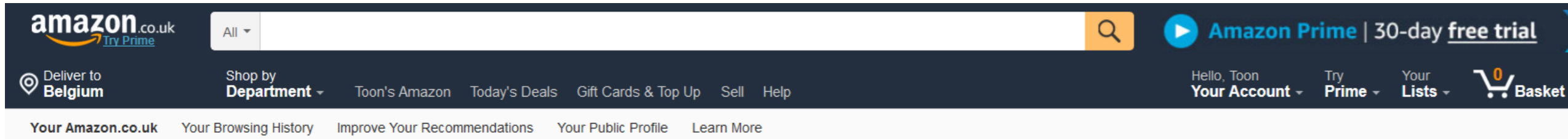


RECOMMENDER SYSTEMS,
THE MAGIC HIDDEN BEHIND THE INTERFACE OF ONLINE SERVICES
OR A CLEVER INTERPLAY BETWEEN SMART ALGORITHMS AND INNOVATIVE
HUMAN-COMPUTER INTERACTION

Toon De Pessemer – May 23, 2018

WHAT ARE RECOMMENDER SYSTEMS?



Your recently viewed items and featured recommendations

Inspired by your purchases

Page 2 of 8 | Start over

Worldwide Travel Adapter, BEZ® the best International Plug [US UK EU AU] with Dual...
★★★★★ 376
£17.99 ✓prime

Tearaway (PlayStation Vita)
PlayStation
★★★★★ 98
PlayStation Vita
£39.99 ✓prime

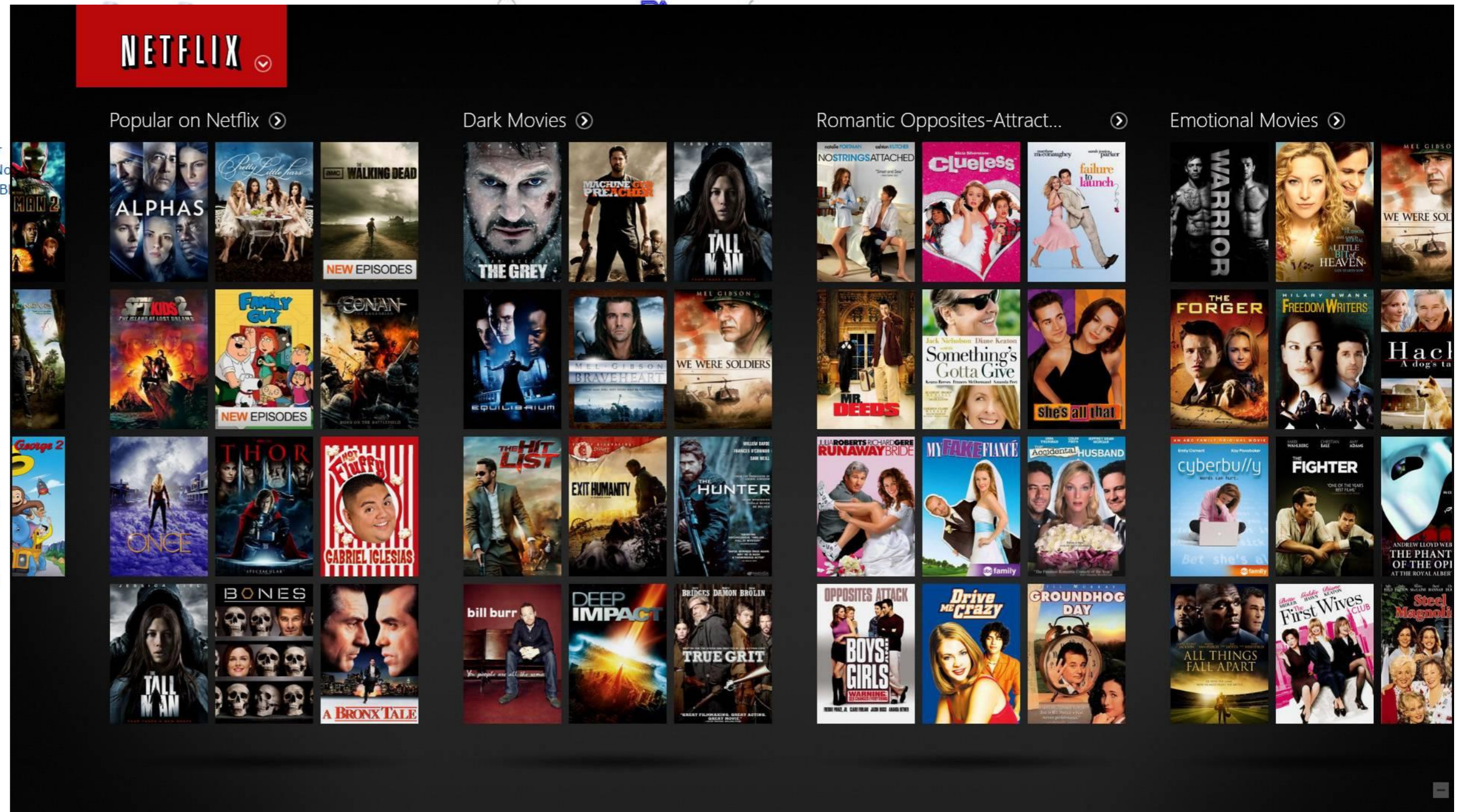
Original battery for Samsung Galaxy Note 3 III LTE, EB-B800B
★★★★★ 21
£11.94

Best Sellers

Come A Little Closer: The breath-taking...
> Rachel Abbott
★★★★★ 282
Kindle Edition
£3.56

The Note: The book everyone's talking...
Zoë Folbigg
★★★★★ 481
Kindle Edition
£1.80

Eleanor Oliphant is Completely Fine: Debut...
> Gail Honeyman
★★★★★ 2,804
Kindle Edition with Audio/Video
£5.67



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

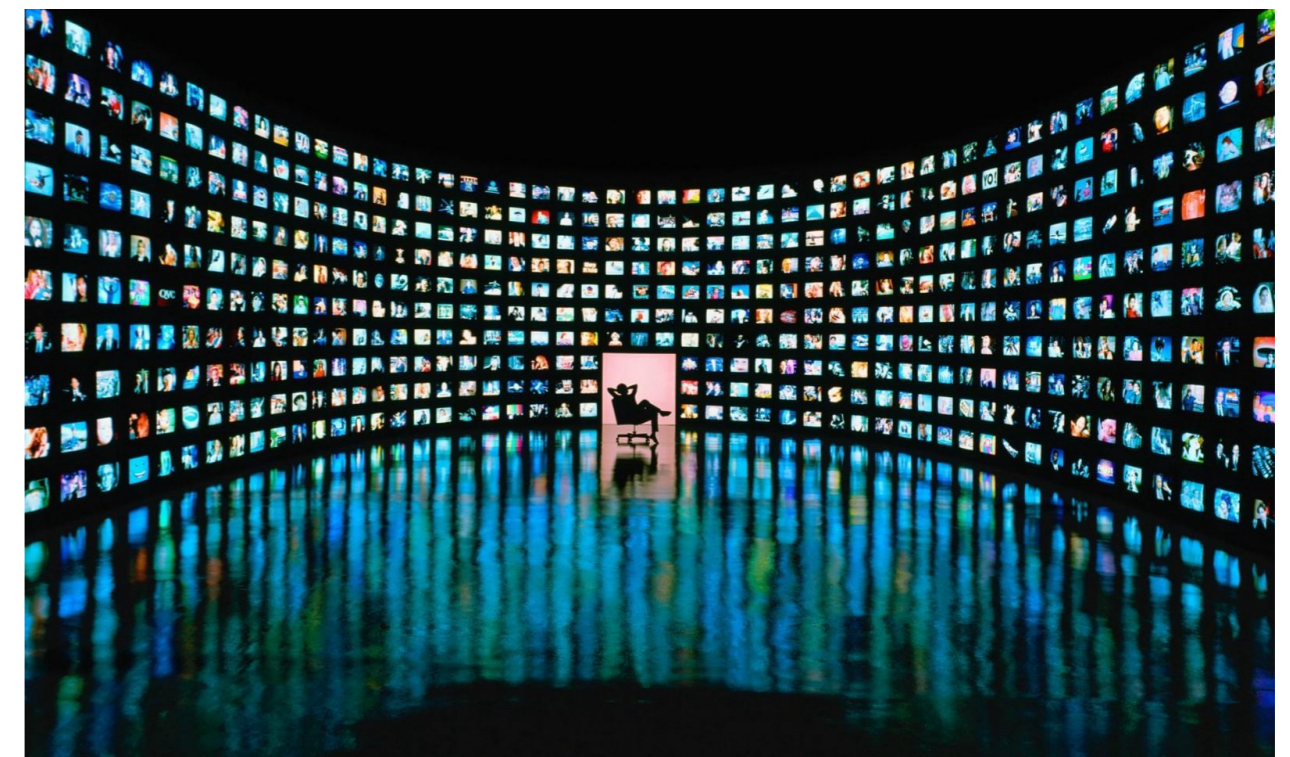
WHY RECOMMENDATIONS ? FROM SCARCITY TO ABUNDANCE

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

Physical world



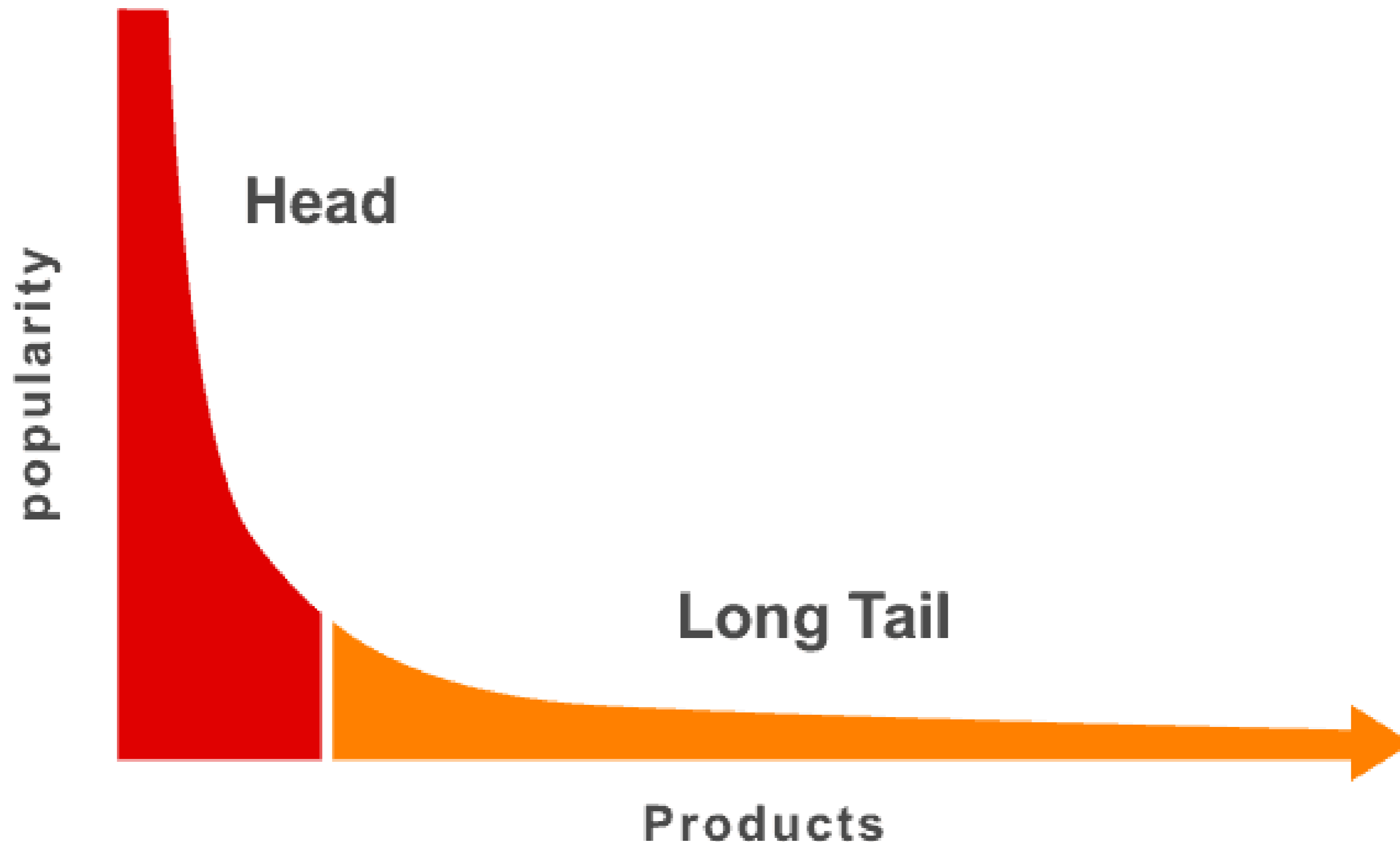
Online world



- A bigger catalog requires better filters

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)

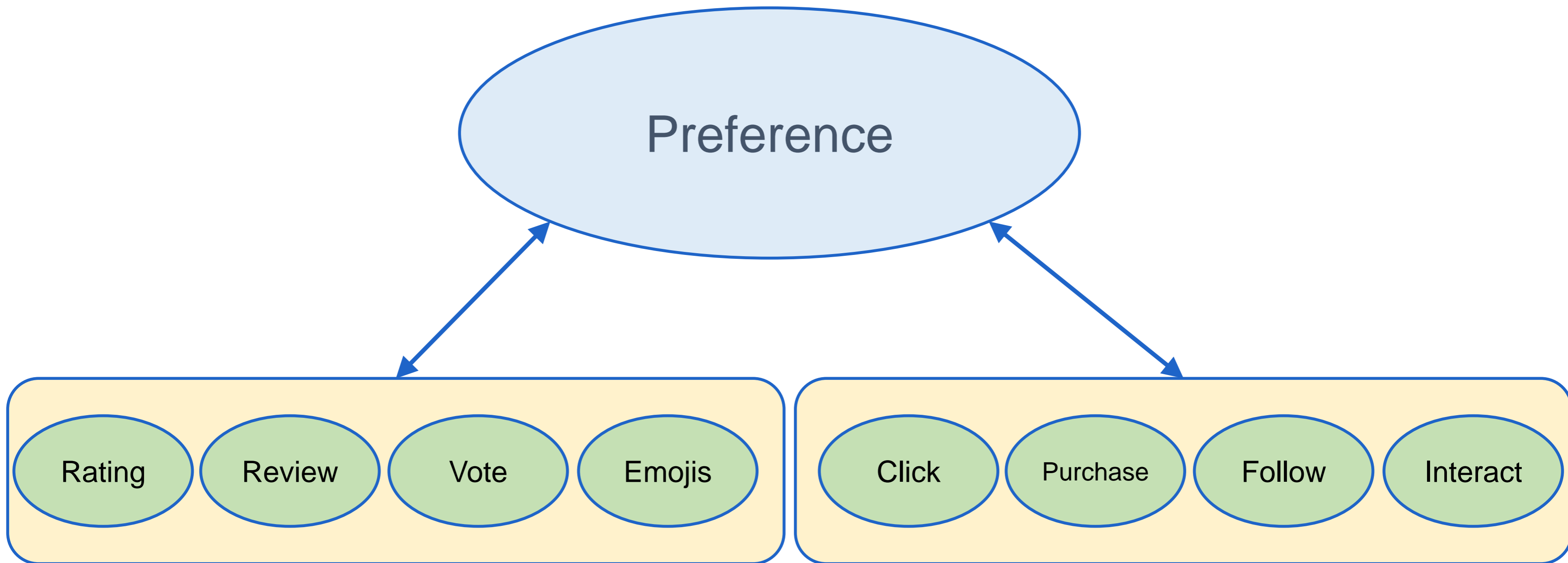
- E.g. Google search
- Information need expressed through a **query**
- Goal: **Retrieve information** which might be useful
- Rather static content base → **indexing** content
- **Dynamic information need**: real-time queries

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

PREFERENCES: THE FUEL OF THE RECOMMENDER ENGINE



DIFFICULTIES WITH RATINGS

IMDb

- 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)



Justice League (2017)
PG-13 120 min - Action | Adventure | Fantasy | Sci-Fi

★★★★★ 7.2/10

46 Metascore

Fueled by his restored faith in humanity and inspired by Superman's selfless act, Bruce Wayne enlists the help of his newfound ally, Diana Prince, to face an even greater enemy.

