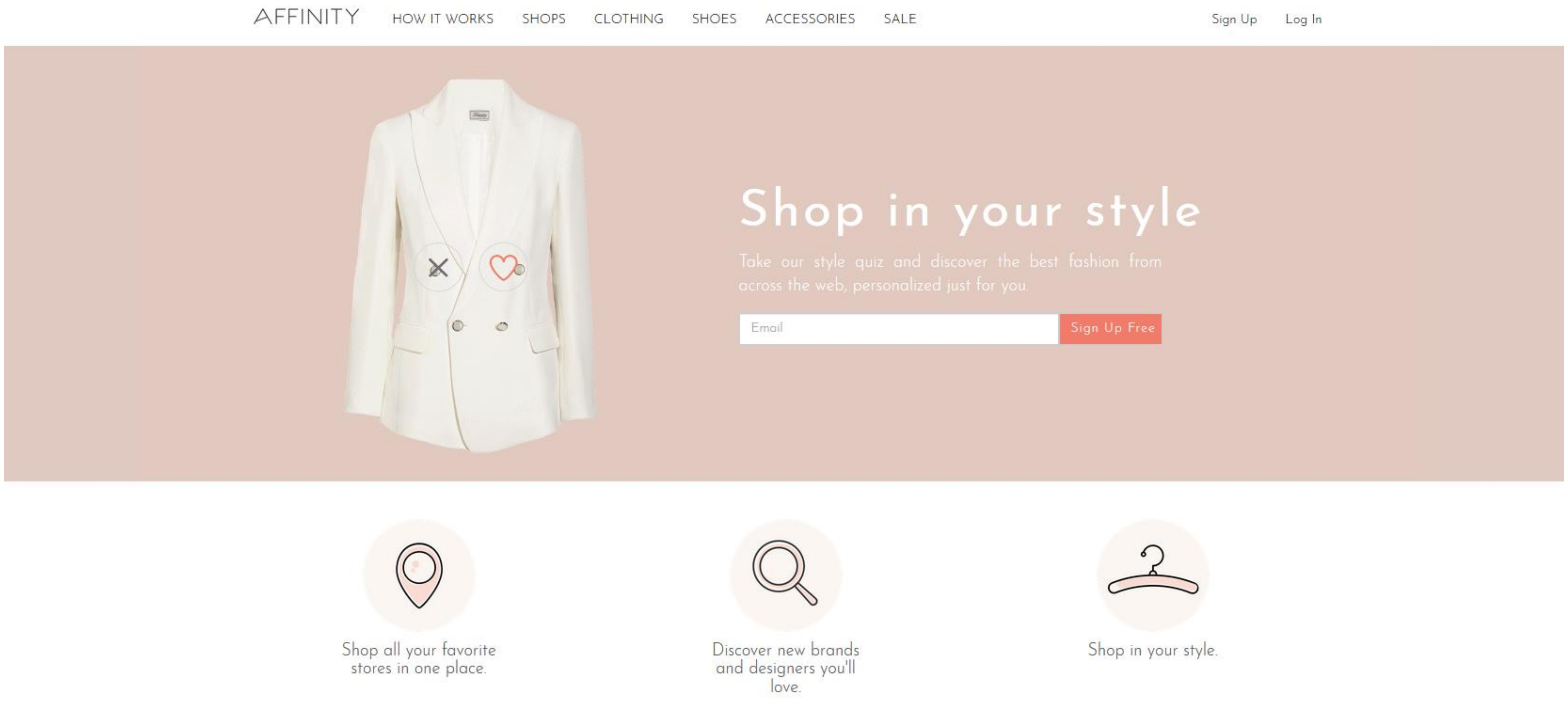


GHENT UNIVERSITY

EXAMPLE: CLOTHING DOMAIN

AFFINITY

Recommendations based on personal style



Often rather stable preferences: Style, size, gender

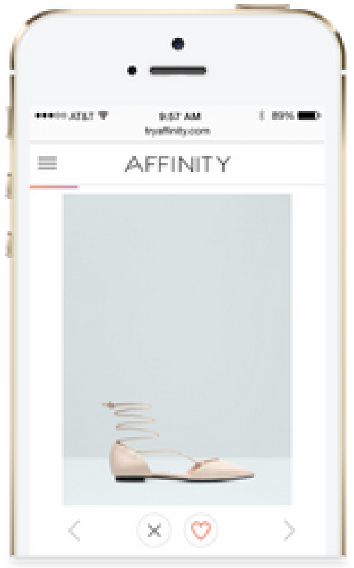
CLOTHING: HOW TO OBTAIN RECOMMENDATIONS?

AFFINITY

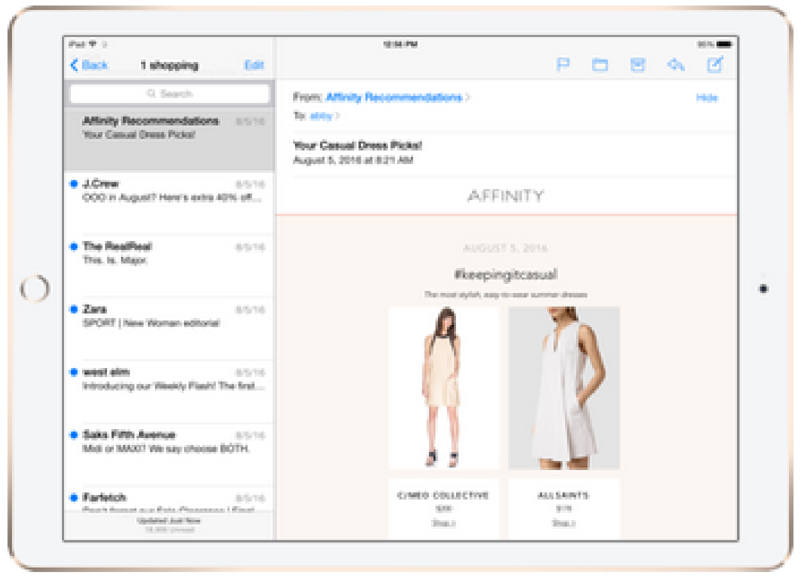
HOW IT WORKS

TAKE OUR STYLE QUIZ

To sign up, all you have to do is rate the items in our quick style quiz. Those ratings teach us about your style and preferences, so we can find you things you'll love.



Initial Profile



GET WEEKLY RECOMMENDATIONS

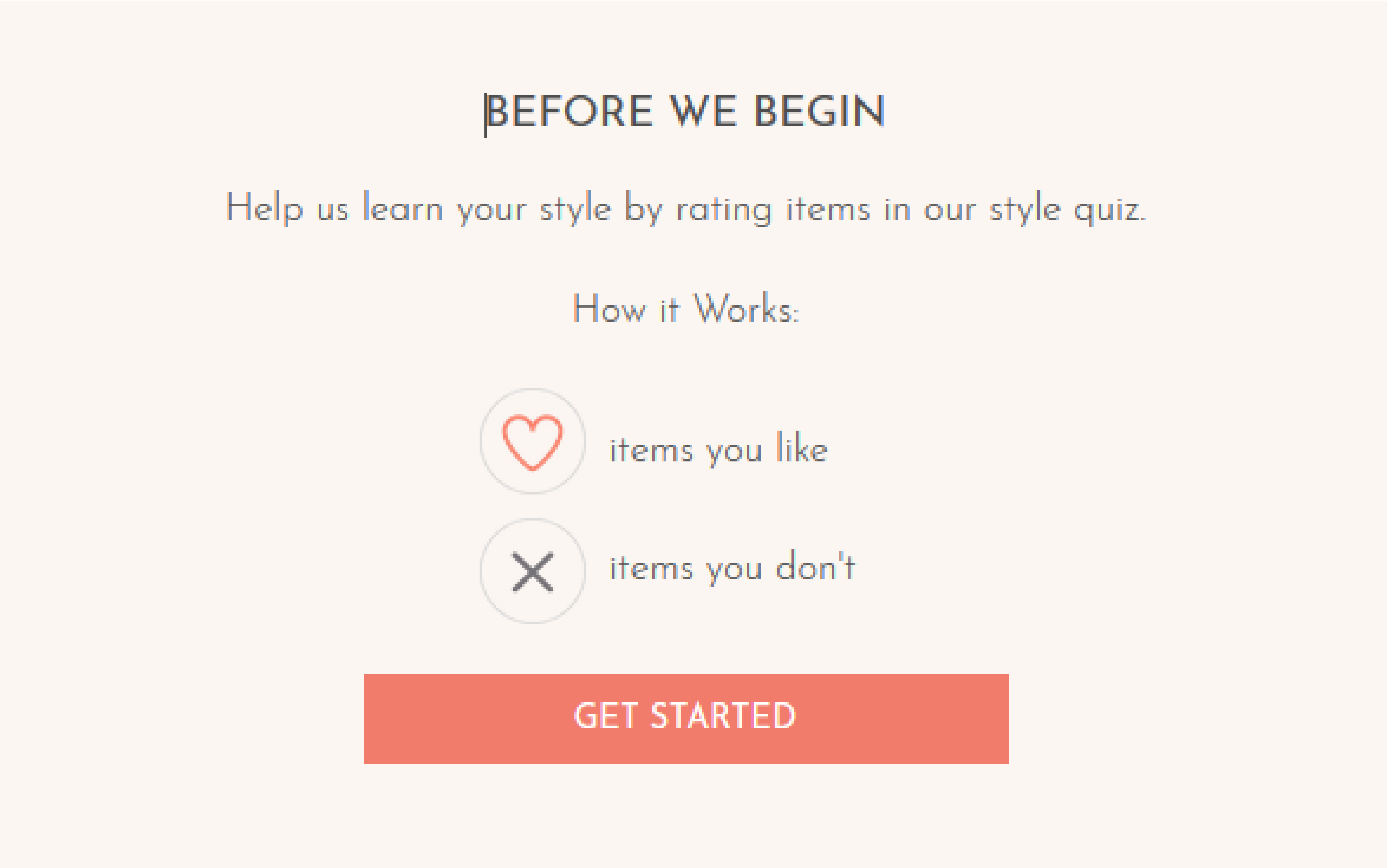
Once you're signed up, you'll receive weekly recommendations with curated items that match your style and preferences – from across thousands of brands, stores and designers.