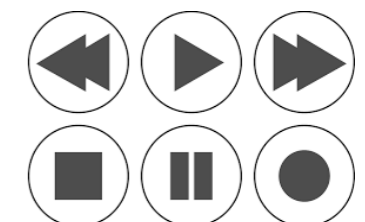
Director: Zack Snyder
Stars: Ben Affleck, Gal Gadot, Jason Momoa, Ezra Miller
Box Office: Weekend: \$0.52M, Gross: \$200.26M

[Watch Trailer](#) [Add to Watchlist](#)

- 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**

News

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



TRIPADVISOR EXAMPLE

The Shed At Dulwich

is this your business?

104 Reviews

#263 of 18,188 Restaurants in London

££ - £££

British, Vegetarian Friendly, Vegan Options

Friem Road, London SE22 0BB, England

+44 7961 568431

Website

Save



All photos (5)

Overview Reviews Location Q&A Details

Overview

5.0

104 reviews



TRAVELLERS TALK ABOUT

- "stews" (3 reviews)
- "tongue" (3 reviews)
- "prawn cocktail" (2 reviews)

Restaurant Hours + Add hours

CUISINES
British, Vegetarian Friendl...
PRICE
££ - £££

RATINGS
Service
Vibe
Food

LOCATION
Friem Road, London SE22 0BB, England

All Details | Improve This Listing



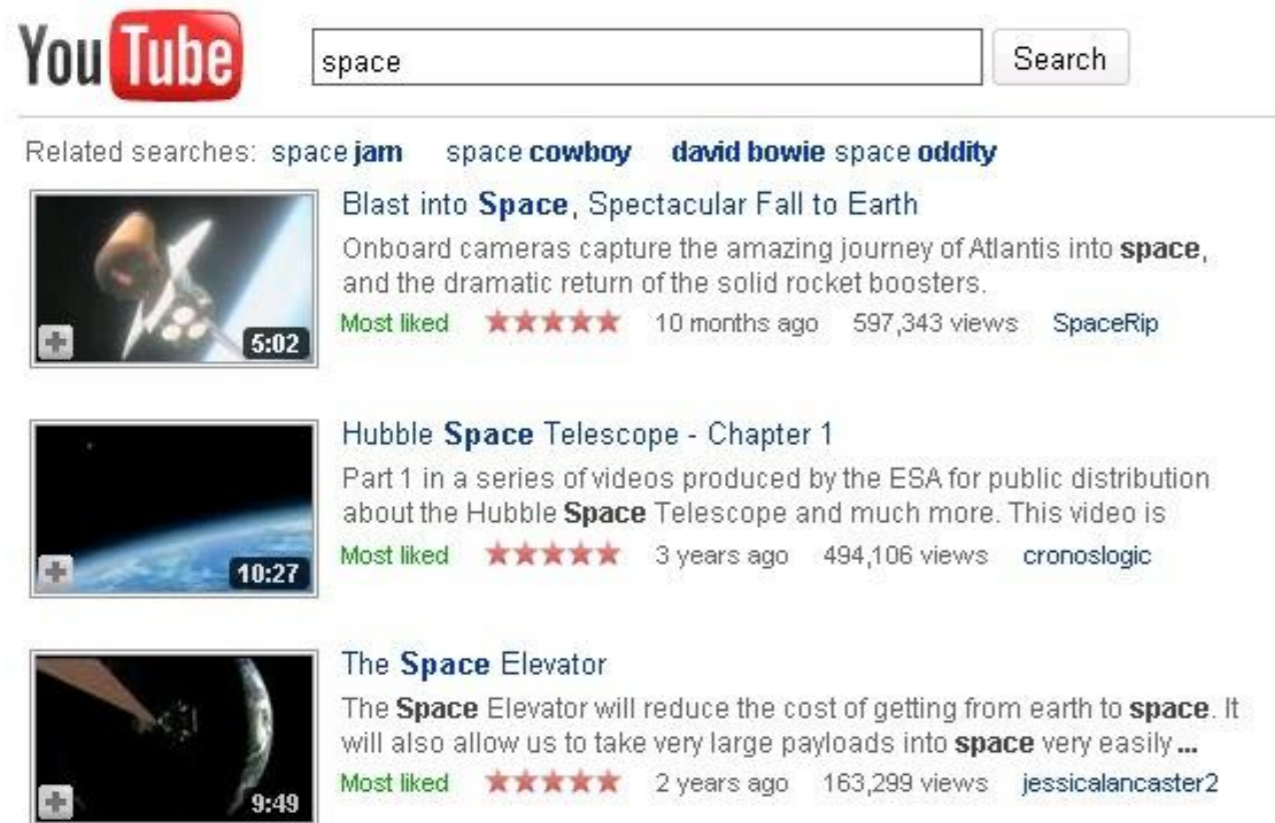
TRIPADVISOR EXAMPLE

- 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

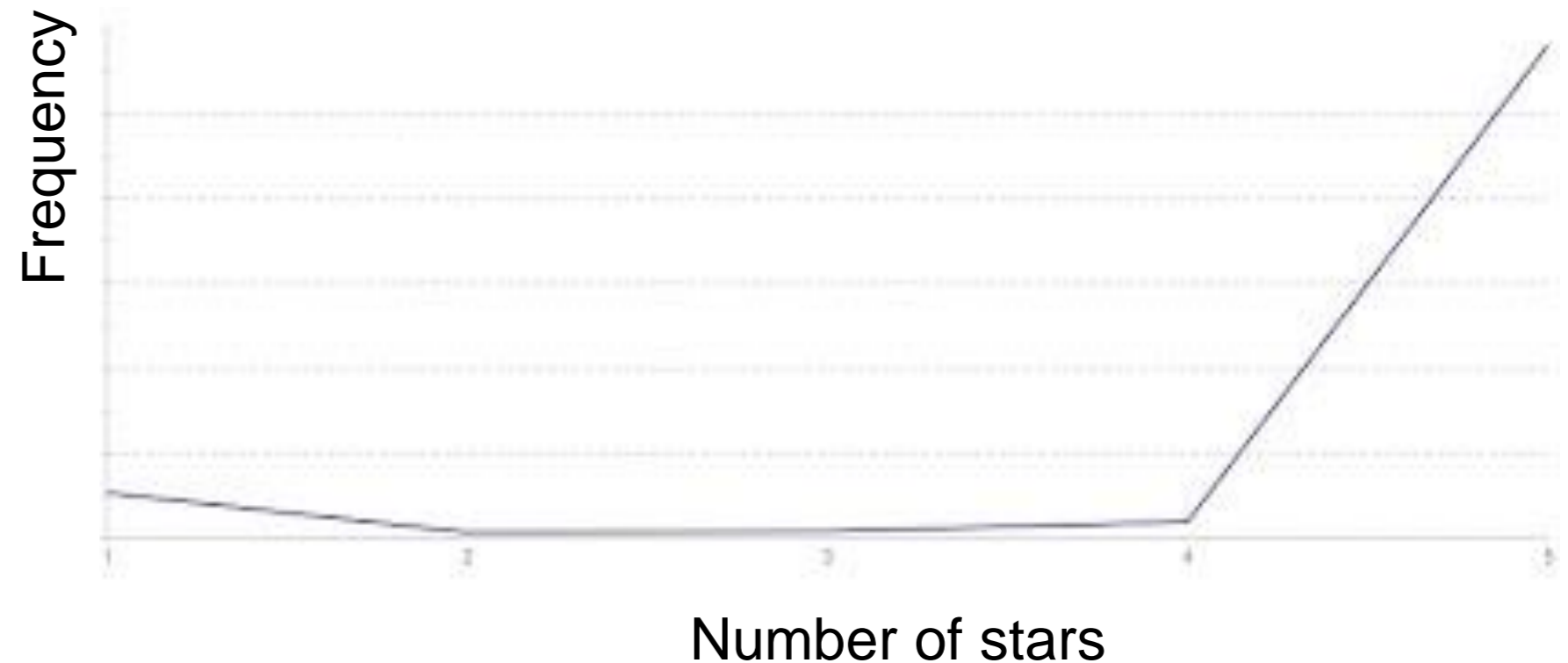


YOUTUBE EXAMPLE (2009)

■ Five Stars Dominate Ratings



The screenshot shows the YouTube search interface for the term 'space'. The search bar contains 'space' and the 'Search' button is visible. Below the search bar, there are related search suggestions: 'space jam', 'space cowboy', 'david bowie', and 'space oddity'. Three video results are displayed, each with a thumbnail, title, description, and a 5-star rating. The first video is 'Blast into Space, Spectacular Fall to Earth' by SpaceRip, with 597,343 views. The second is 'Hubble Space Telescope - Chapter 1' by cronologic, with 494,106 views. The third is 'The Space Elevator' by jessicalancaster2, with 163,299 views.



- 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

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



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



INTUITIVE HUMAN-DEVICE INTERACTIVITY

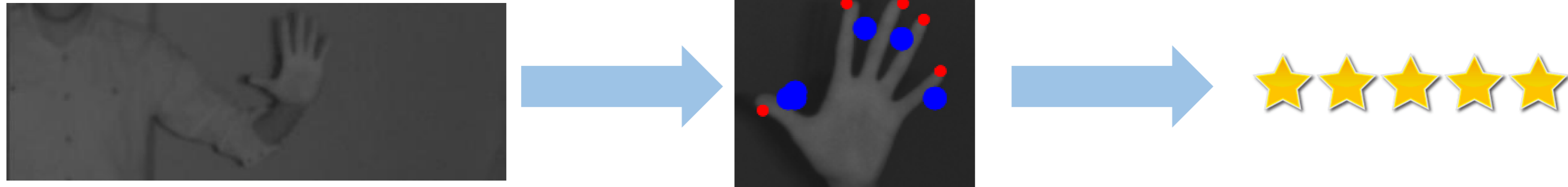
Example: Microsoft Kinect as motion sensing input device



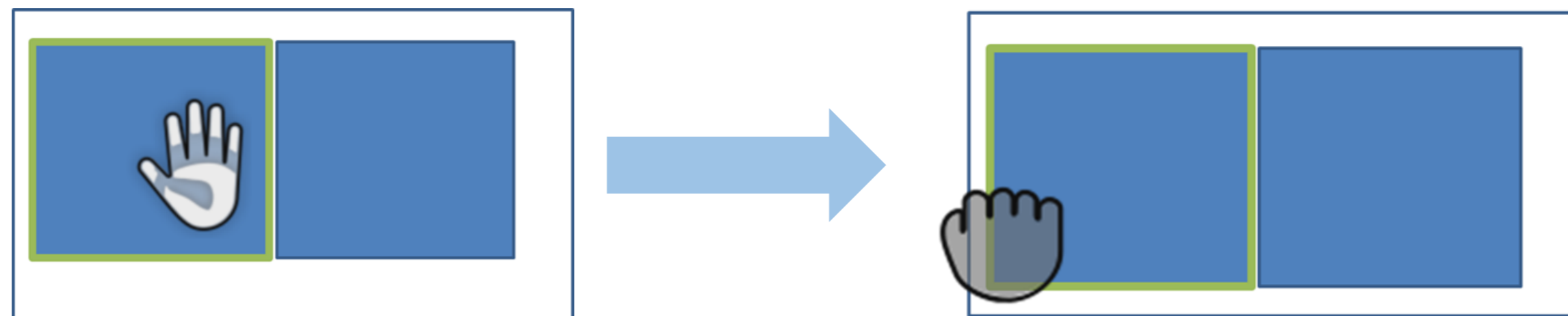
- **Speech recognition:** video control
- **Text-to-speech:** feedback to the user
- **Tracking movements:** browsing, content selection, explicit feedback
- **Facial recognition:** user identification (authentication)
- **Localization of a sound source:** context detection
- **Emotion recognition:**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



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



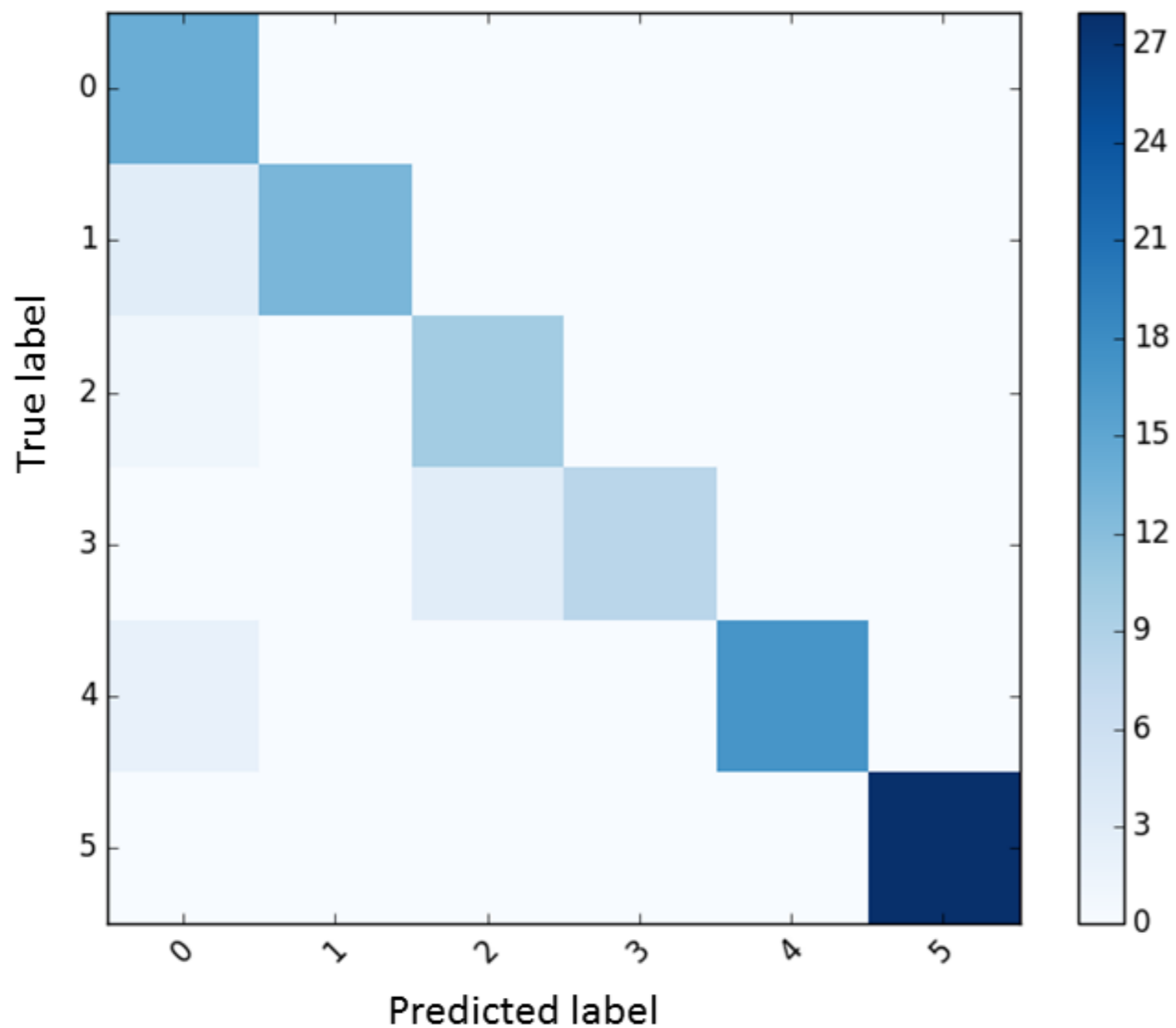
PROBLEM OF INTUITIVE INTERACTION

- 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

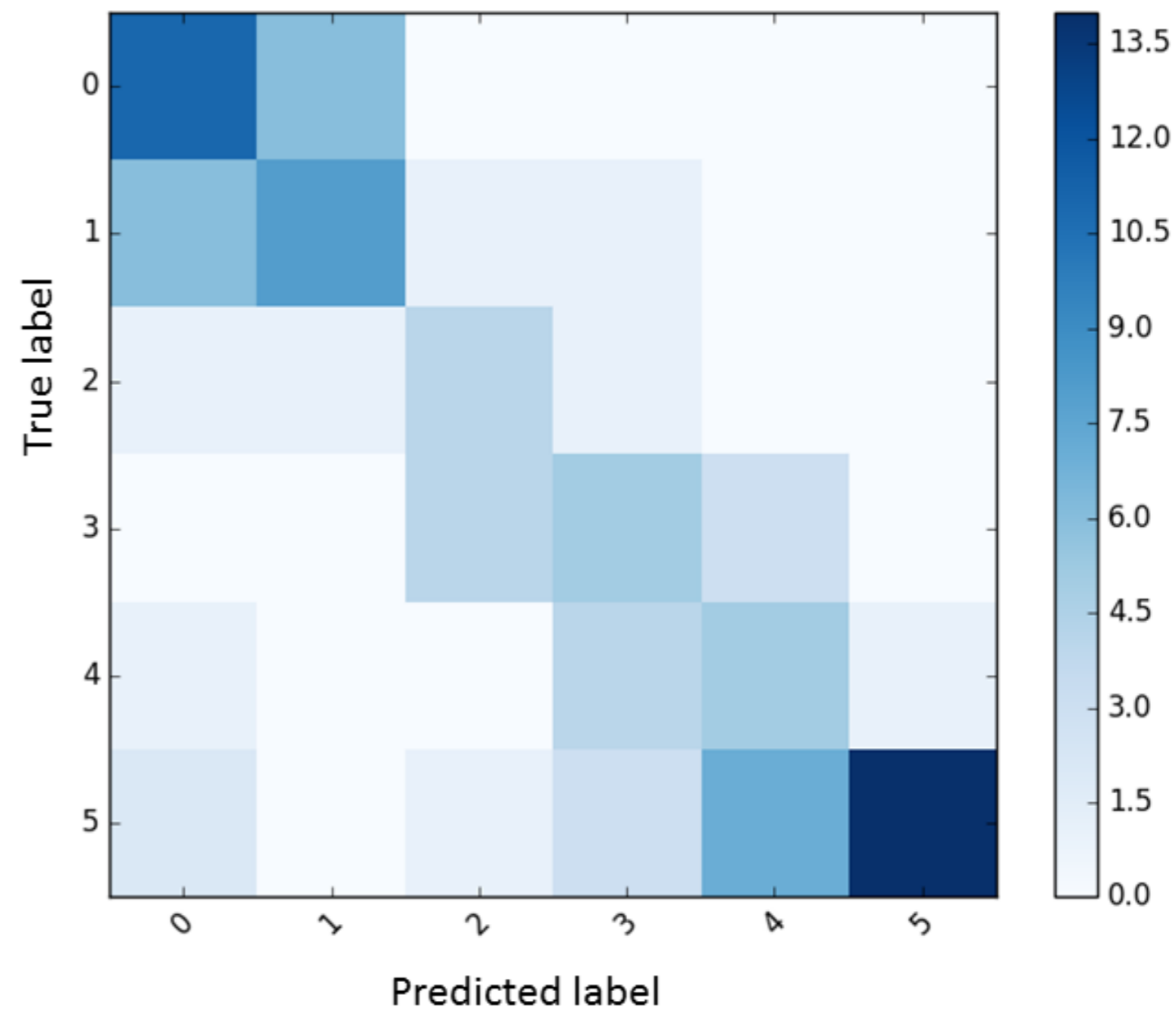


FINGER TRACKING

Distance: 1 to 2.5 meters

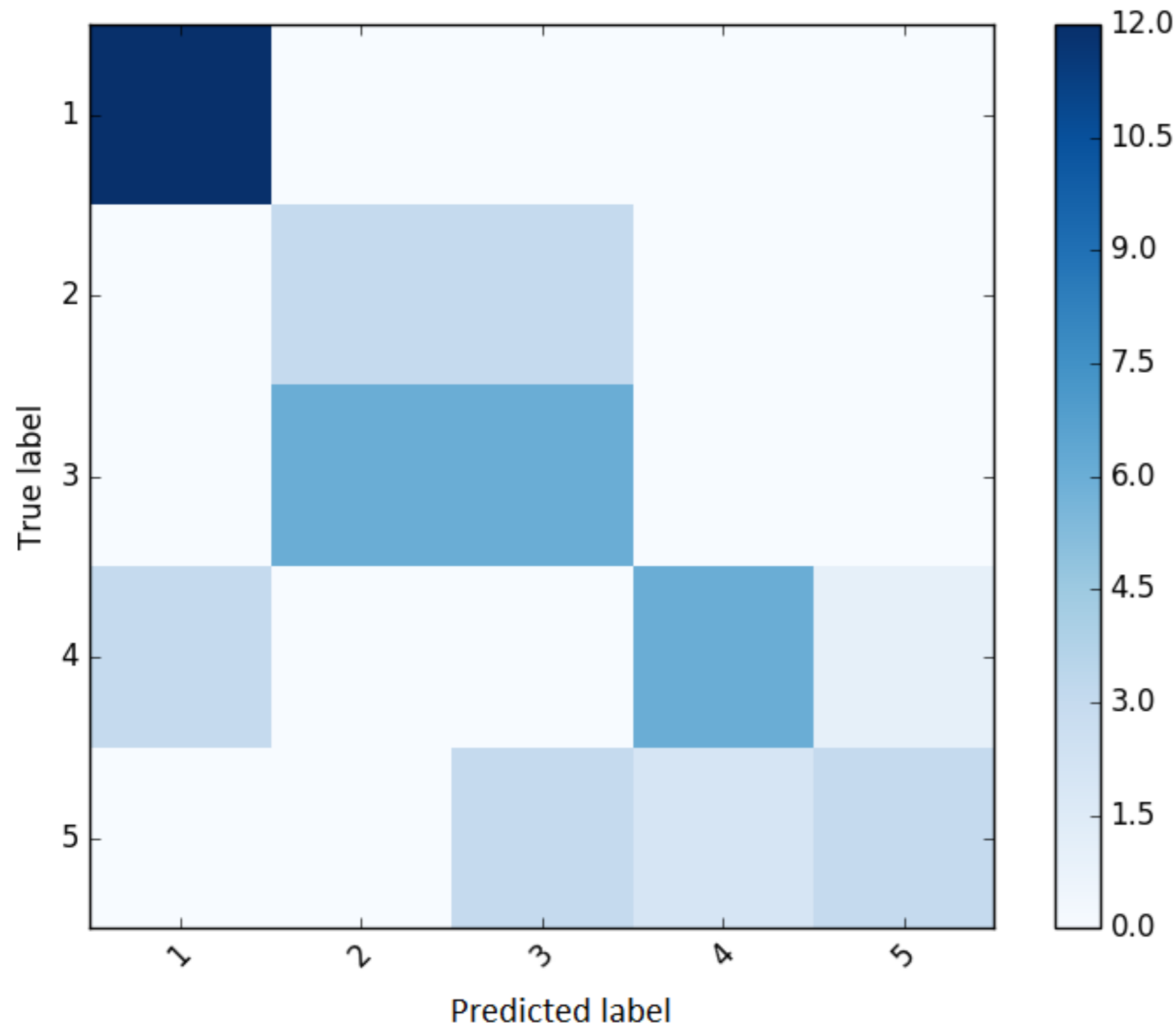


Distance: further than 2.5 meters



HAND WRITING

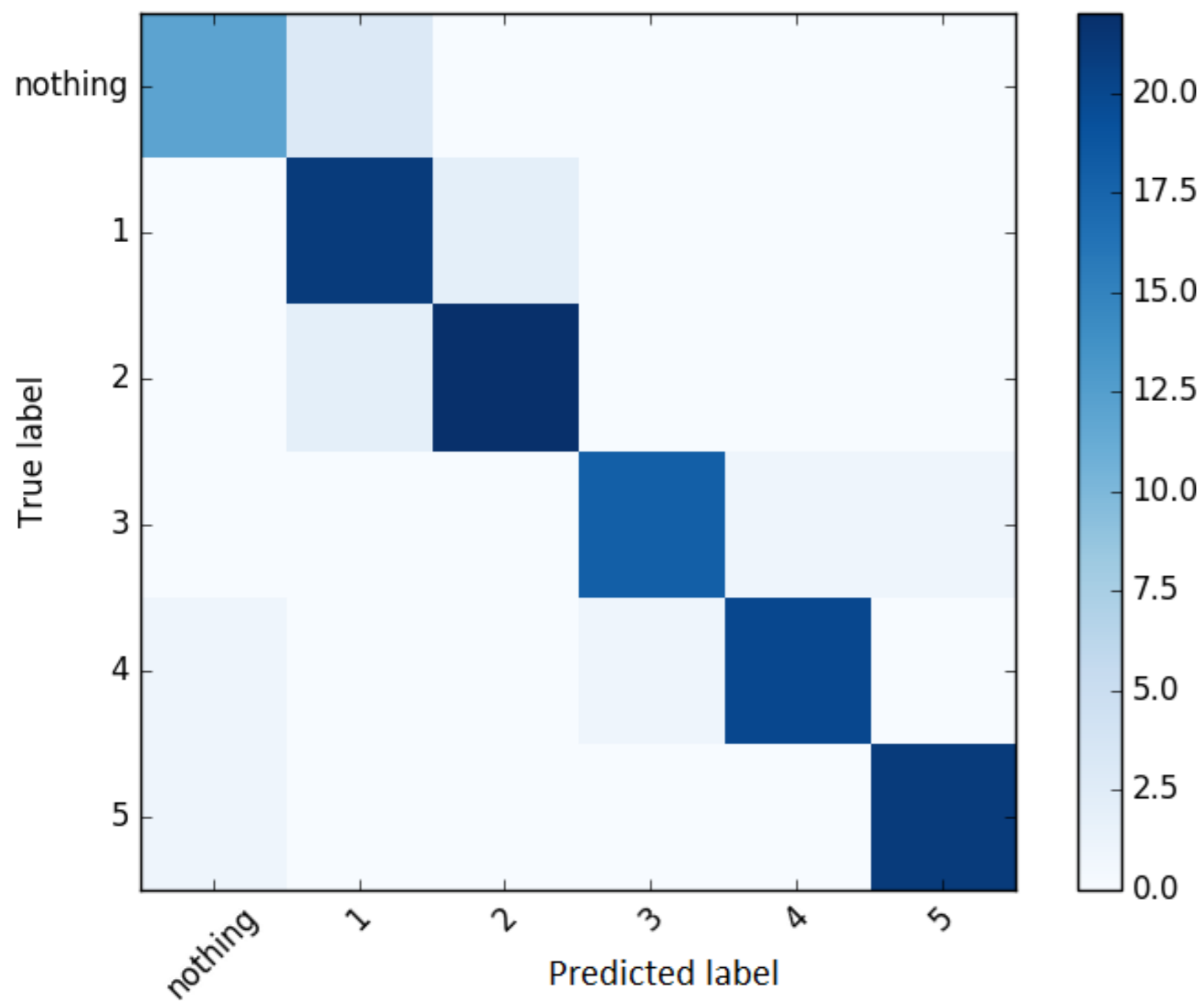
Distance: 1 to 2.5 meters



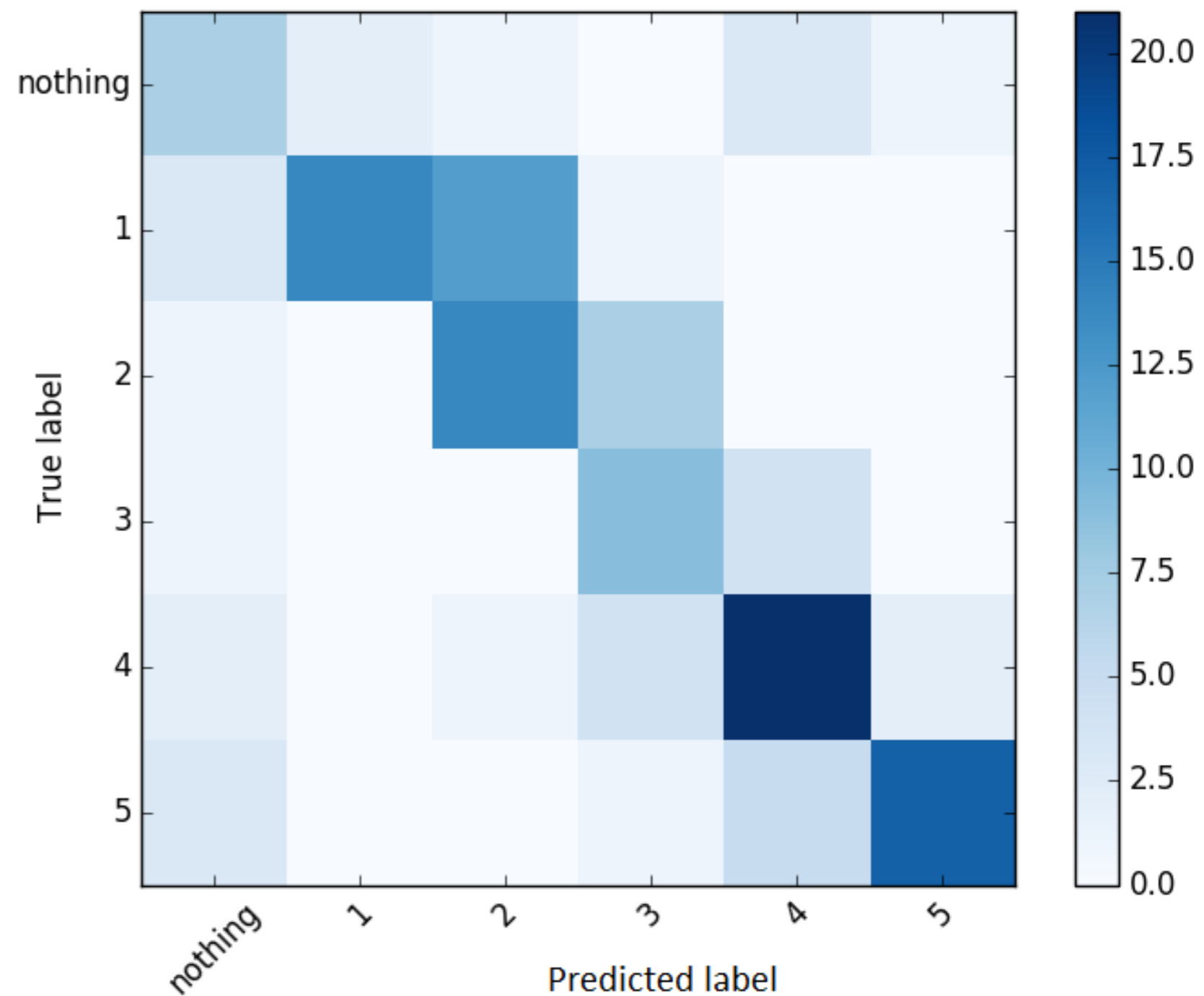
- Rather low accuracy:
30/48 correct
- Difficulties to recognize the
begin and end of a hand writing
gesture