Personal Recommendations

GHENT UNIVERSITY

CLOTHING: STYLE QUIZ

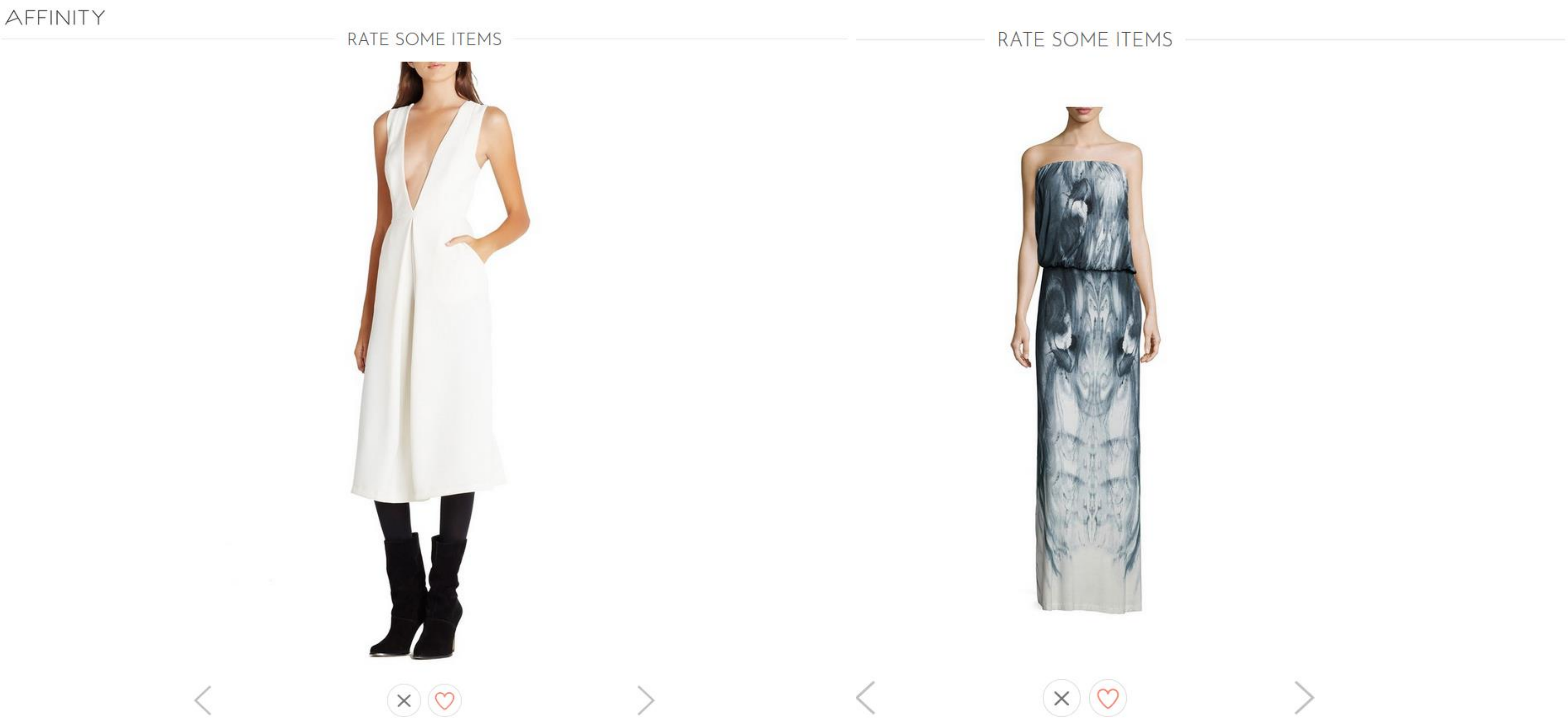
AFFINITY



Profile Building = Explicit feedback for specifically selected items

CLOTHING: STYLE QUIZ

AFFINITY



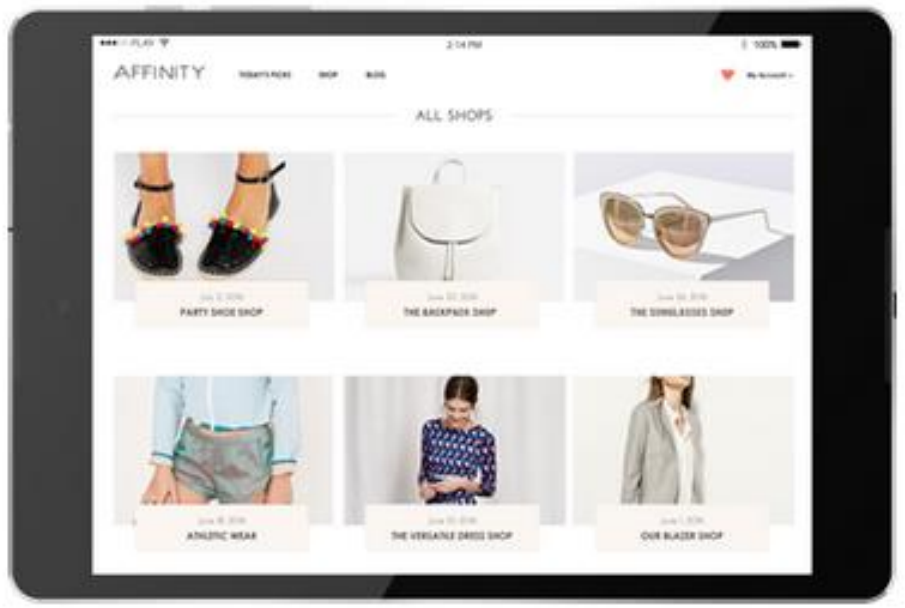
Profile Building = Explicit feedback for specifically selected items

CLOTHING: HOW TO OBTAIN RECOMMENDATIONS?

AFFINITY

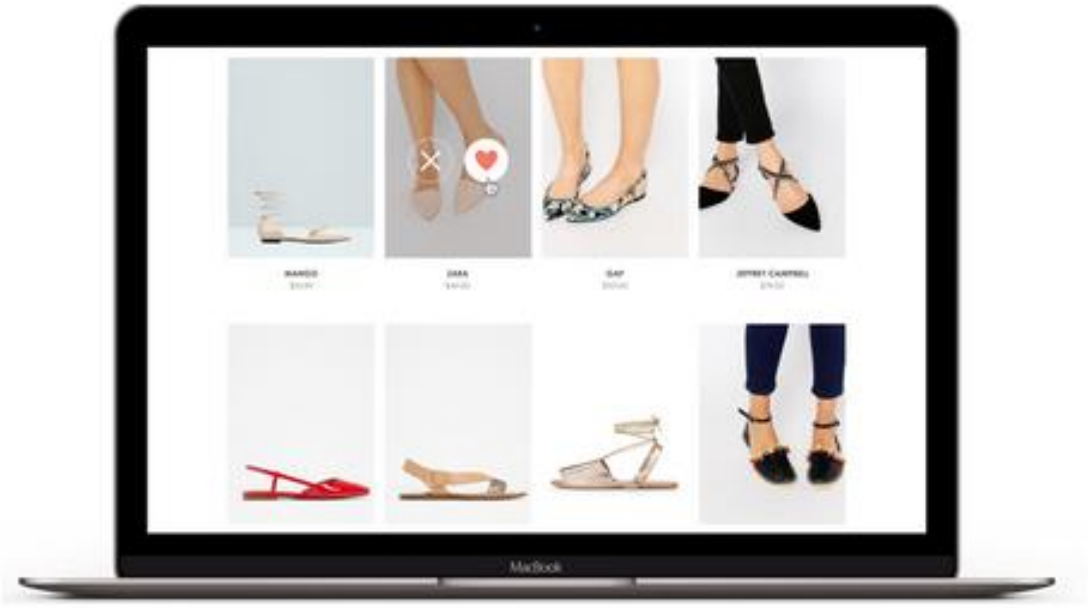
SHOP IN YOUR STYLE

If you're looking for something specific, you can browse our selection of the best items from across the web by category, season or trend. And since it's all personalized to your taste, finding things you'll love is easy.