SPEECH RECOGNITION

Distance: within 3.5 meters



Distance: further than 3.5 meters



USER IDENTIFICATION

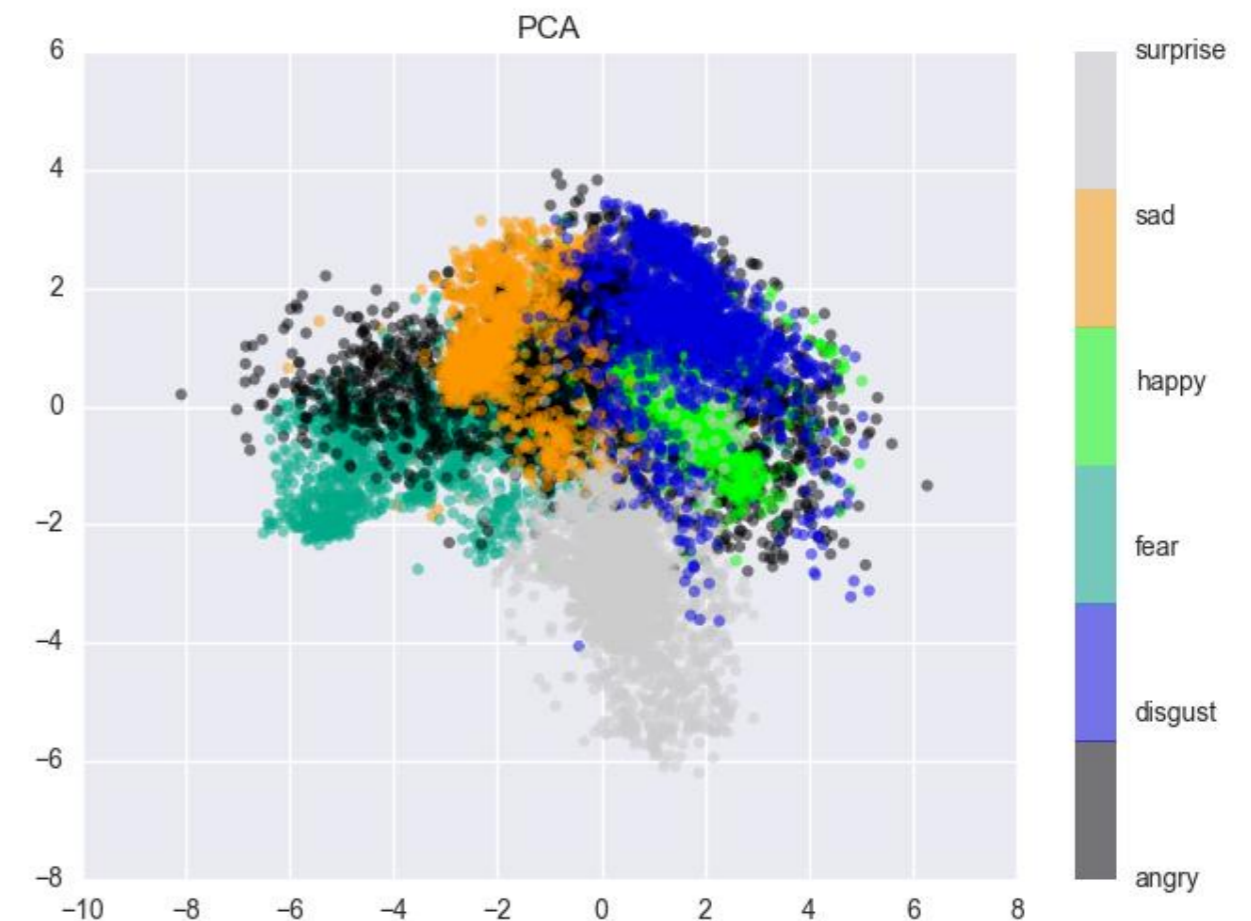
- Classic solution: username + password
- Nobody wants to input username/password on TV
 - Especially not with a remote control
- But TV is a shared device
- Solution: automatic user identification
- Disadvantage: privacy aspects



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



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 identify **key scenes** in the content



NON-PERSONALIZED RECOMMENDERS

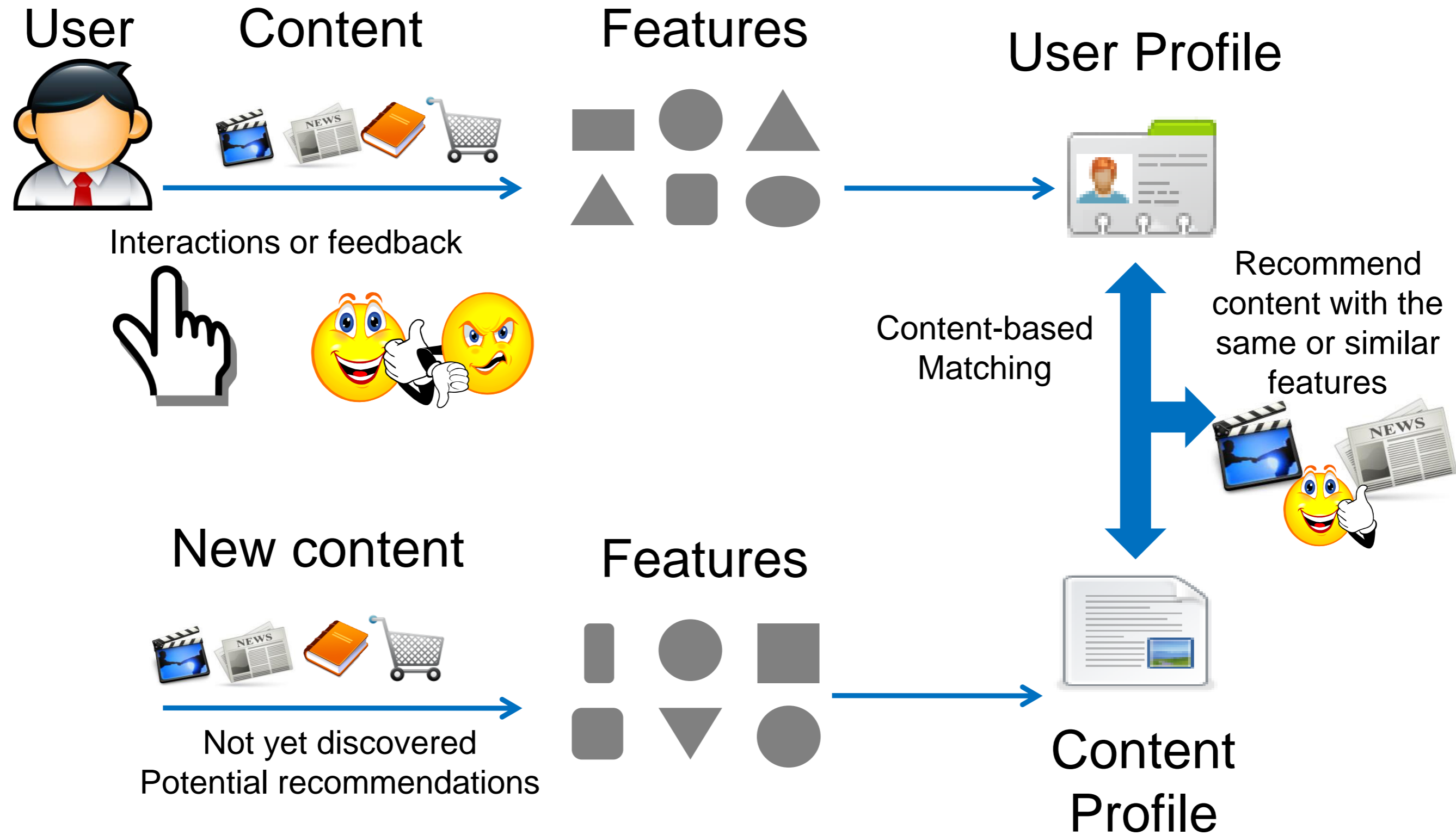
TYPE OF NON-PERSONALIZED RECOMMENDATIONS

- **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)



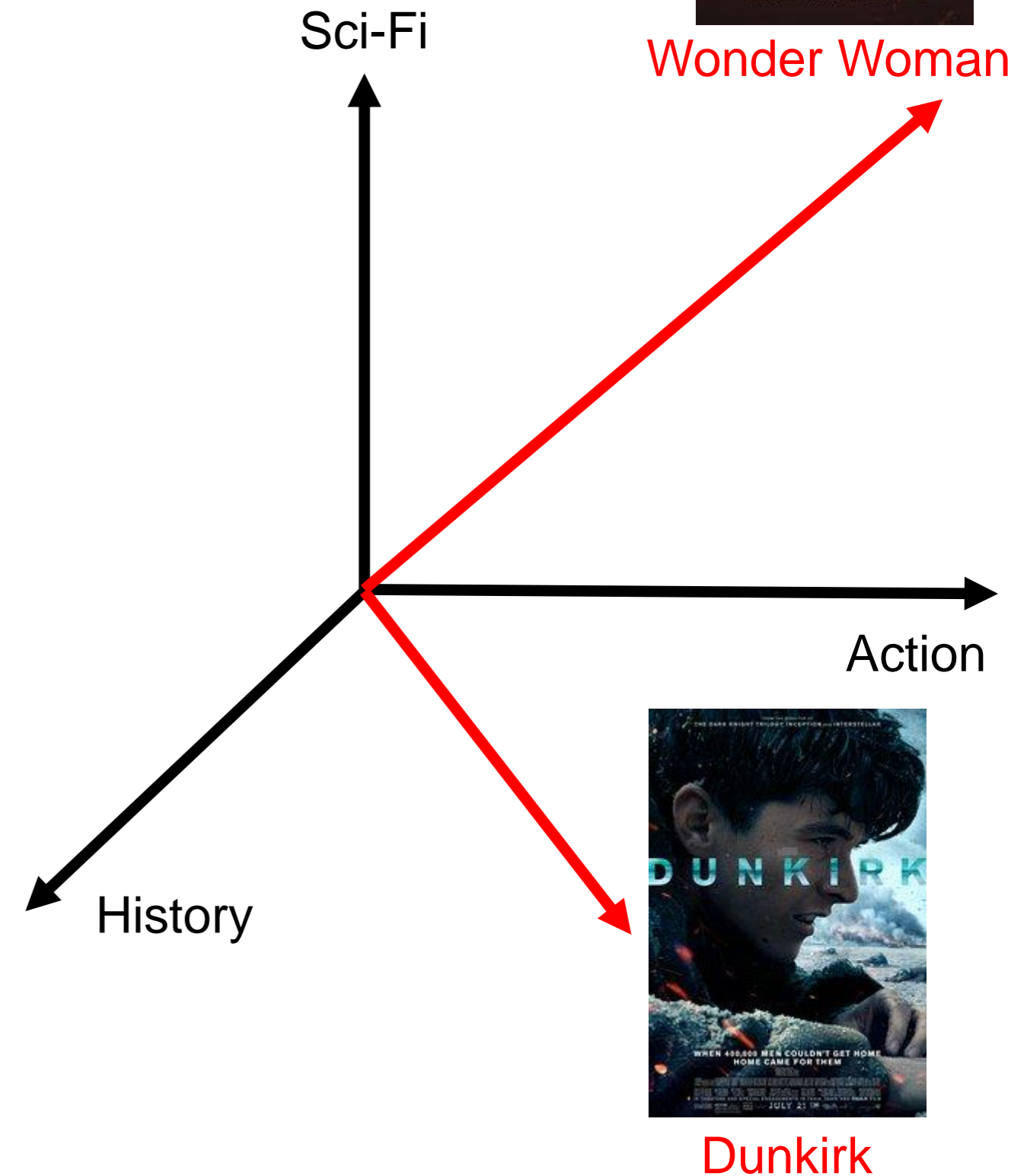
CONTENT-BASED FILTERING

CONTENT-BASED FILTERING



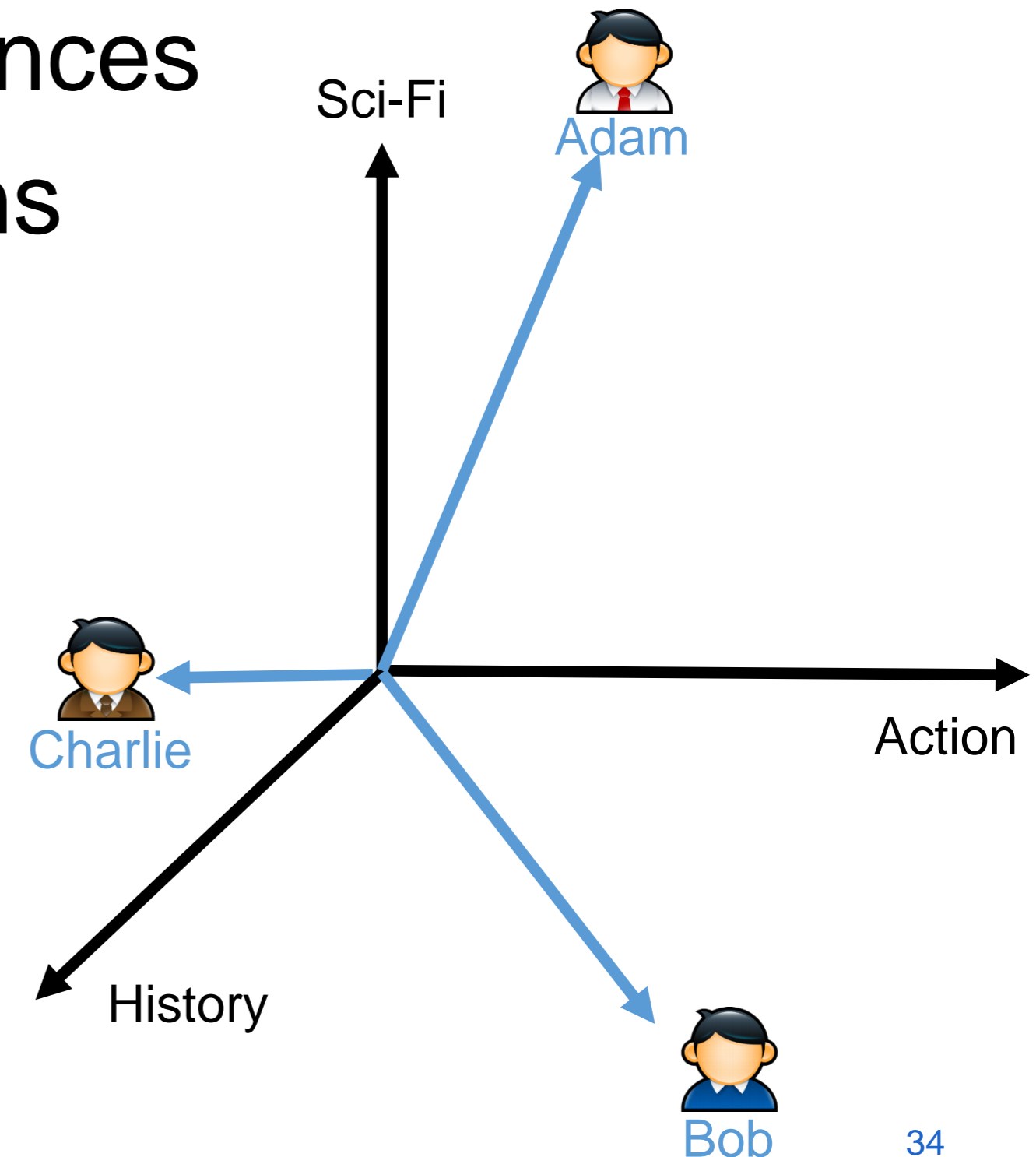
VECTOR SPACE MODEL: ITEMS

- Universe of attributes defines a content space
- Each attribute is one dimension
- Item = set of attributes
=position in the content space
→ position defines vector



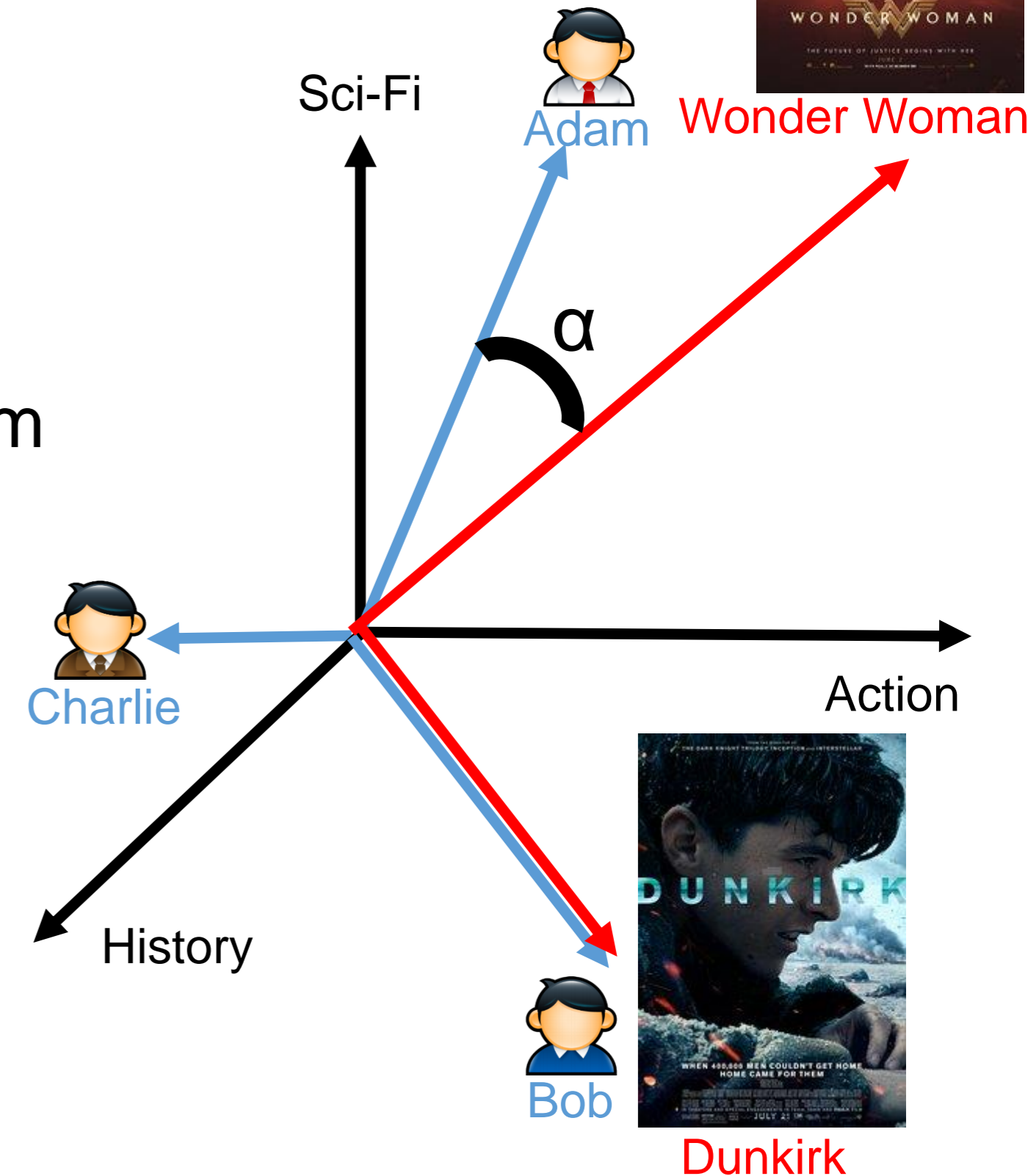
VECTOR SPACE MODEL: USERS

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



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




EXAMPLE: CLOTHING DOMAIN

AFFINITY

Recommendations based on personal style

AFFINITY HOW IT WORKS SHOPS CLOTHING SHOES ACCESSORIES SALE

Sign Up Log In



Shop in your style

Take our style quiz and discover the best fashion from across the web, personalized just for you.



Shop all your favorite stores in one place.



Discover new brands and designers you'll love.



Shop in your style.

Often rather stable preferences: Style, size, gender

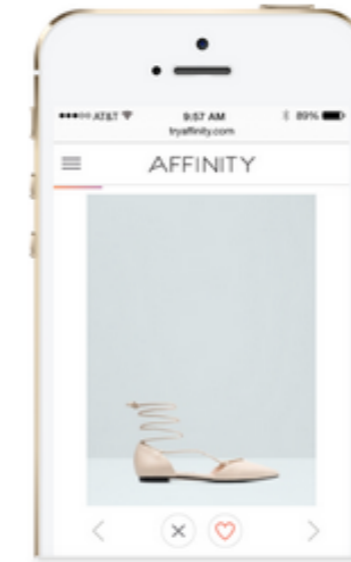
CLOTHING: HOW TO OBTAIN RECOMMENDATIONS?

AFFINITY

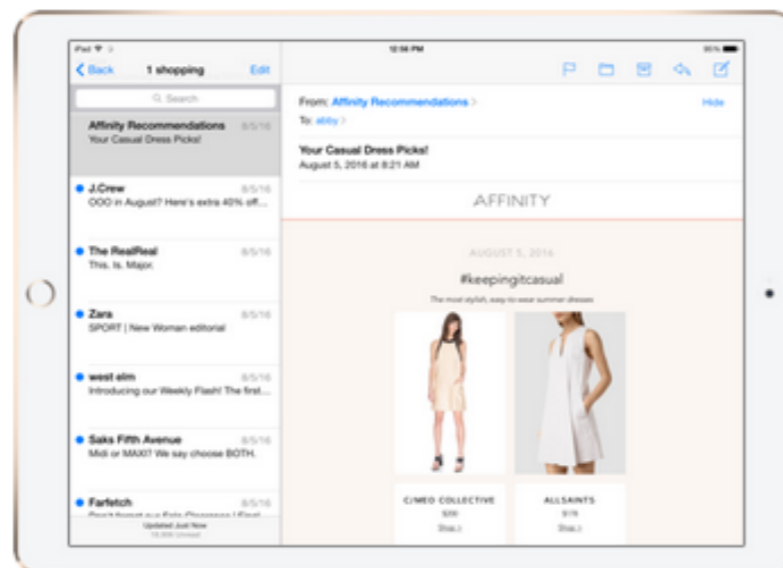
HOW IT WORKS

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

CLOTHING: STYLE QUIZ

AFFINITY

BEFORE WE BEGIN

Help us learn your style by rating items in our style quiz.

How it Works:



items you like



items you don't

GET STARTED

*Profile Building =
Explicit feedback
for specifically
selected items*

CLOTHING: STYLE QUIZ

AFFINITY

AFFINITY

RATE SOME ITEMS



RATE SOME ITEMS



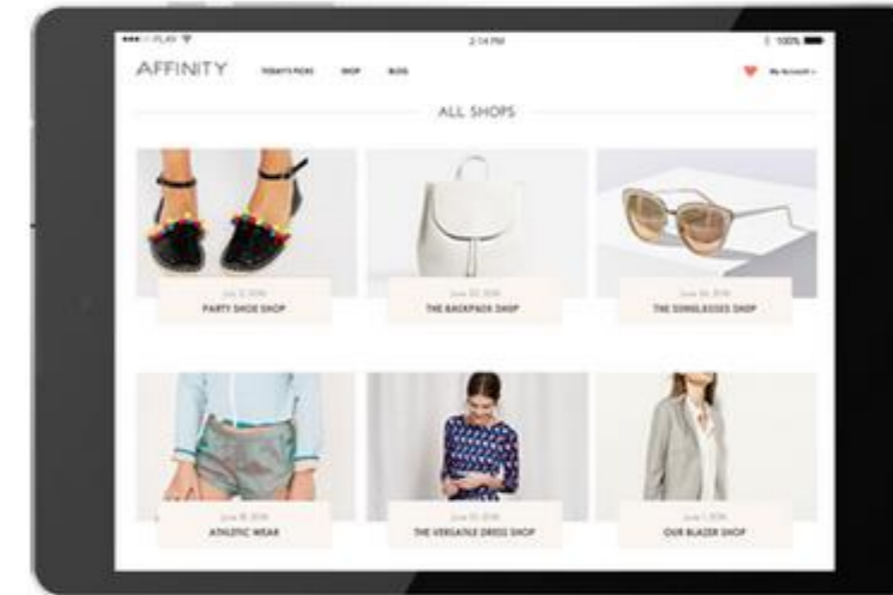
*Profile Building =
Explicit feedback
for specifically
selected items*

CLOTHING: HOW TO OBTAIN RECOMMENDATIONS?

AFFINITY

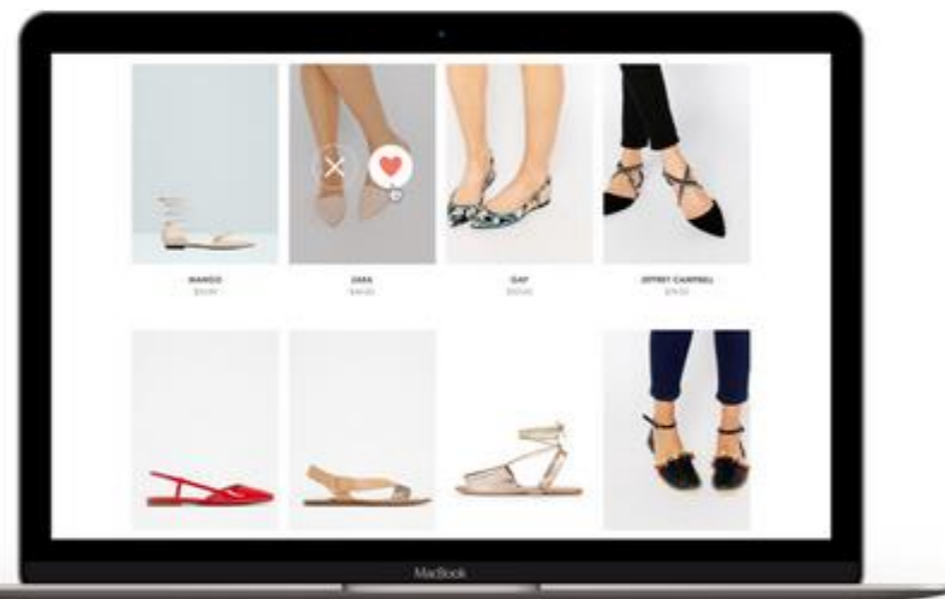
SHOP IN YOUR STYLE

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

EXAMPLE: NEWS DOMAIN

News Recommendation



Facebook shut down 1.3 billion fake accounts in the last six months
Yesterday



Boris Johnson battles to save the Iran Nuclear Deal as he discusses how to deal with Trump
7 hours ago



Villagers flee civil war in Myanmar by ELEPHANT
8 hours ago



Instagram testing feature to tell how much time you spend on app
Yesterday



QUENTIN LETTS: Tom Watson gave a display greasy enough to test one's digestion
9 hours ago



Katherine Parkinson's war in Georgia
Yesterday



World Cup 2018: Mario Gotze misses out on Germany's provisional squad
Yesterday



Sign in

News

Sport

Weather

Shop

Earth

Travel

More

Search

SPORT 2018 FIFA World Cup™



Home

Football

Formula 1

Cricket

Rugby U

Tennis

Golf

Athletics

Cycling

World Cup

All Sport

World Cup > Groups & Schedule | Scores & Fixtures | Top Scorers

All Scores & Fixtures | All Teams | Leagues & Cups

World Cup 2018: Mario Gotze misses out on Germany's provisional squad

15 May 2018 | World Cup | 144

Share



World Cup moments: Glory for Gotze

Top Stories



Allardyce leaves Everton after six months

5m | Everton | 104



Southgate set to reveal England squad with Hart & Wilshere to miss out

11h | England | 16



'Lomachenko having biggest impact on a sport since Tiger Woods'