Personalized browsing

Filtered content offer



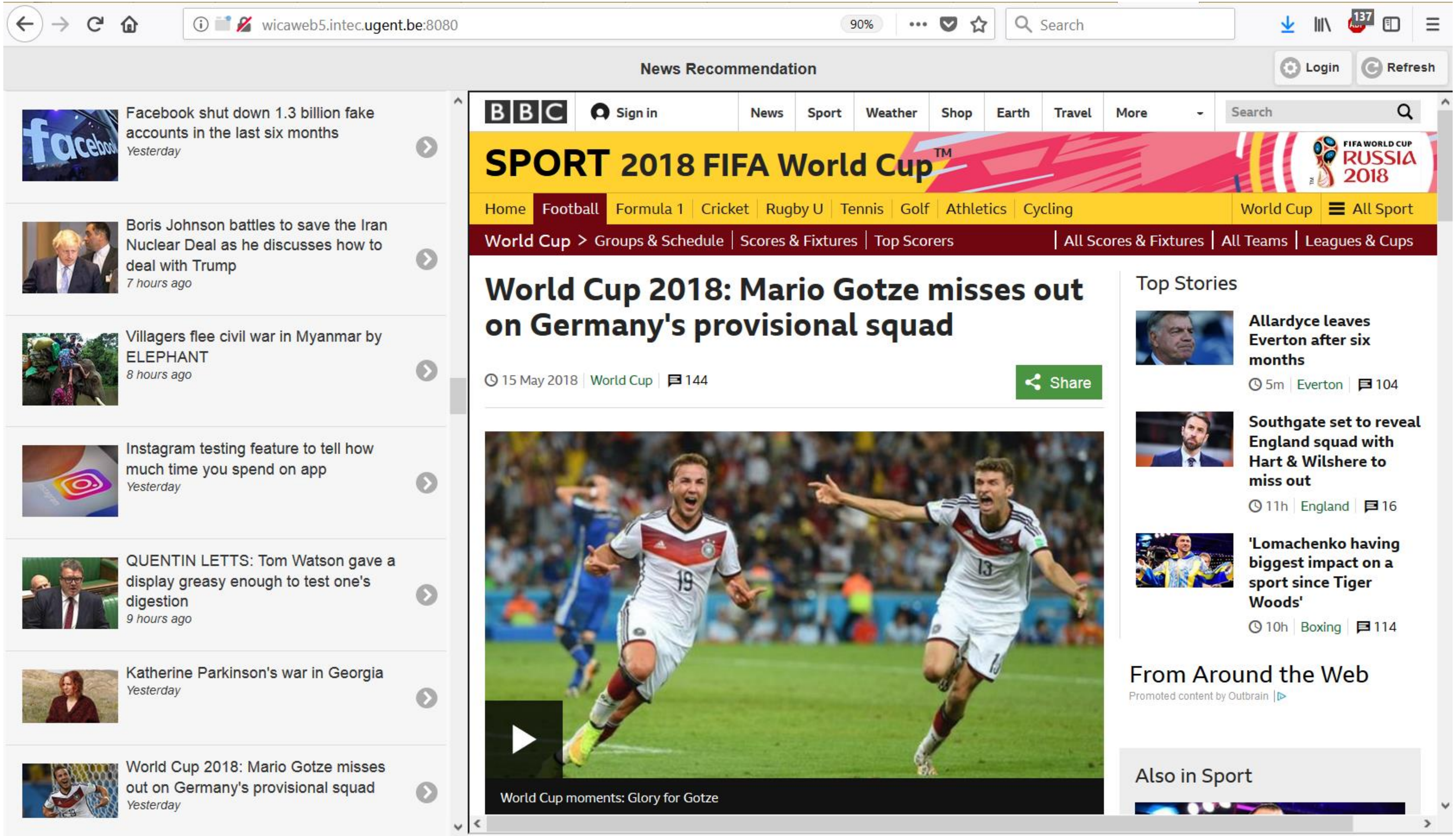
GIVE CONTINUOUS FEEDBACK

You can tell us what you think about any item on our site with one click – and we use that feedback to improve your recommendations and better personalize your shopping experience.

Adjusting profile

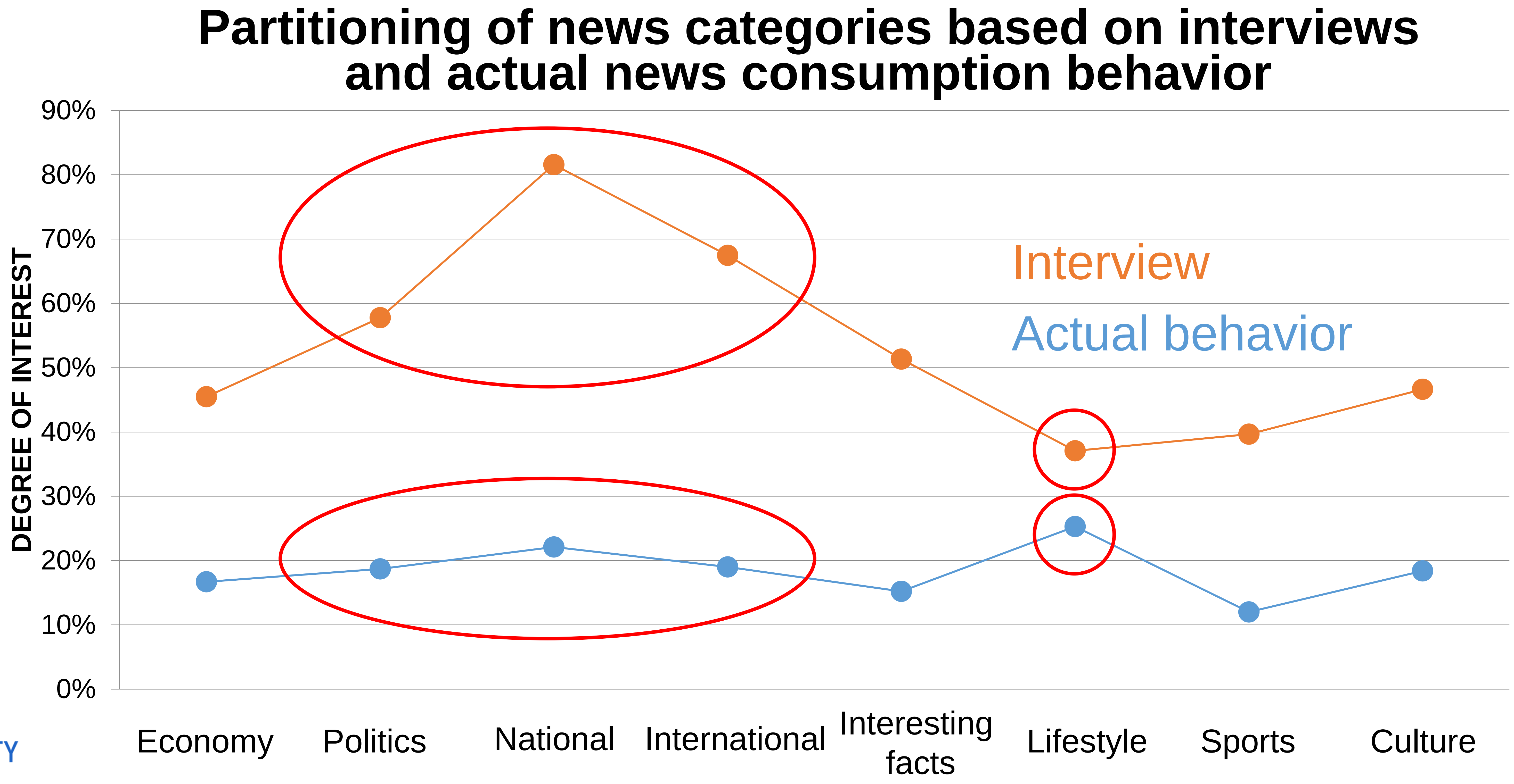
GHENT UNIVERSITY

EXAMPLE: NEWS DOMAIN



BUILDING A NEWS PROFILE BASED ON CATEGORIES

Example: Case study with + 100 users consuming digital news content



EXAMPLE: CASE-BASED REASONING

chooseabeerfor.me

Search

🍻Beer Recommendation Engine🍻

Search for a Beer and Select from the Menu

Guinness Black Lager

Guinness Black Lager by Guinness Ltd. can be described as: stout coffee chocolate euro schwarzbier

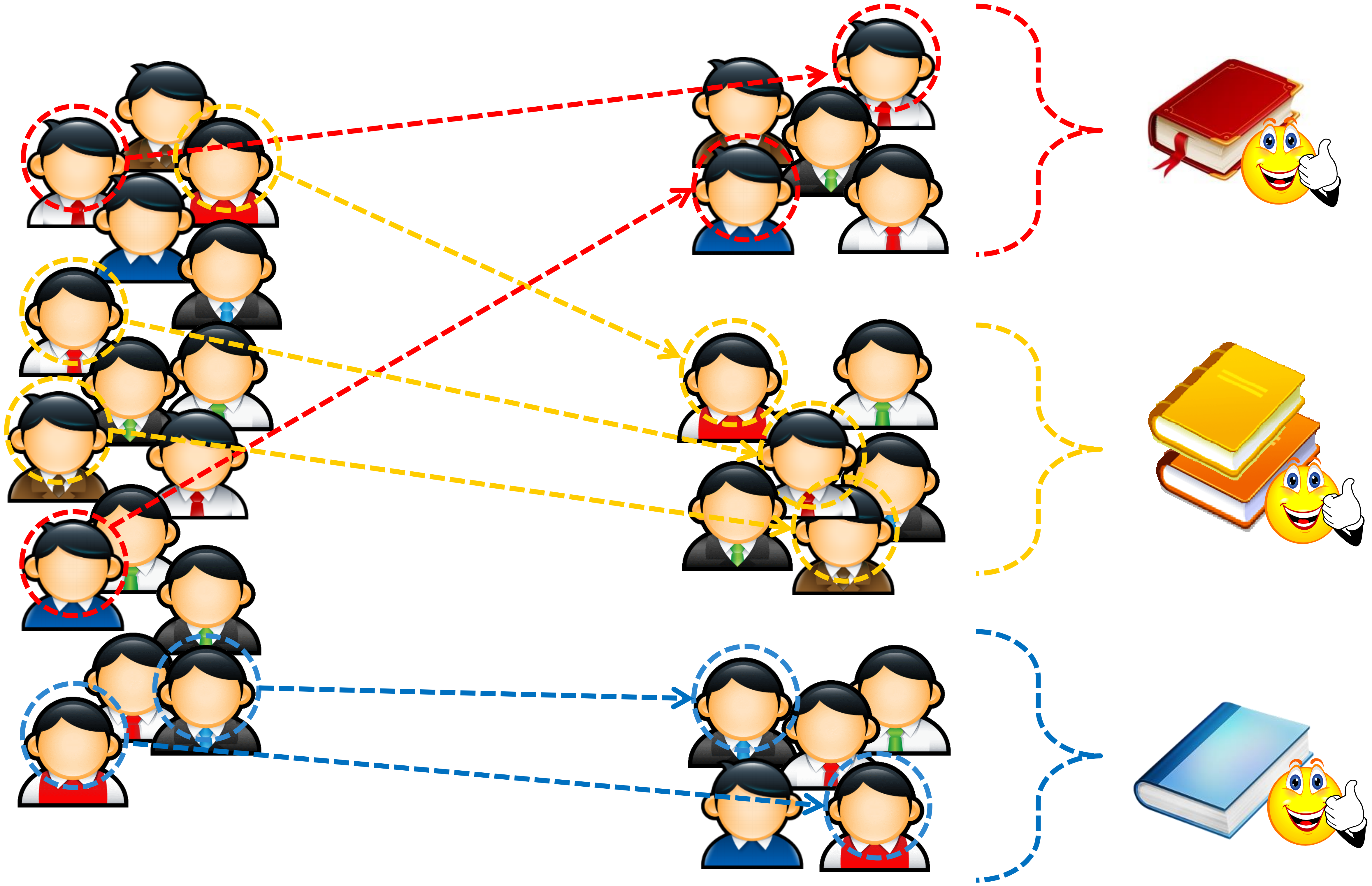
Here are some beers like Guinness Black Lager:

Similar Beer	Keywords
Xingu Black Beer by Cervejaria Kaiser	schwarzbier brazilian brazil coffee chocolate
Widow Maker Black Ale by Keweenaw Brewing Company	coffee roasty chocolate roastiness schwarzbier
Saranac Black & Tan by Matt Brewing Company / Saranac Brewery	coffee chocolate stout tan roasty
Asahi Black (Kuronama) by Asahi Breweries Ltd	sushi chocolate coffee toronto roasty
Leinenkugel's Creamy Dark by Jacob Leinenkugel Brewing Company	chocolate coffee leinie leinies euro

COLLABORATIVE FILTERING

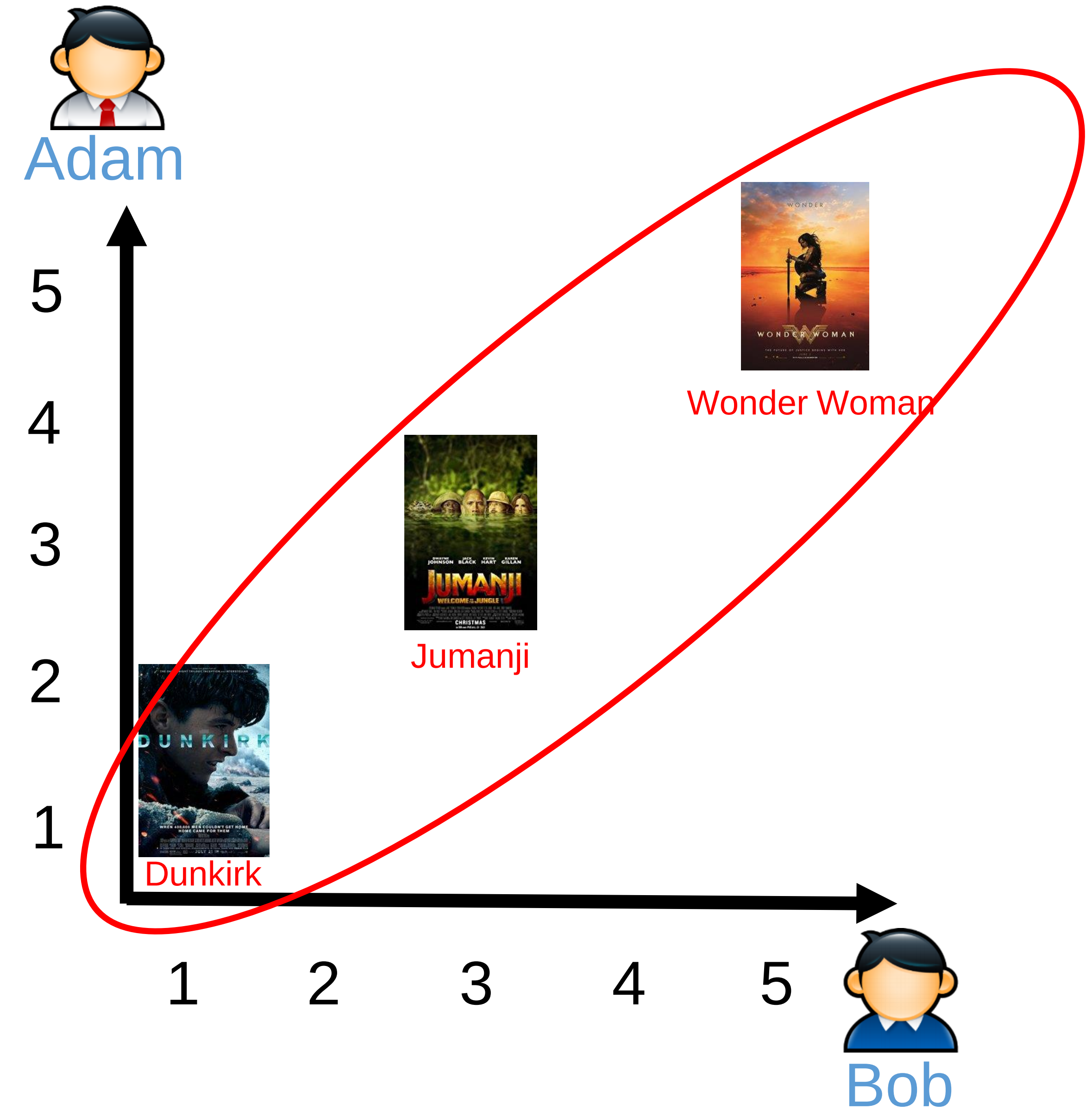
GHENT UNIVERSITY

COLLABORATIVE FILTERING



COLLABORATIVE FILTERING

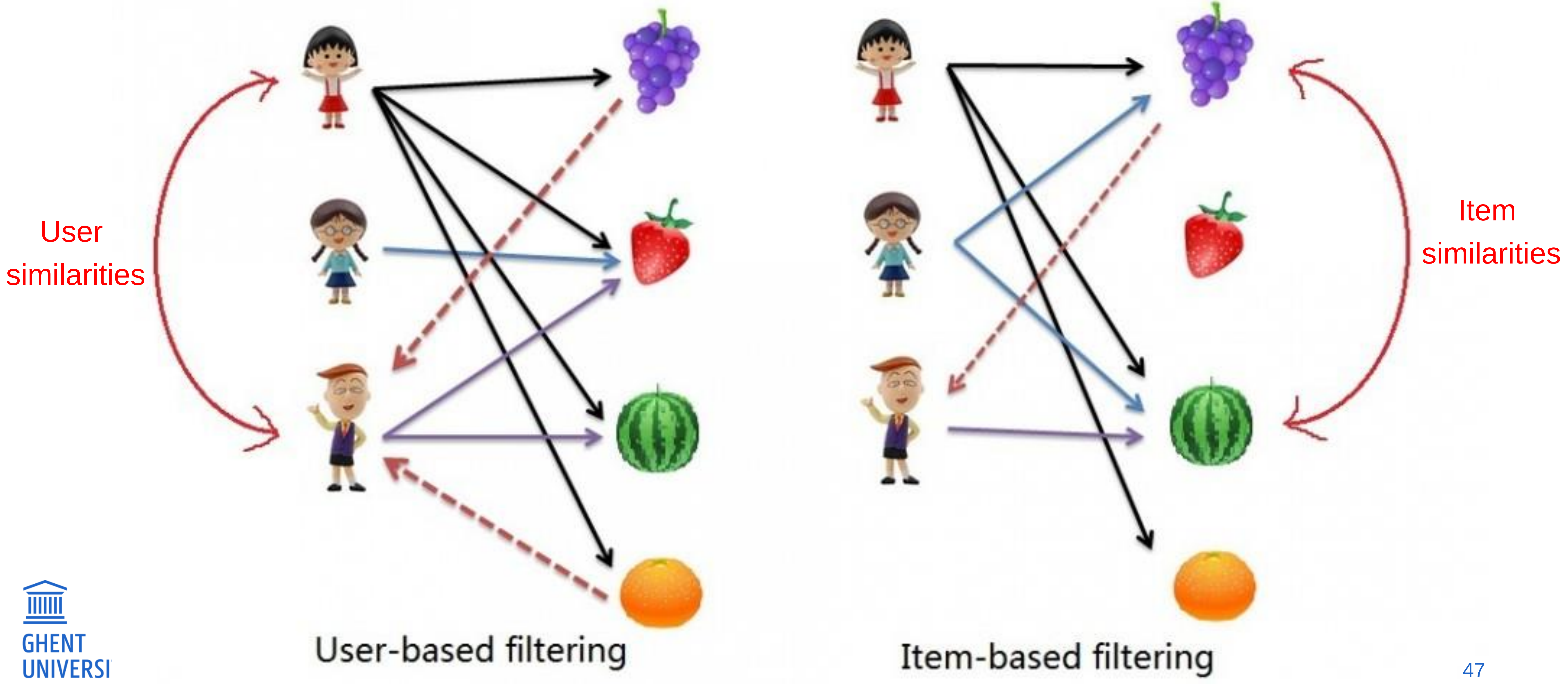
- Rating prediction for a target user and target item
  = Weighted average of similar users' ratings for that item
- Weight reflects agreement between the two users
  = Correlation in rating behavior