10h | Boxing | 114

From Around the Web

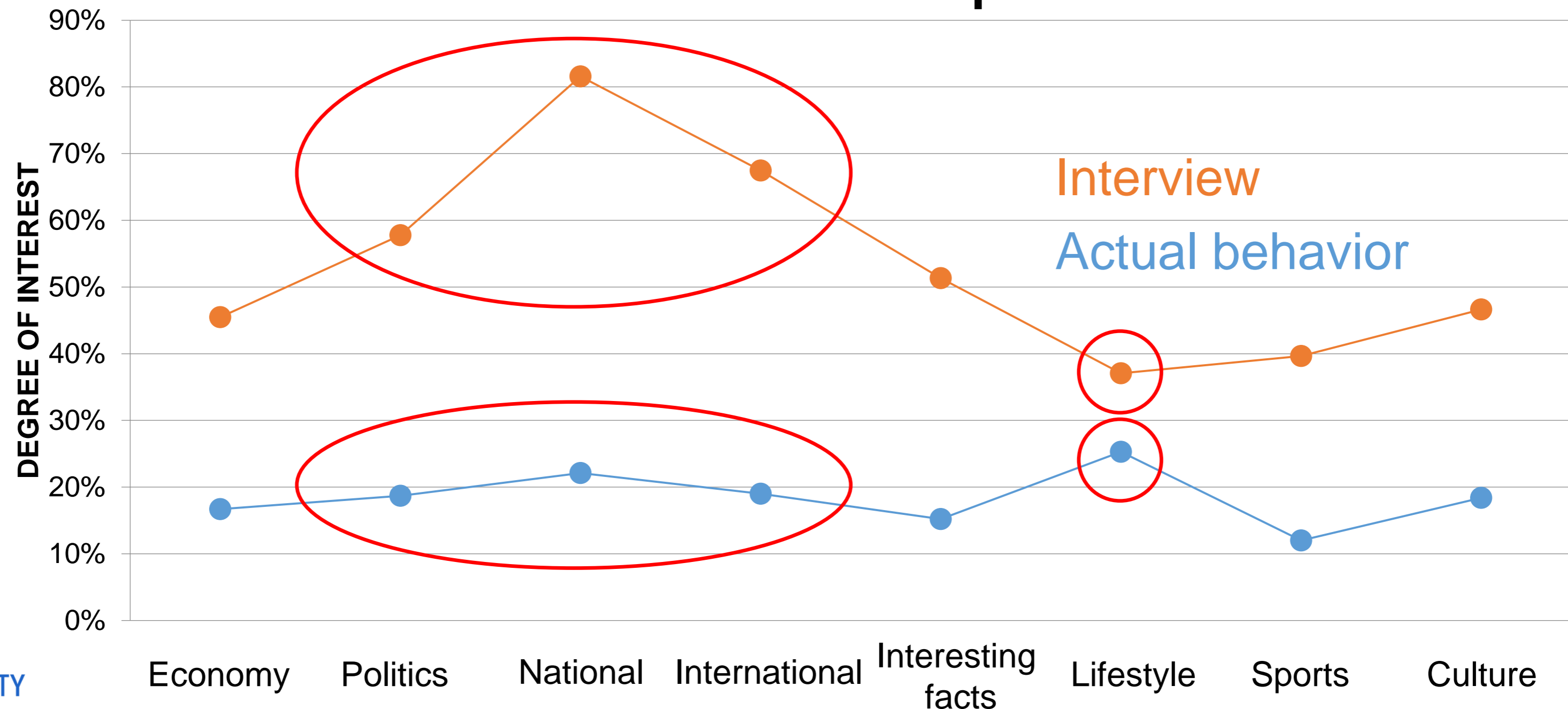
Promoted content by Outbrain

Also in Sport

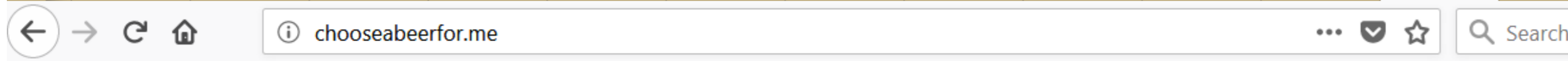
BUILDING A NEWS PROFILE BASED ON CATEGORIES

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

Partitioning of news categories based on interviews and actual news consumption behavior



EXAMPLE: CASE-BASED REASONING



Beer Recommendation Engine

Search for a Beer and Select from the Menu

Guinness Black Lager

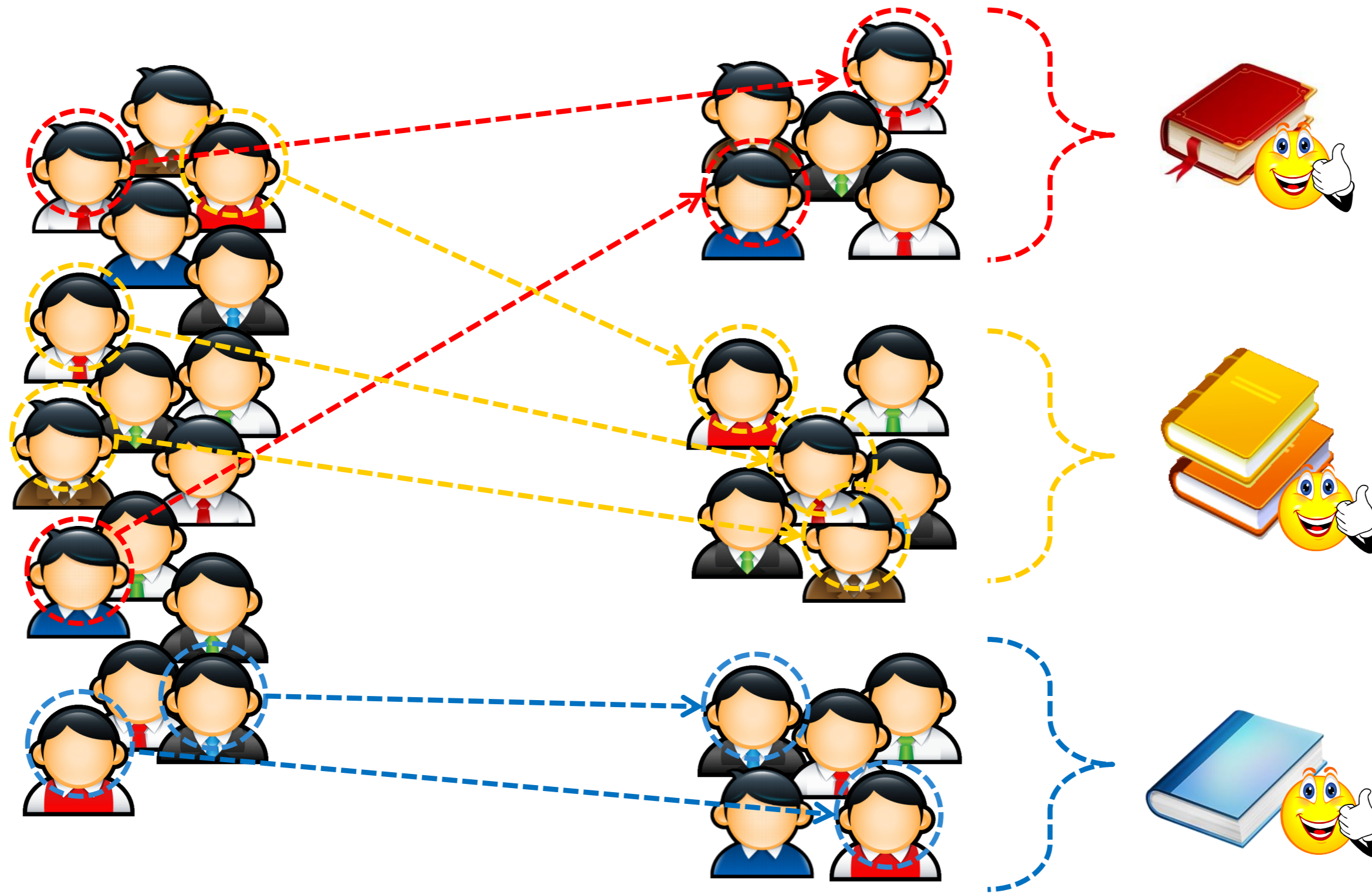
[Guinness Black Lager](#) by [Guinness Ltd.](#) can be described as:
stout coffee chocolate euro schwarzbier

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 |

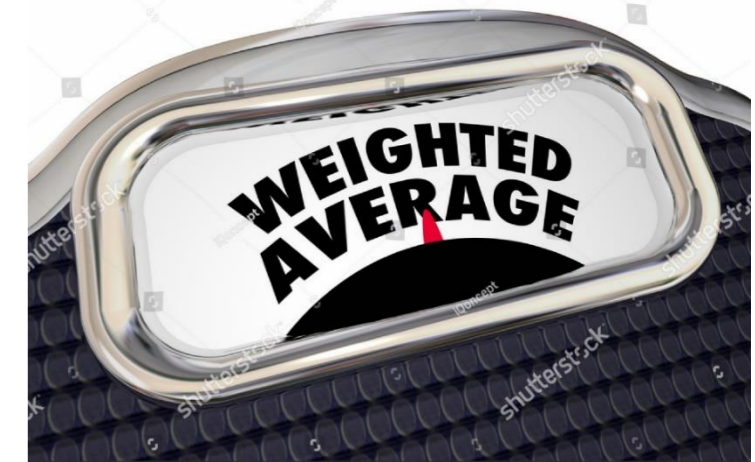
COLLABORATIVE FILTERING

COLLABORATIVE FILTERING



COLLABORATIVE FILTERING

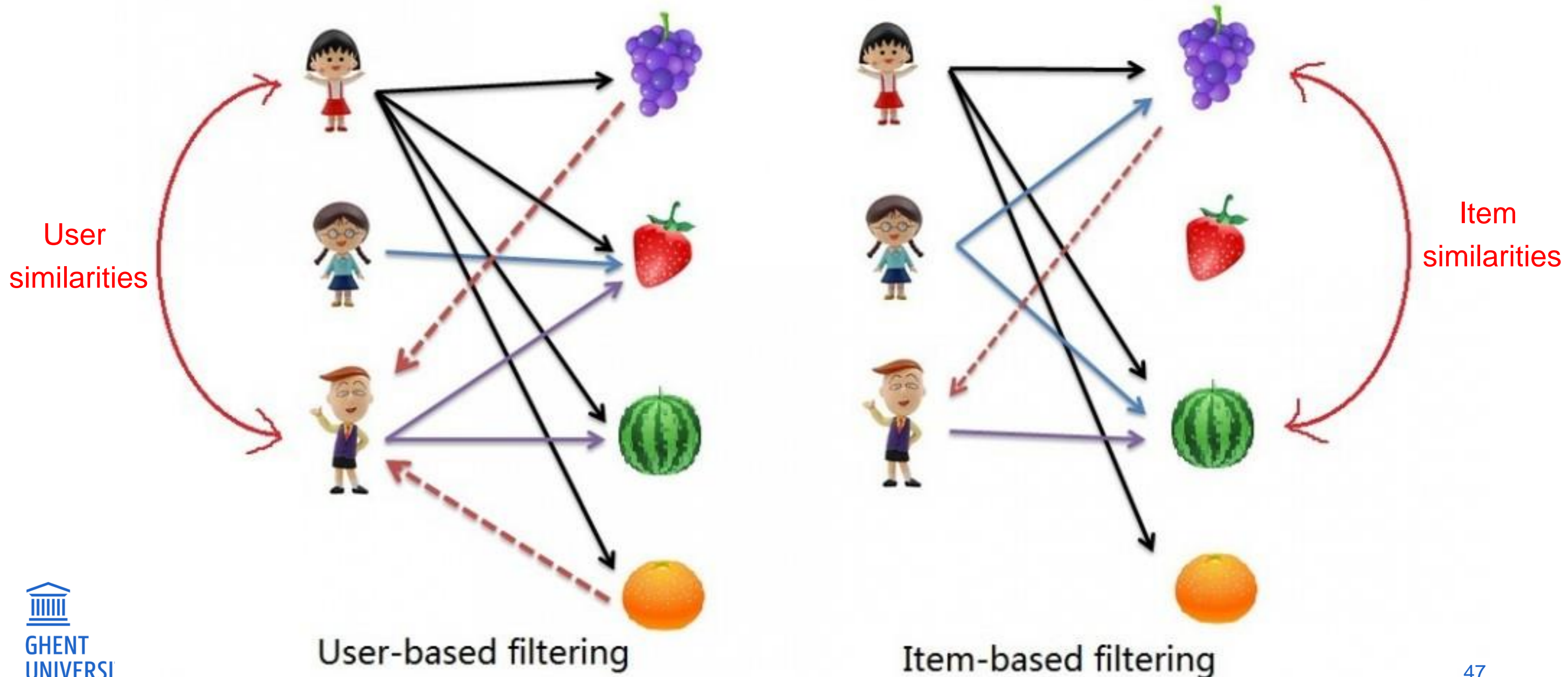
- Rating prediction for a target user and target item
= Weighted average of similar users' ratings for that item
- Weight reflects agreement between the two users
= Correlation in rating behavior



Users agree on items they like
and on items they dislike

ITEM-ITEM COLLABORATIVE FILTERING

- Idea: compare items instead of users



HYBRID RECOMMENDERS

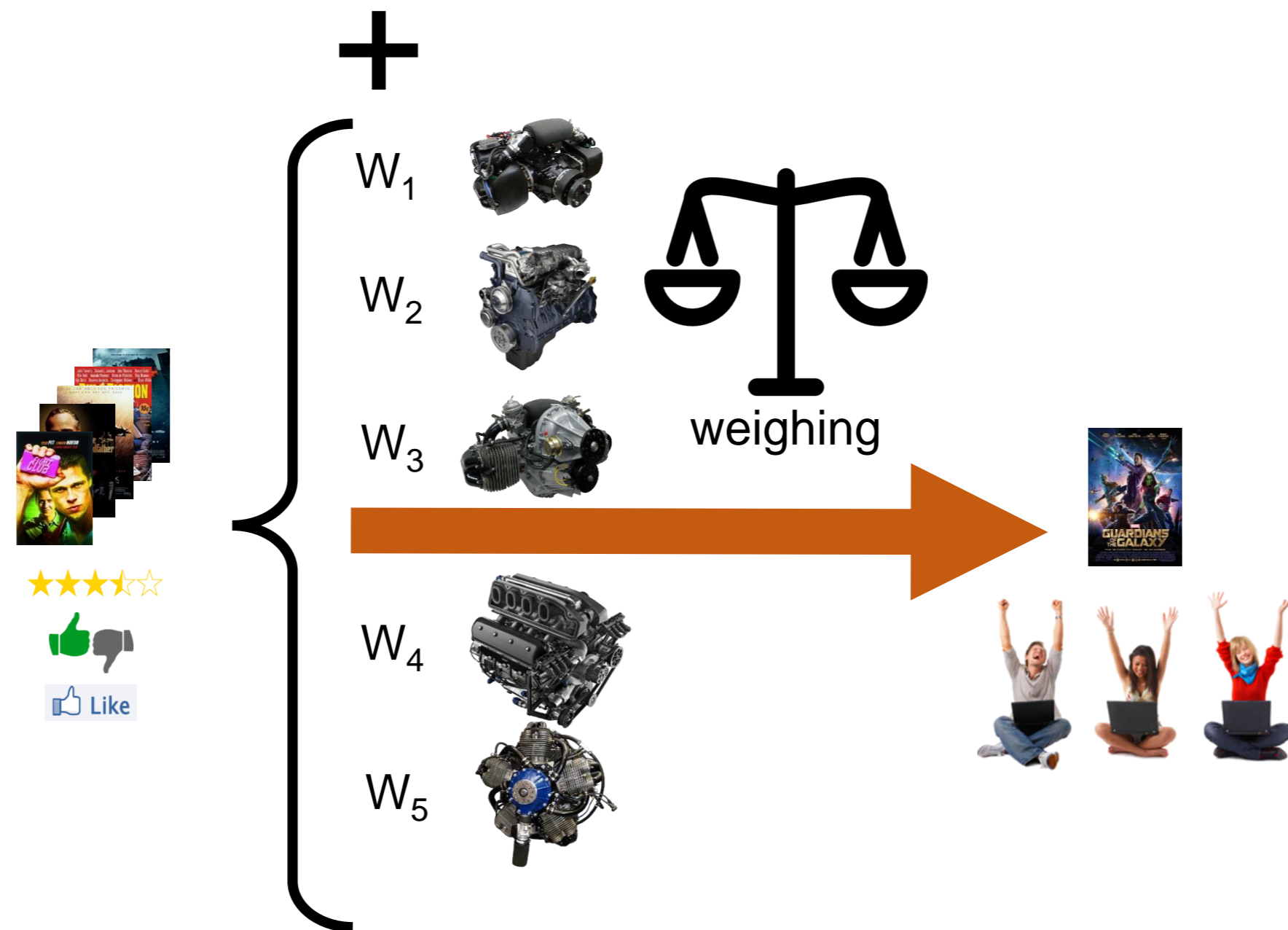
HYBRID RECOMMENDERS?