Users agree on items they like and on items they dislike

ITEM-ITEM COLLABORATIVE FILTERING

- Idea: compare items instead of users



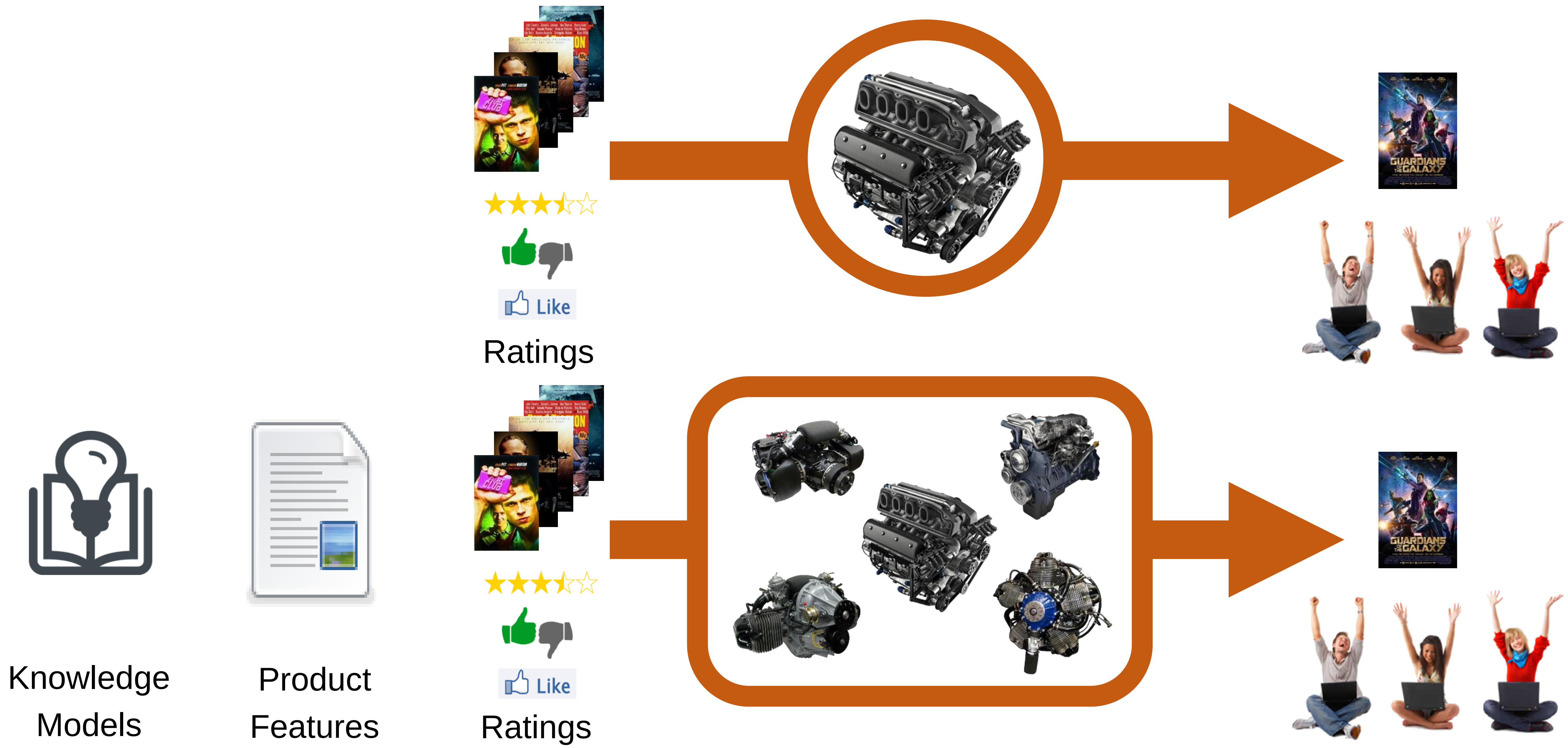
Source: http://www.salemmarafi.com/code/collaborative-filtering-with-python/

HYBRID RECOMMENDERS

GHENT UNIVERSITY

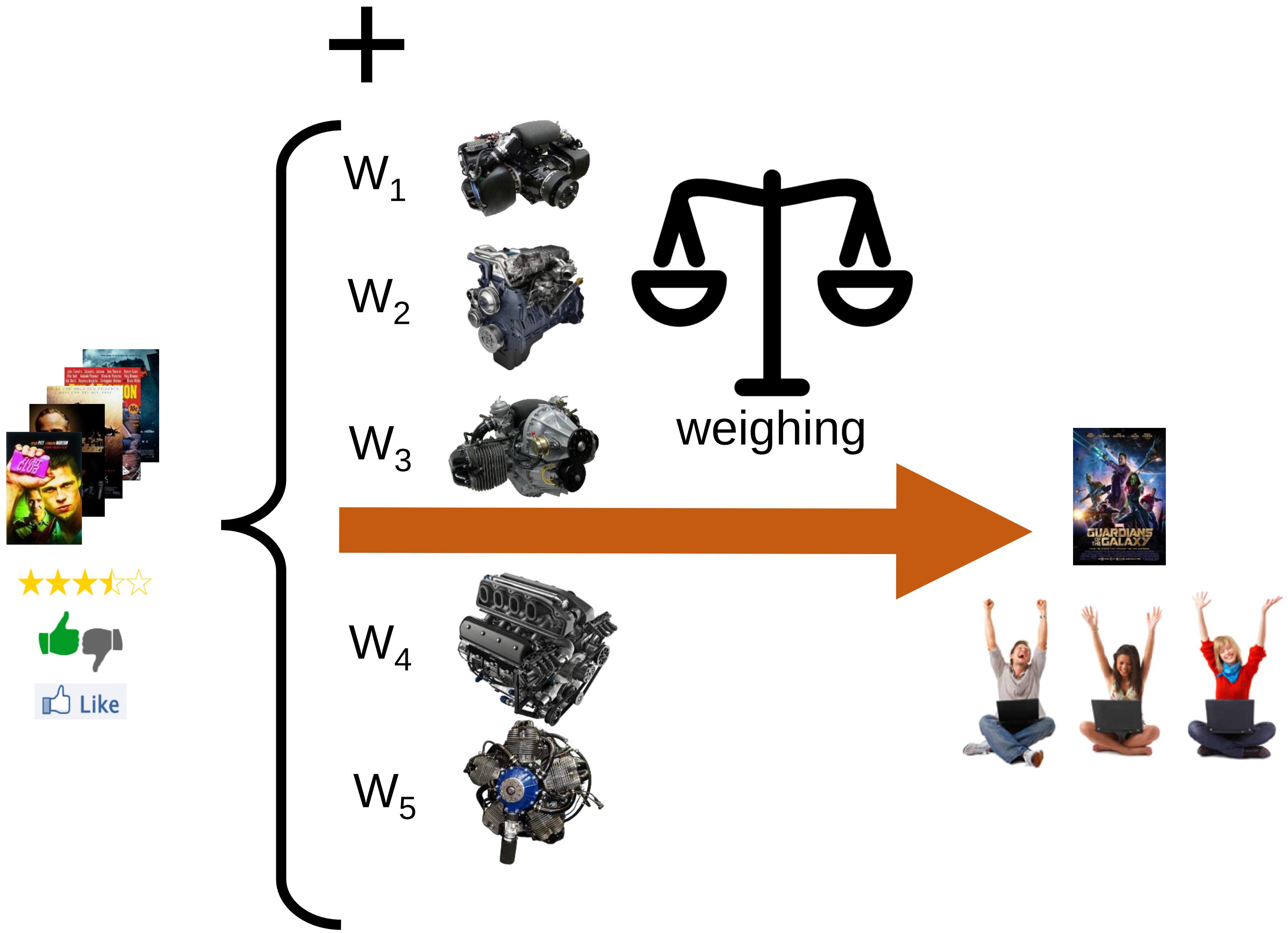
HYBRID RECOMMENDERS?

- A recommender that combines various inputs and/or various algorithms
- Similar approaches in machine learning



HOW TO COMBINE ALGORITHMS? WEIGHING

– Weighing the algorithm scores (rating prediction) or votes (recommendation)



HYBRIDS IN PRACTICE: NETFLIX COMPETITION

- Competition for all researchers and recommender enthusiasts
- Goal: find the best recommendation algorithm for Netflix
- Challenge: improve Netflix's own algorithm for predicting ratings by 10%
- Prize: 1 Million $
- Winner:
  - Weighted hybrid design based on > 100 algorithms
  - Adaptive switching of weights based on user and item features (user model and metadata)

GHENT UNIVERSITY

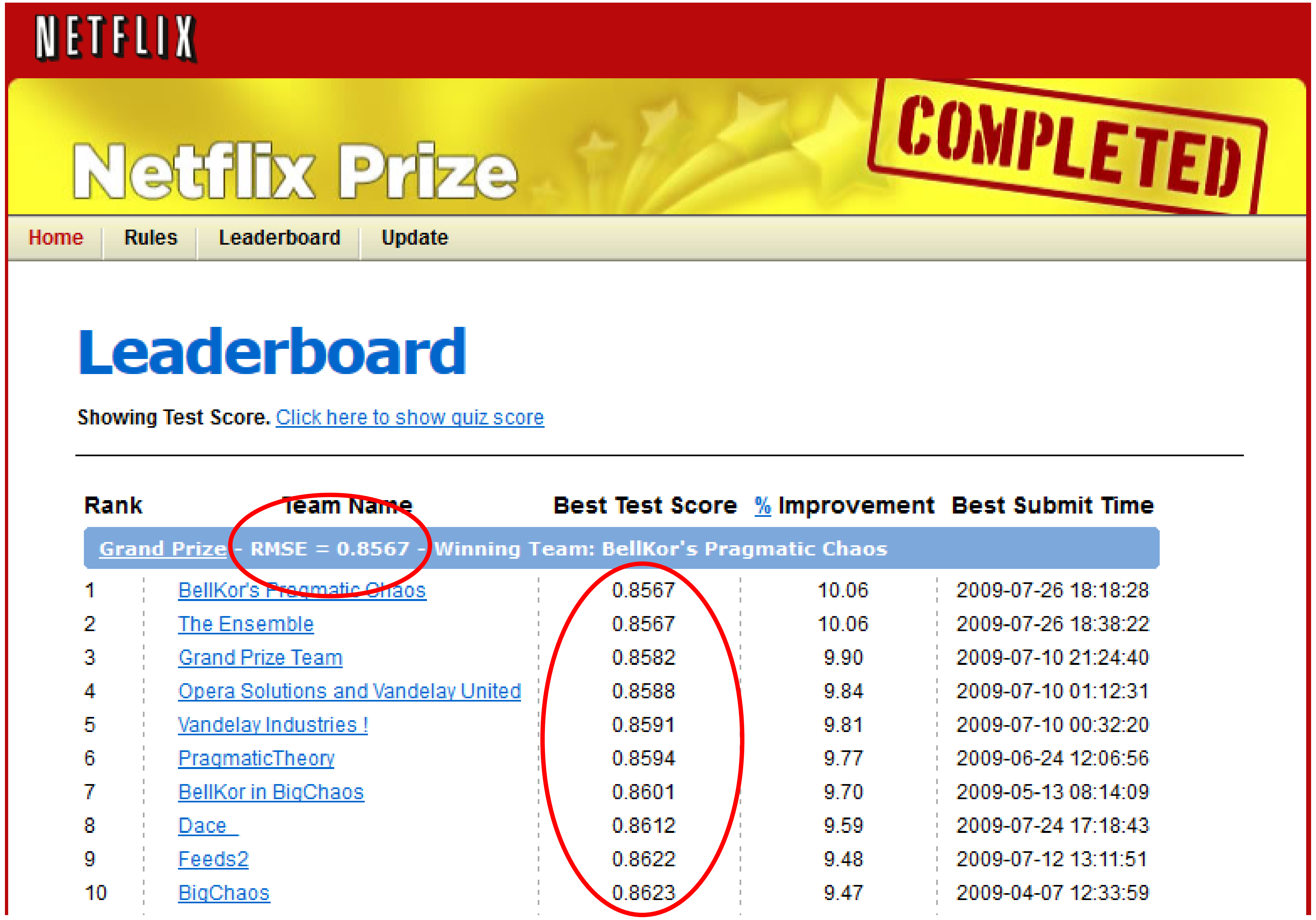
EVALUATING RECOMMENDATIONS

GHENT UNIVERSITY

RATING PREDICTION: ROOT MEAN SQUARED ERROR

- $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_{u,i} - r_{u,i})^2}$

- Difference between rating prediction $p_{u,i}$ and true rating $r_{u,i}$
- Used as the only metric in the Netflix contest
- Many other accuracy metrics, e.g. MAE and MSE are strongly correlated


NETFLIX

Netflix Prize COMPLETED

Home | Rules | Leaderboard | Update

Leaderboard

Showing Test Score. Click here to show quiz score

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59

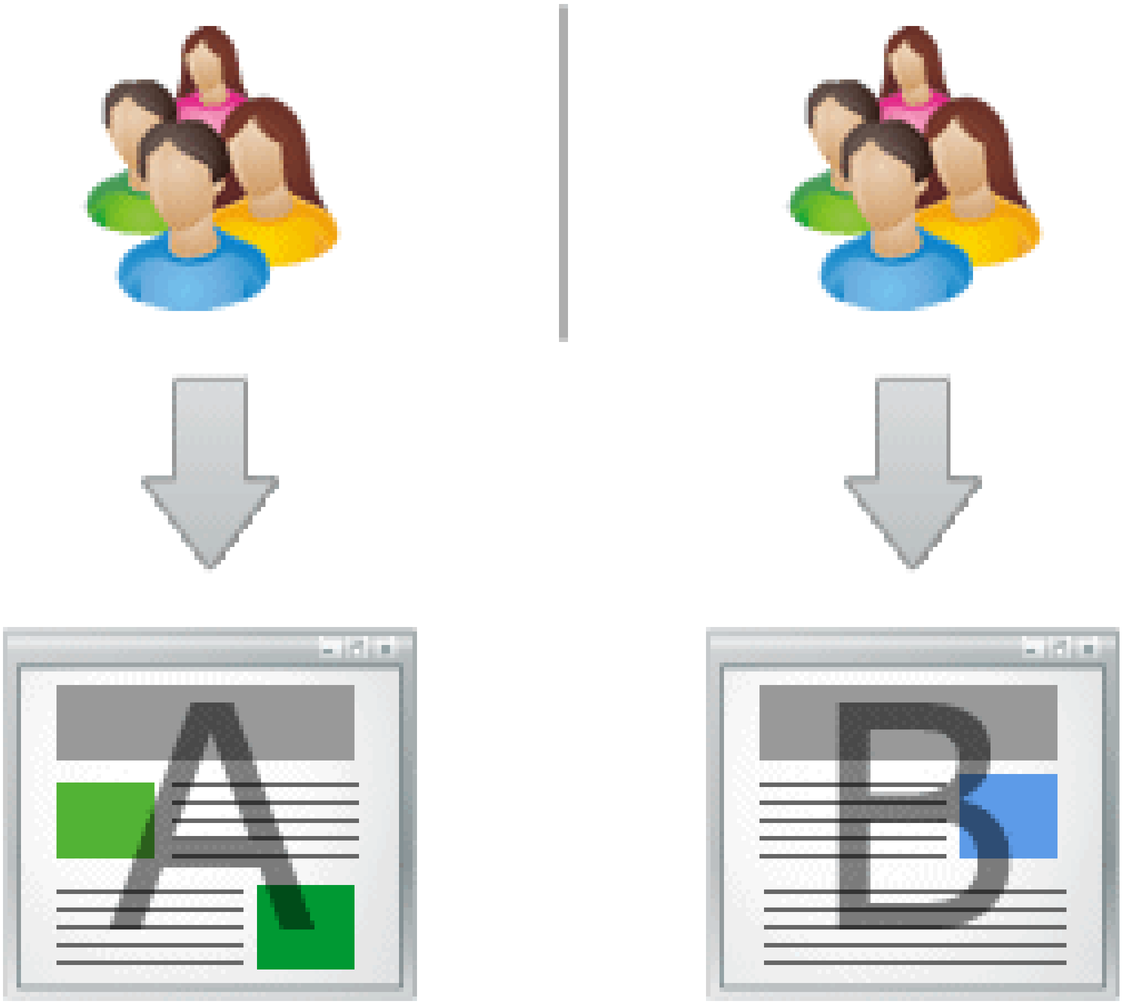
CONCERNS ABOUT RATING PREDICTION METRICS: PREDICTIONS FOR BAD MOVIES

- Important to predict ratings of 1 and 2 stars correctly ?
  - Difference between 1 and 2 stars is equally important as the difference between 4 and 5 stars ?
  - Accuracy improvement (for low ratings) might not be visible for users

- Observed problem: people watch more 3 star movies (romantic comedies, thrillers) than 5 star movies (documentaries)
  - High ratings ≠ high usage

ONLINE TESTING

- Evaluating the recommender system within the real application on real users
- One or more test systems (e.g., different algorithms) are compared
- Users get assigned to one of the alternative systems (uniformly, to avoid biasing) ➔ AB-testing
- Averaging over large enough user sets



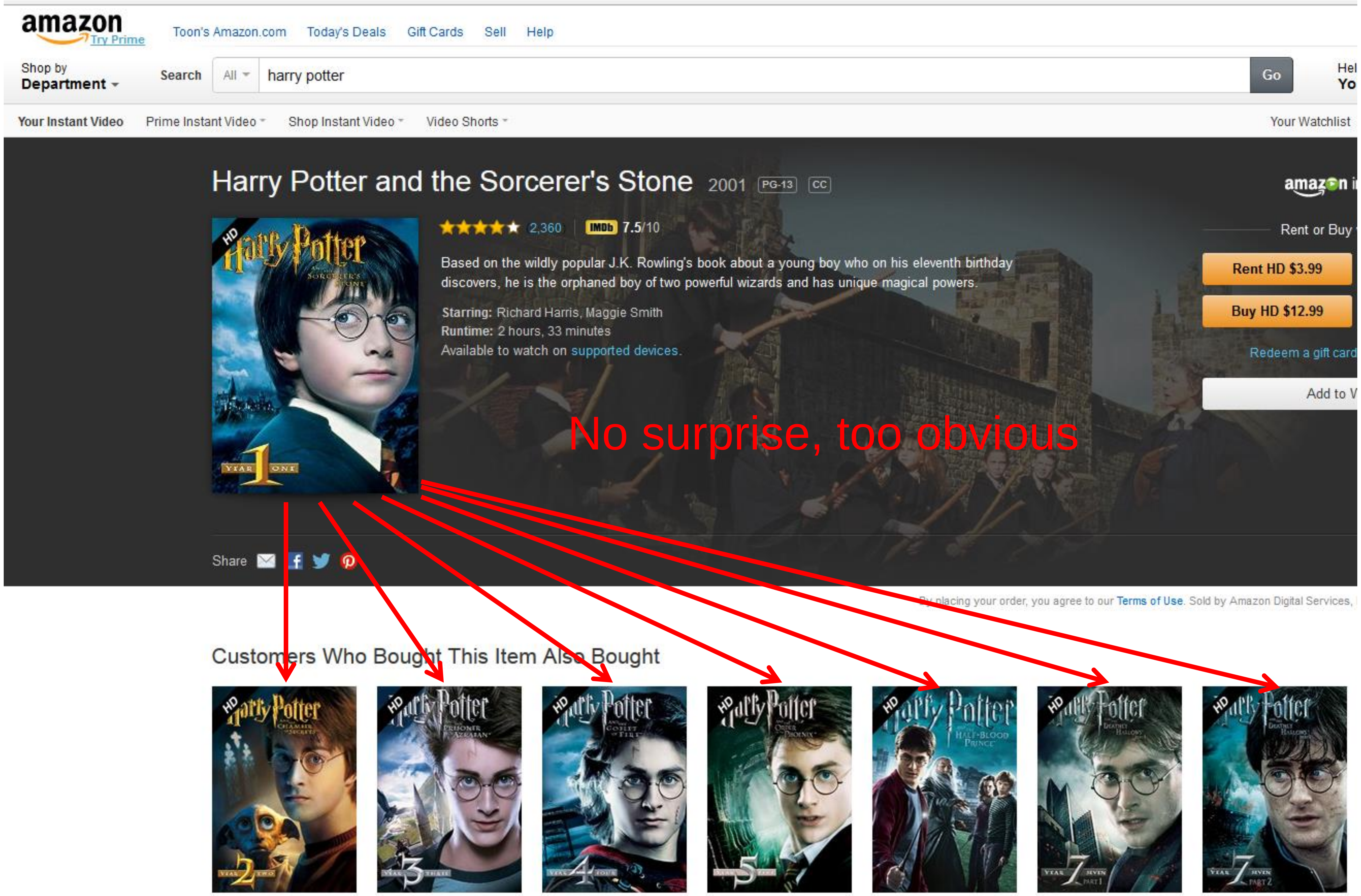
What do you want to measure?

- Immediate behavior
- Long-term behavior

Take this into account during test setup

GHENT UNIVERSITY

WHAT ELSE TO MEASURE?