- 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

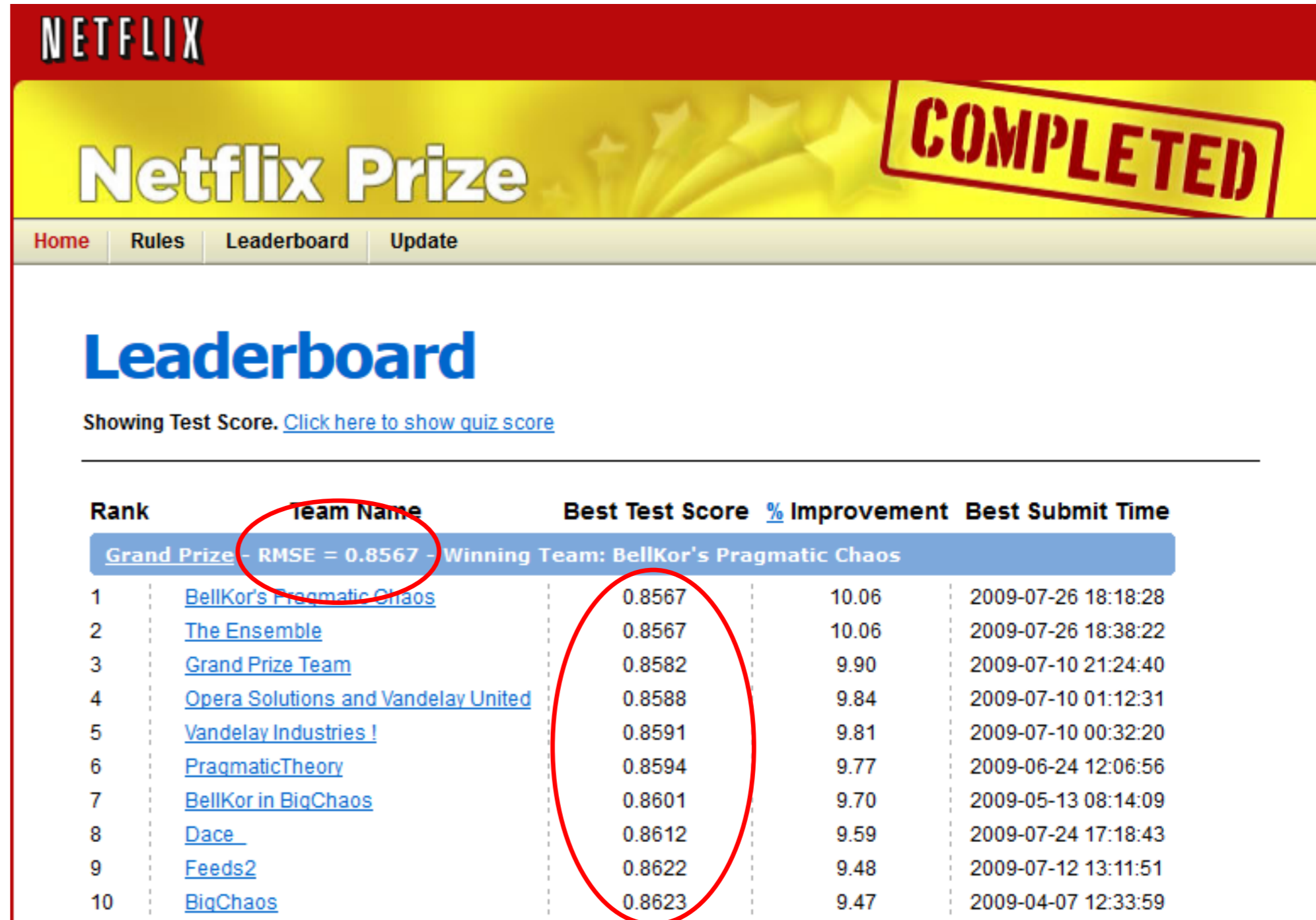
- Competition for all researchers and recommender enthusiasts
- Goal: find the best recommendation algorithm for Netflix
- Challenge: improve Netflix's own algorithm for predicting ratings by 10%
- Prize: 1 Million \$
- Winner:
 - Weighted hybrid design based on > 100 algorithms
 - Adaptive switching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



EVALUATING RECOMMENDATIONS

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



The screenshot shows the Netflix Prize Leaderboard. At the top, there is a red banner with the Netflix logo and a yellow banner with 'Netfix Prize' and a 'COMPLETED' stamp. Below the banner is a navigation bar with 'Home', 'Rules', 'Leaderboard', and 'Update'. The main heading is 'Leaderboard' with a sub-heading 'Showing Test Score. [Click here to show quiz score](#)'. The table below lists the top 10 teams with their Rank, Team Name, Best Test Score, % Improvement, and Best Submit Time. The Grand Prize RMSE is 0.8567, and the winning team is BellKor's Pragmatic Chaos.

| 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 BiqChaos | 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 | BiqChaos | 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 \neq high usage

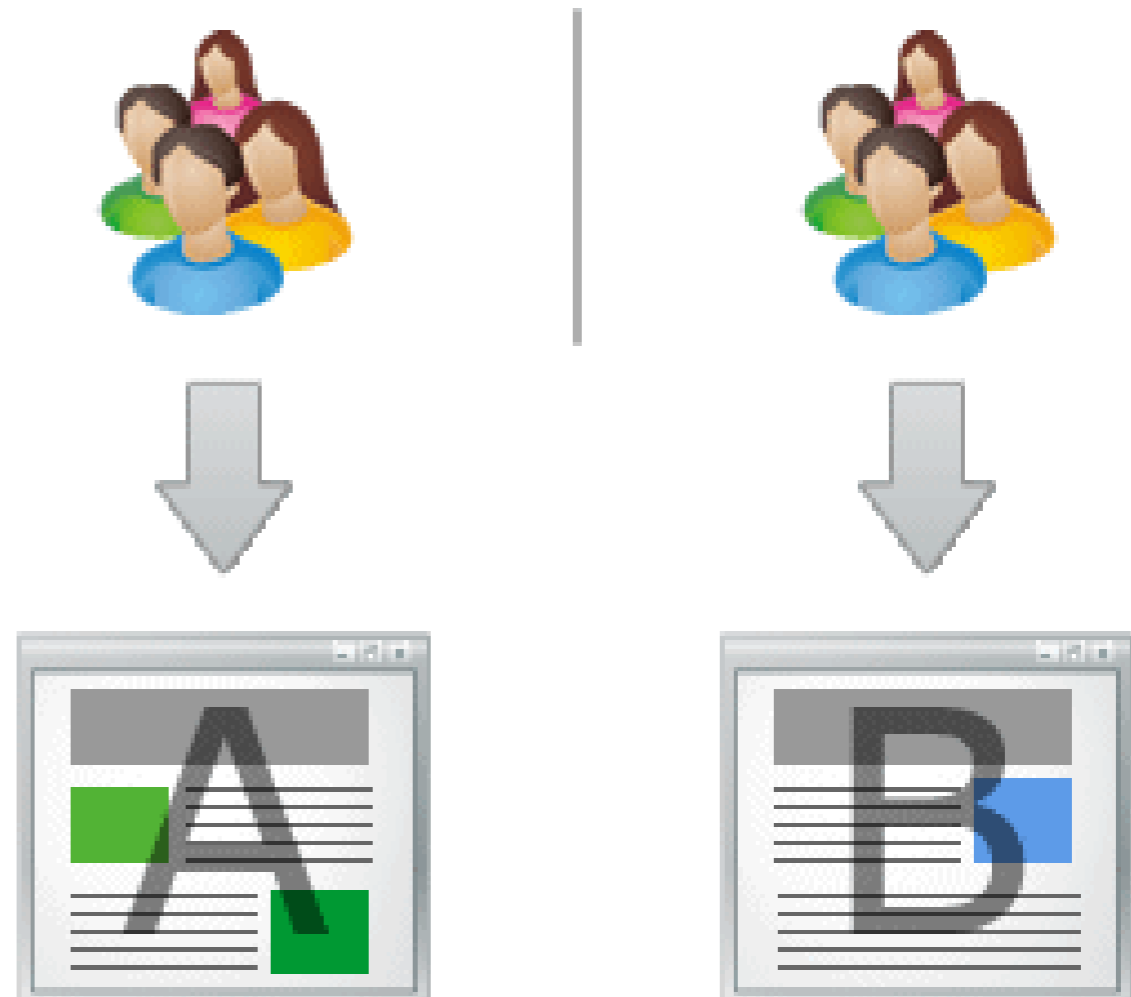
ONLINE TESTING

- 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



WHAT ELSE TO MEASURE?

amazon
Try Prime
Toon's Amazon.com Today's Deals Gift Cards Sell Help

Shop by Department Search All ▼ harry potter Go Hel Yo

Your Instant Video Prime Instant Video Shop Instant Video Video Shorts Your Watchlist

Harry Potter and the Sorcerer's Stone 2001 PG-13 CC

★★★★★ 2,360 IMDb 7.5/10

Based on the wildly popular J.K. Rowling's book about a young boy who on his eleventh birthday discovers, he is the orphaned boy of two powerful wizards and has unique magical powers.

Starring: Richard Harris, Maggie Smith
Runtime: 2 hours, 33 minutes
Available to watch on supported devices.

Rent or Buy

Rent HD \$3.99

Buy HD \$12.99

Redeem a gift card

Add to V

Share

By placing your order, you agree to our Terms of Use. Sold by Amazon Digital Services,

No surprise, too obvious

Sometimes even recommendations for different versions of the same book/item.

E.g. hardcover, paperback edition, collection box, ...