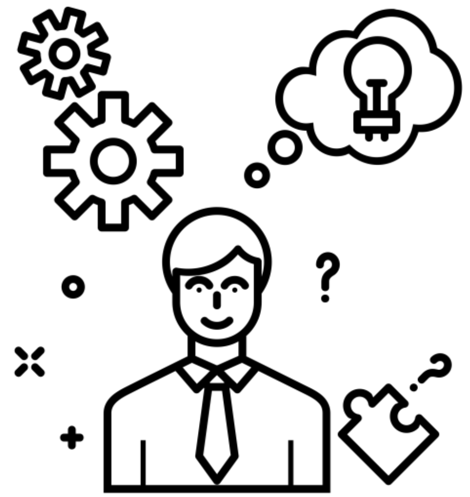
Sometimes even recommendations for different versions of the same book/item. E.g. hardcover, paperback edition, collection box, …

SERENDIPITY

- How surprising are the successful recommendations?
- Serendipity:
  - Not yet discovered, and not be expected by the user
  - Interesting, relevant and useful to the user
- Risk to lead users to unsatisfying or useless items

GHENT UNIVERSITY

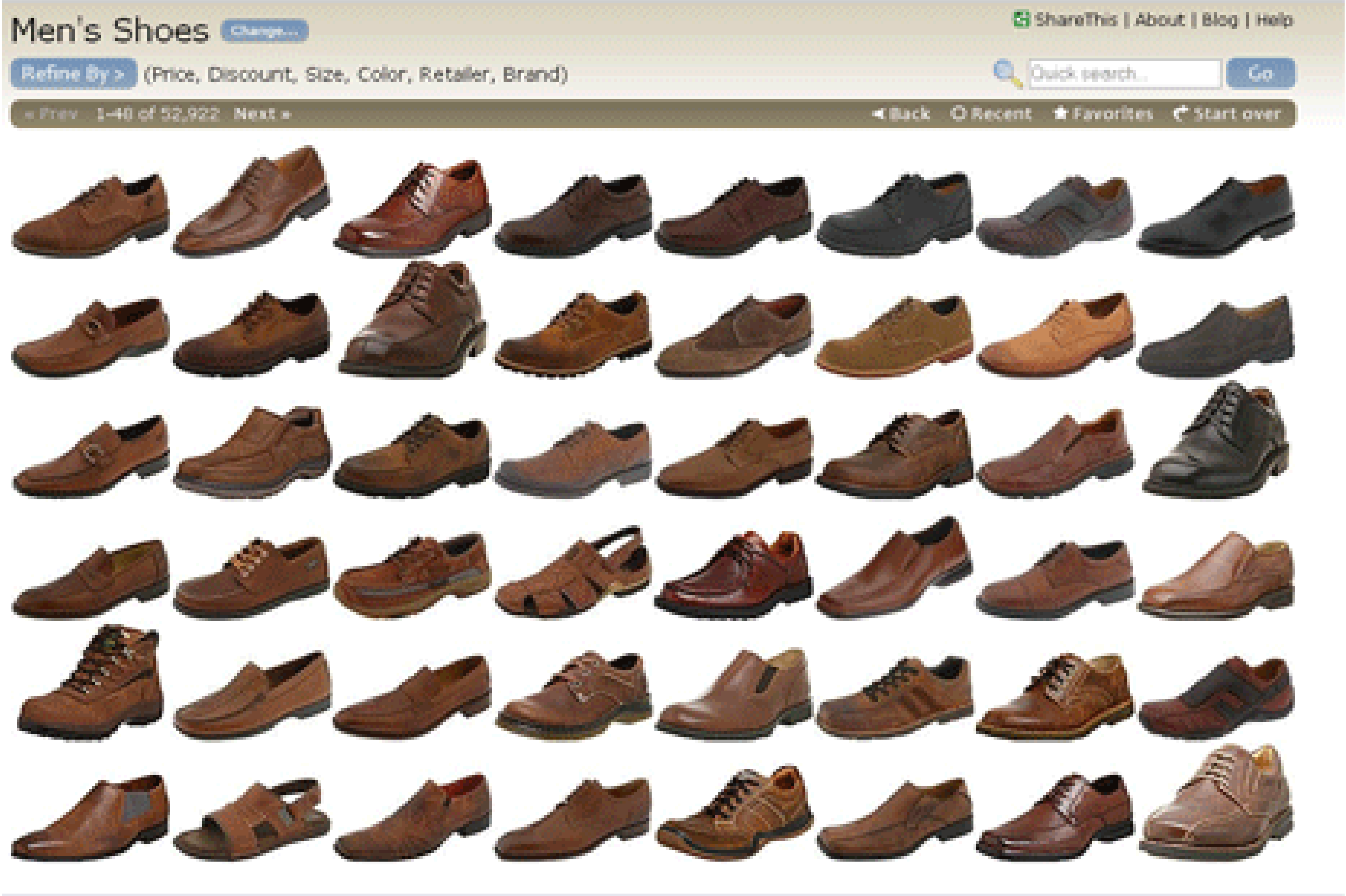
MEASURING SERENDIPITY



- Difficult to measure
- Manual: ask users through a questionnaire
- Automatic:
  - Score a successful recommendation based on **how far** it is (content-based similarity) **from the known items** in the user's profile.
  - **Unexpectedness** of an item: difference in prediction score between the surprising recommender and a **primitive recommender** (e.g., popular recommender)

DIVERSITY

- Measure of how different the items in a Top-N recommendation list are
- Diversity of a set of items:
  - The diversity of the most similar pairs
  - The average diversity of all pairs of the list



- Measuring
  - With similarity metric (Diversity is the opposite of similarity)
  - Asking the user's opinion

FILTER BUBBLE

- A state of intellectual isolation as a result from personalization
  - Algorithms selectively assume what information users would want to see
  - Users get only information according to this assumption
- “Algorithms feed users with tastes / opinions that reinforce the ones they already got”
- Serious problem for domains such as news

GHENT UNIVERSITY

CHALLENGE OF NEWS DIVERSITY

Sender – Receiver

offerings - consumption

Content Diversity

Categories, topics, ideas,
(political) viewpoints

Source Diversity

Newspapers, journalists,
news agencies

Gatekeeping diversity

Audience = secondary gatekeepers
Distributed through various social media
Augmented with user content (comments)

GROUP RECOMMENDATIONS

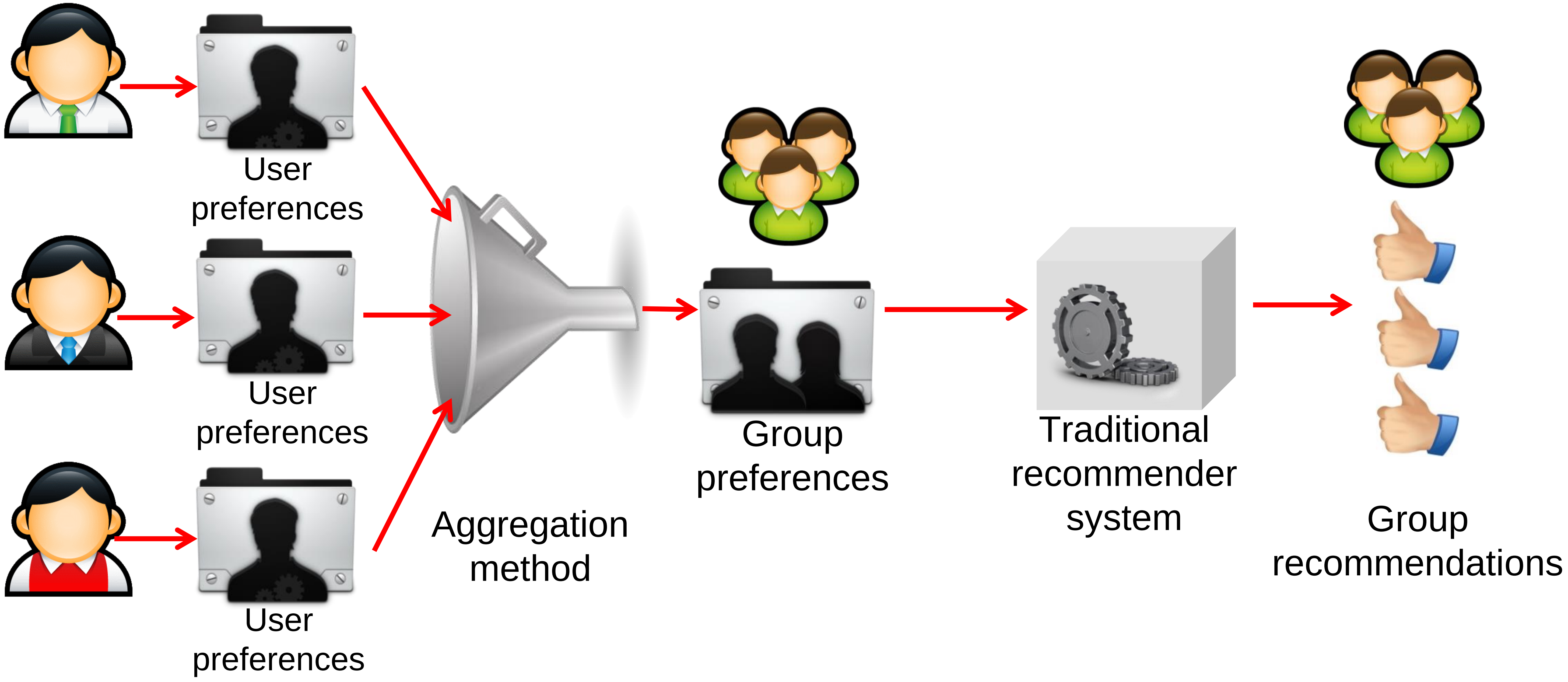
GHENT UNIVERSITY

WHY DO WE NEED GROUP RECOMMENDATIONS?