Customers Who Bought This Item Also Bought



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



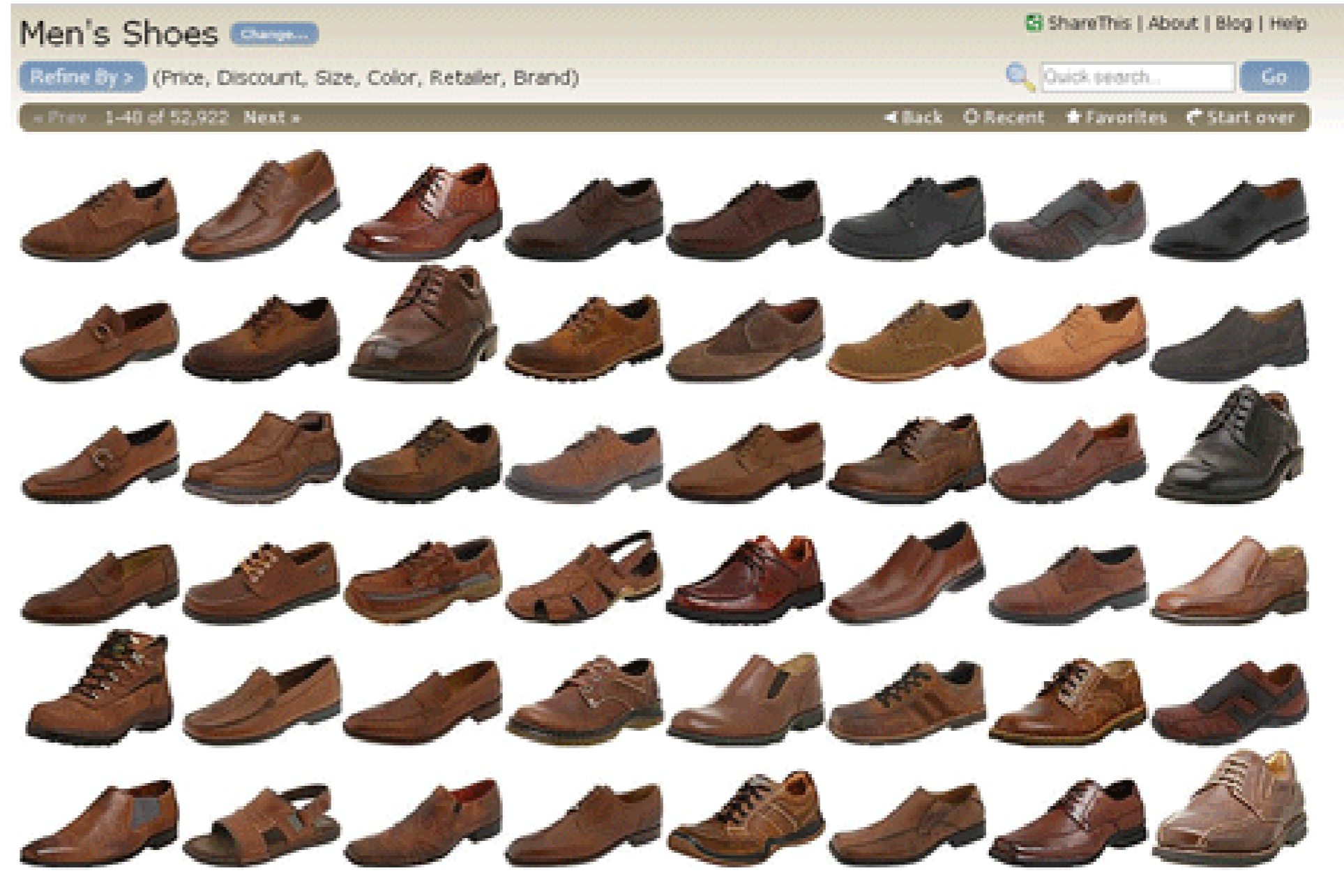
MEASURING SERENDIPITY



- 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



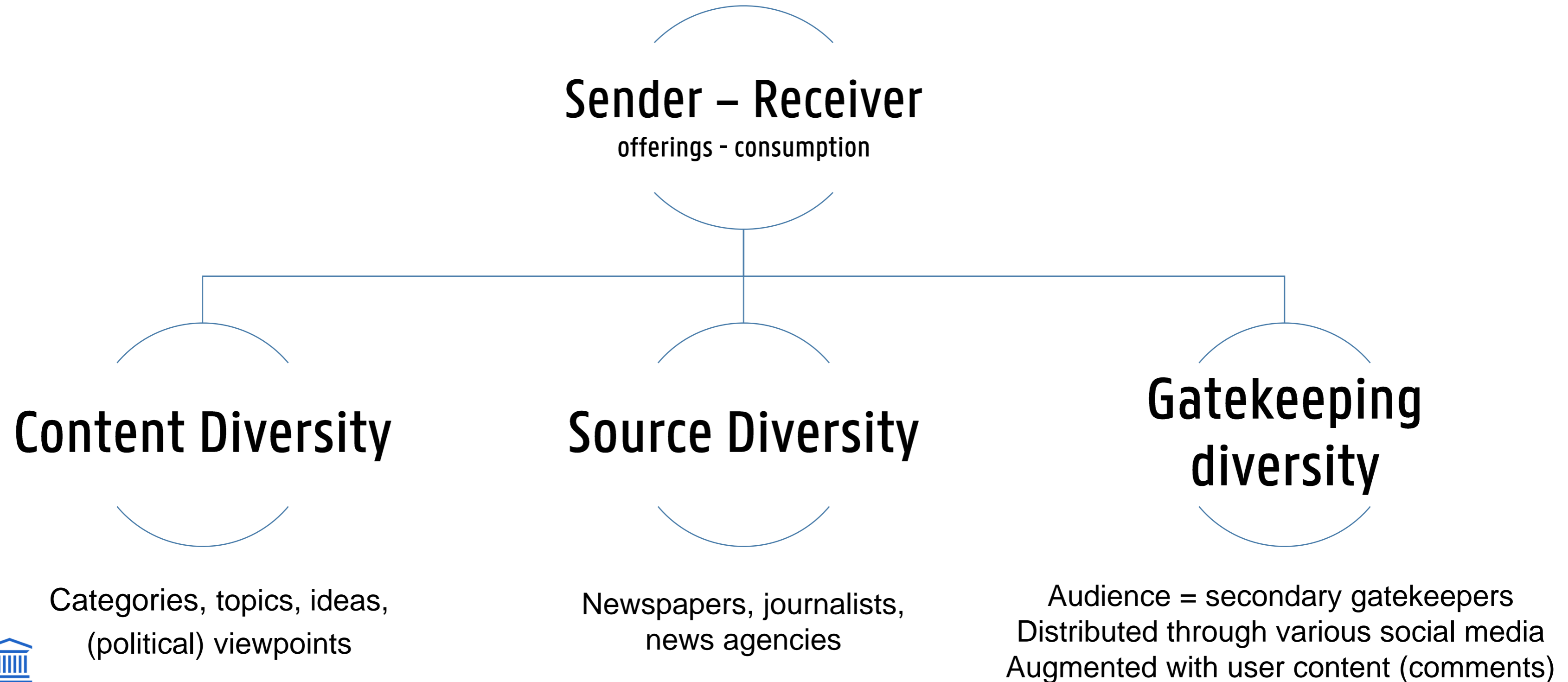
- Measuring
 - With similarity metric (Diversity is the opposite of similarity)

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



CHALLENGE OF NEWS DIVERSITY



GROUP RECOMMENDATIONS

WHY DO WE NEED GROUP RECOMMENDATIONS?



Consumption and selection of content in group

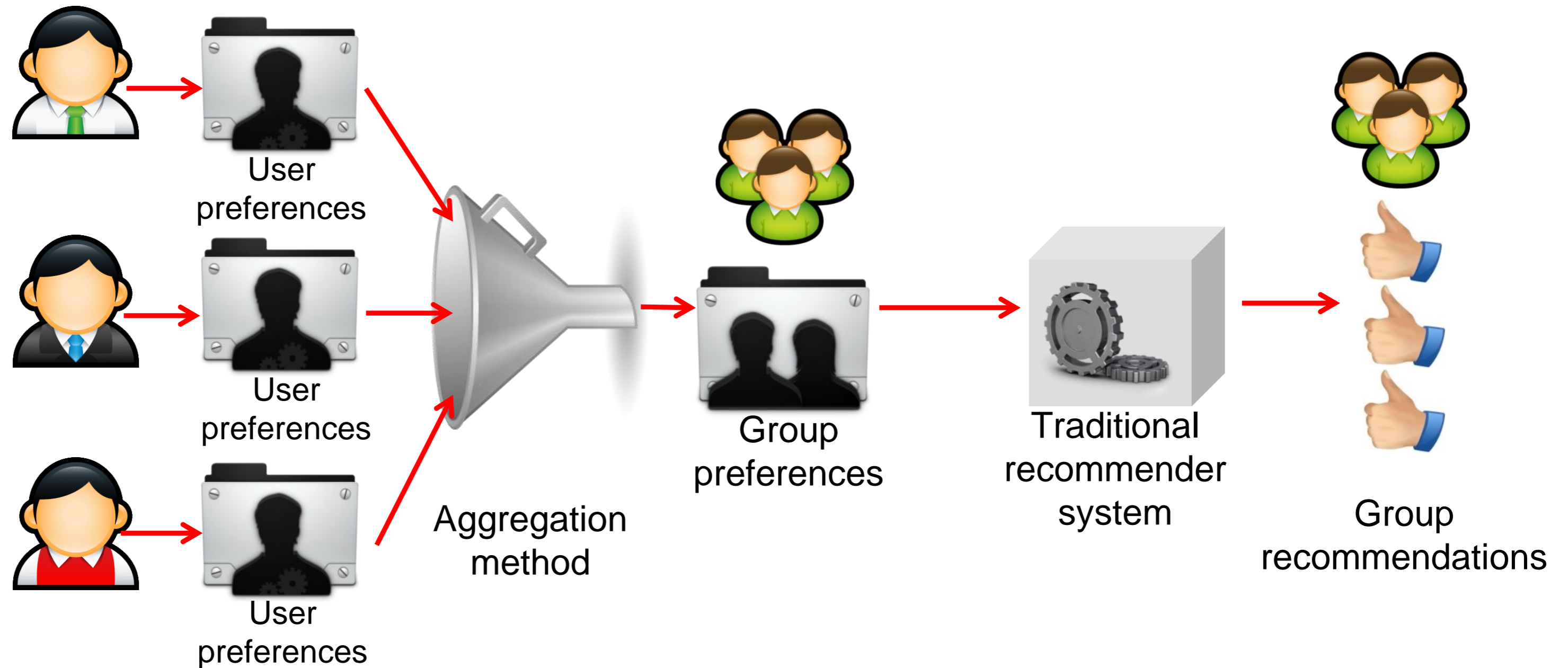
CHALLENGES FOR GROUP RECOMMENDATIONS



Conflicting interests

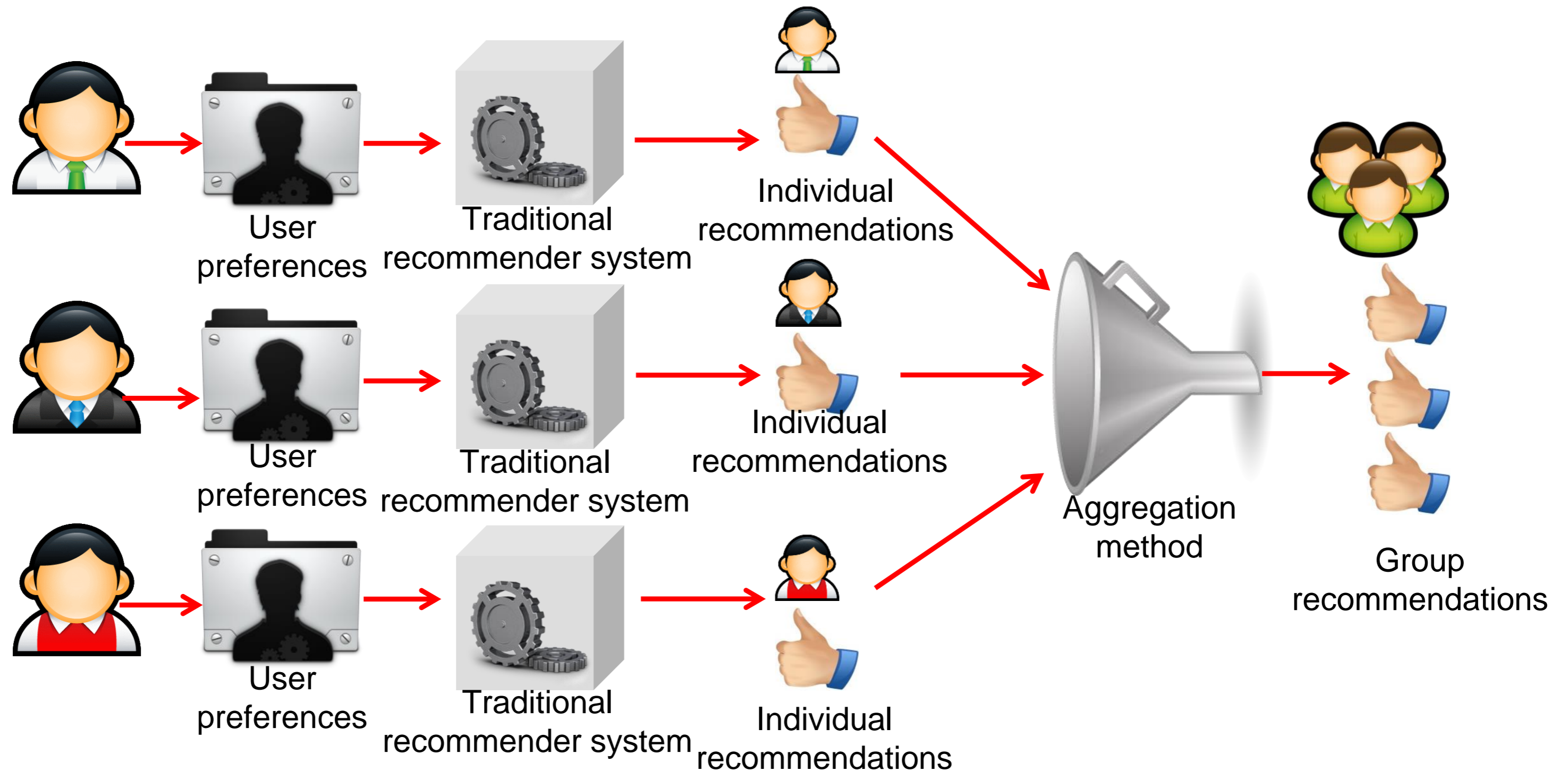
HOW TO GENERATE GROUP RECOMMENDATIONS?

Strategy 1: aggregating user preferences



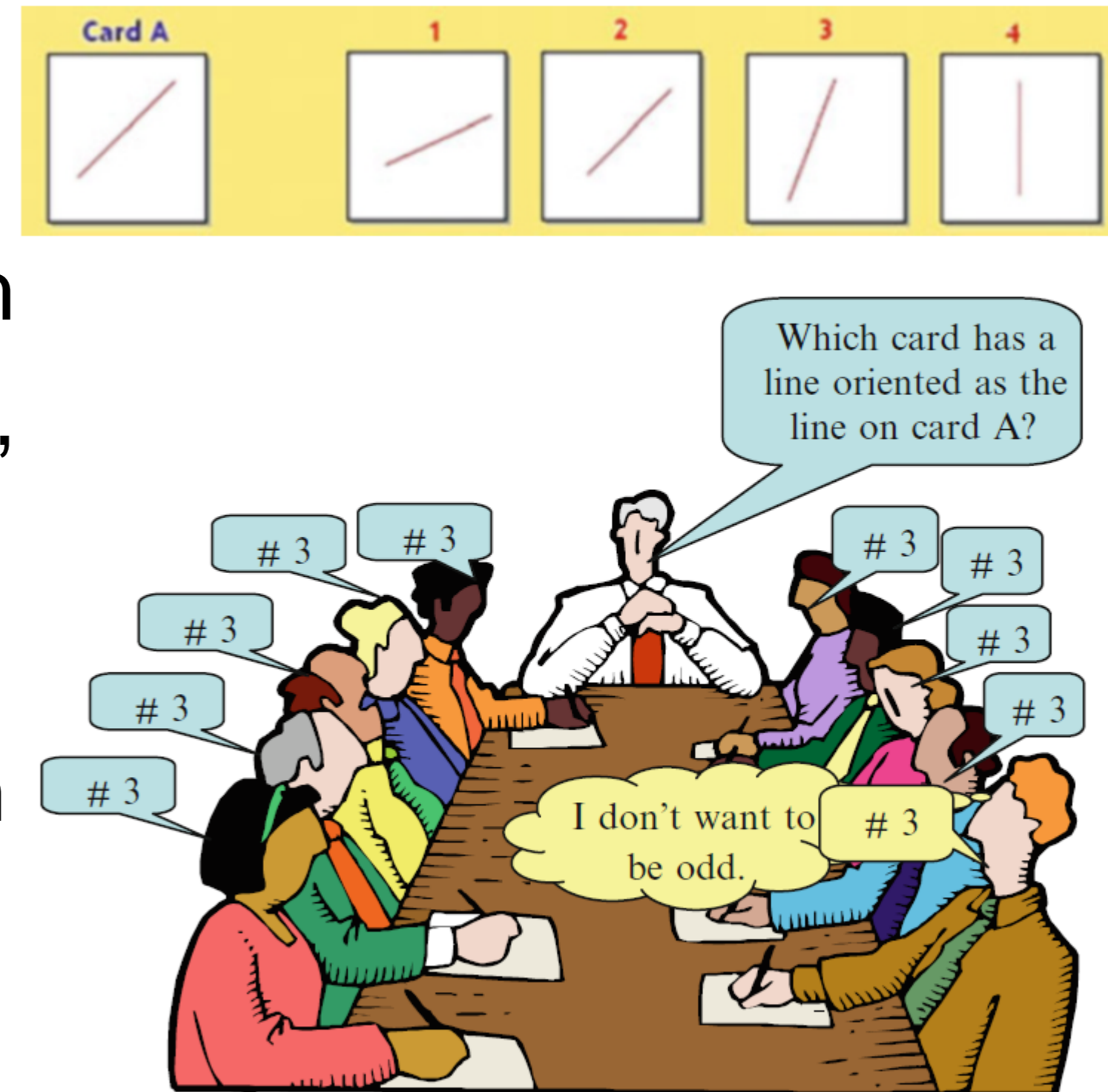
HOW TO GENERATE GROUP RECOMMENDATIONS?

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

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

Toon De Pessemer
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](http://www.ugent.be)