CHALLENGES FOR GROUP RECOMMENDATIONS

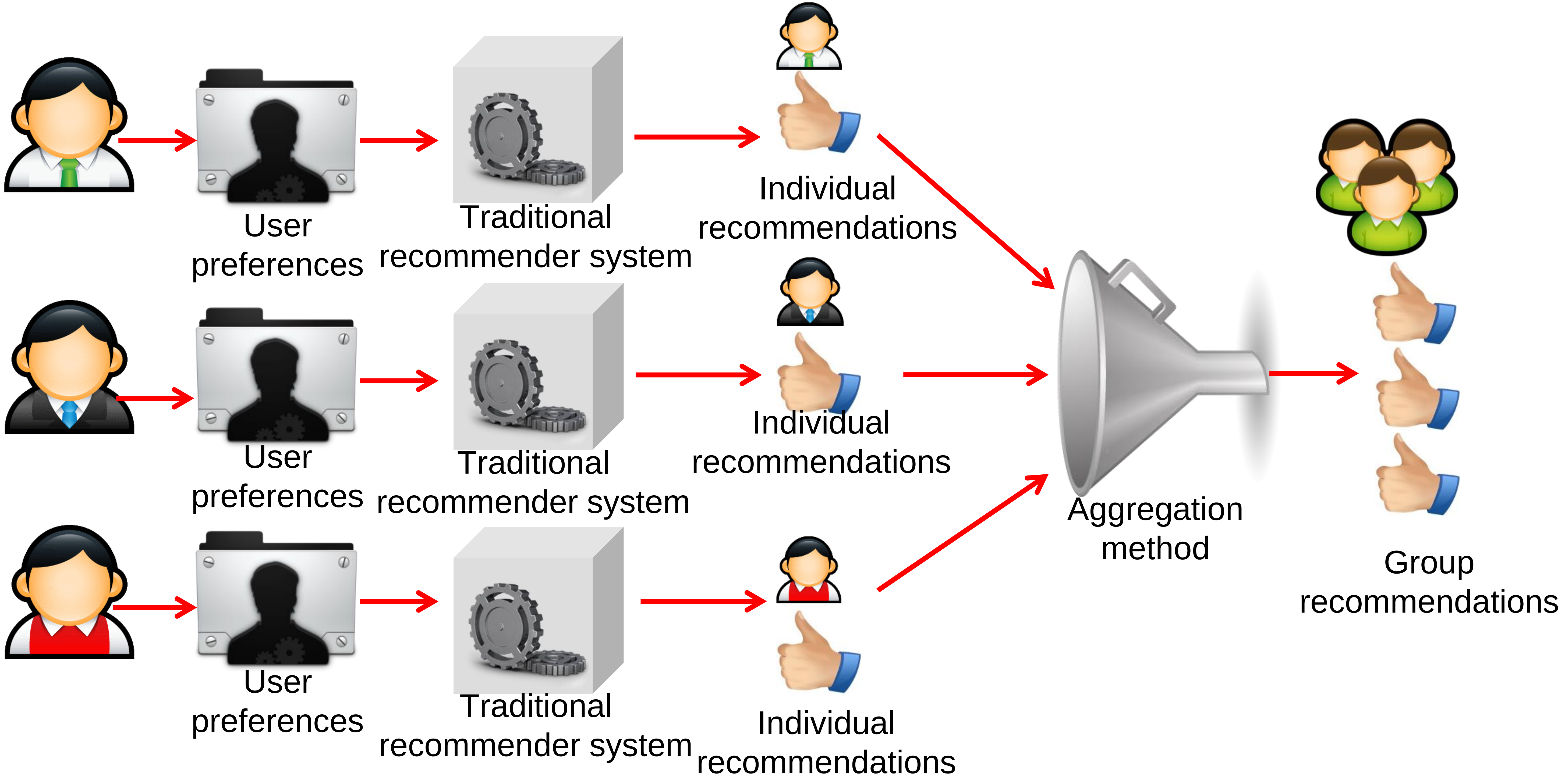
HOW TO GENERATE GROUP RECOMMENDATIONS?

Strategy 1: aggregating user preferences



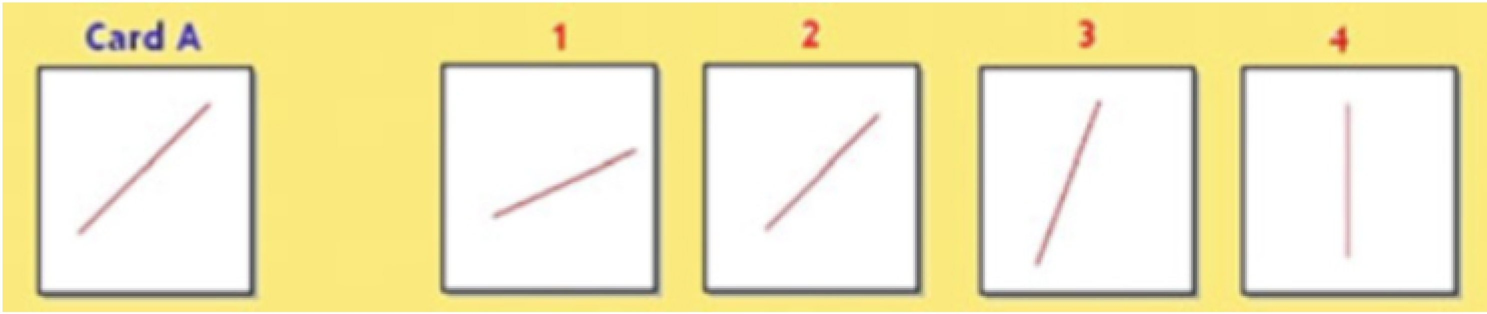
HOW TO GENERATE GROUP RECOMMENDATIONS?

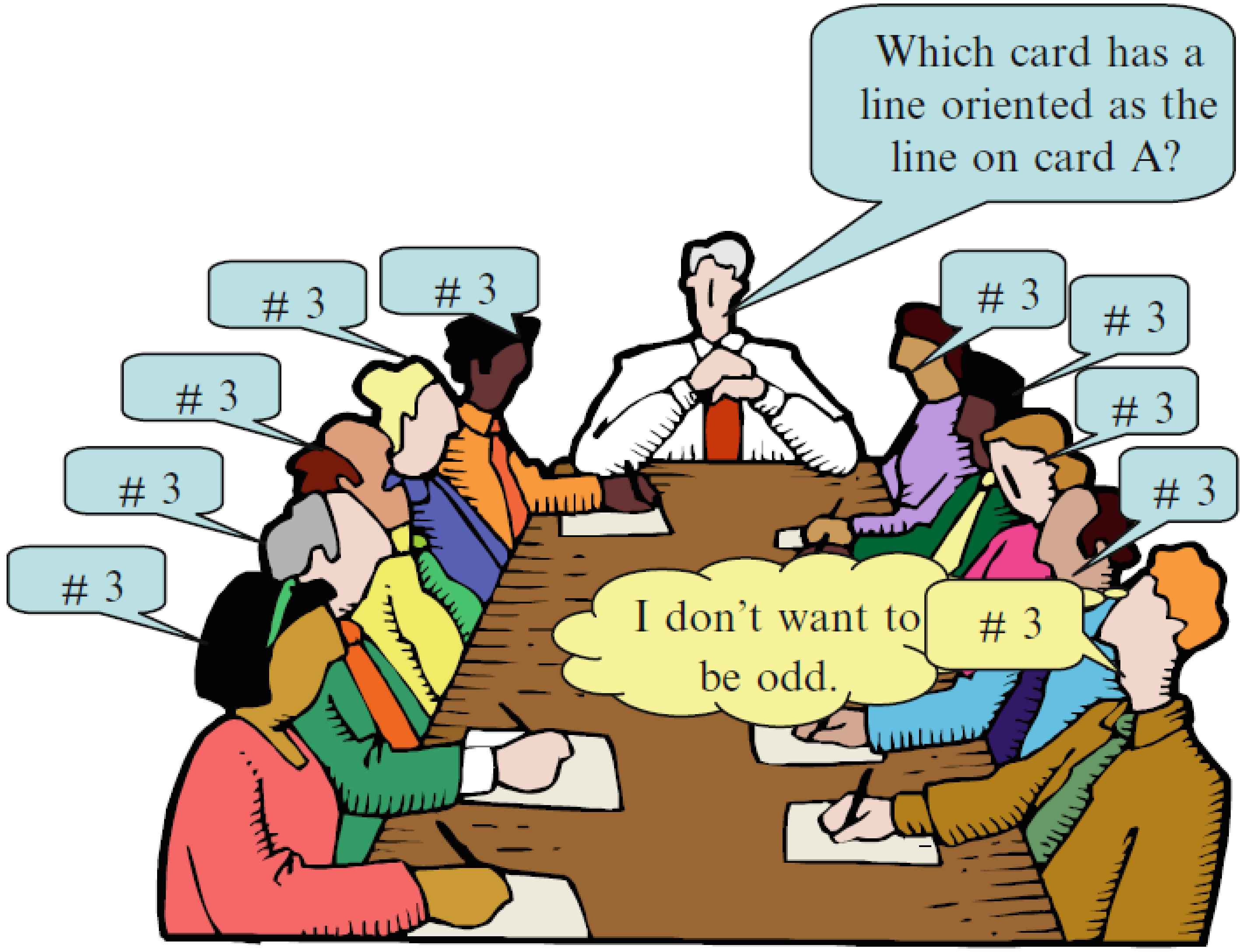
Strategy 2: aggregating recommendations



USERS MAY BE INFLUENCED BY THE GROUP

- Group members may influence each other: process of conformity
- E.g. Conformity experiment by Asch
- People want to be part of the group, even though they have a different opinion
- Or people change their own opinion because they believe the group must be right





Source: Recommender Systems Handbook

GHENT UNIVERSITY

CONCLUSION

- In the past: a lot of (too much?) attention to recommendation algorithms and accuracy
- User experience influenced by many other aspects:
  - Diversity, serendipity, usefulness, …
  - Interface:
    - Explanations, transparency, user control, …
- Many challenges
  - Human-computer interaction
    - Content browsing, selection, …
    - Feedback (ratings)
  - Privacy

FACULTY OF ENGINEERING
AND ARCHITECTURE

Toon De Pessemier
imec – Waves – Ghent University

TECHNOLOGIEPARK-ZWIJNAARDE 15
9052 GHENT, BELGIUM

Email: toon.depessemier@ugent.be

http://www.waves.intec.ugent.be/
http:// www.ugent.be

GHENT
UNIVERSITY