Psychology-informed Recommender Systems

RECOMMENDER SYSTEMS AND SOCIAL COMPUTING LAB

INSTITUTE OF COMPUTATIONAL PERCEPTION

MULTIMEDIA MINING AND SEARCH GROUP

Elisabeth Lex

Graz University of Technology, Austria

elisabeth.lex@tugraz.at

https://elisabethlex.info

@elisab79

Markus Schedl

Johannes Kepler University Linz, Austria

Linz Institute of Technology, Austria

markus.schedl@jku.at

www.mschedl.eu

@m_schedl

JOHANNES KEPLER
UNIVERSITY LINZ
Altenberger Straße 69
4040 Linz, Austria
jku.at

About Today's Presenters

Elisabeth Lex

Associate Professor
Graz University of Technology, Austria
Institute of Interactive Systems and Data Science

Head of Recommender Systems and Social Computing Lab

Contact: elisabeth.lex@tugraz.at | https://elisabethlex.info | @elisab79

Markus Schedl

Full Professor
Johannes Kepler University Linz, Austria
Institute of Computational Perception, Head of Multimedia Mining and Search Group
Linz Institute of Technology, AI Lab, Head of Human-centered AI Group

Contact: markus.schedl@jku.at | www.mschedl.eu | @m_schedl

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Tutorial is Based on a Recent Article

Lex, E., Kowald, D., Seitlinger, P., Tran, T.N.T., Felfernig, A., and Schedl, M.
Psychology-informed Recommender Systems, Foundations and Trends in Information Retrieval, 15(2):134-242, 2021. http://dx.doi.org/10.1561/1500000090

Preprint available from: https://bit.ly/37u0o31

Agenda

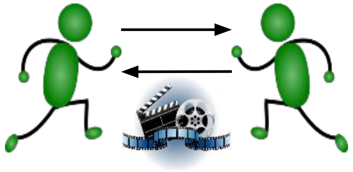
- Part I: Introduction and Motivation (EL+MS)
- Part II: Overview of Types of Psychology-informed Recommender Systems (PIRSs)
  - Cognition-inspired Recommender Systems (EL)
  - Personality-aware Recommender Systems (MS)
  - Affect-aware Recommender Systems (MS)
- Part III: User-centric Evaluation (EL)
- Part IV: Grand Challenges (EL)

Part I: Introduction and Motivation

Main Flavors of Recommender Systems

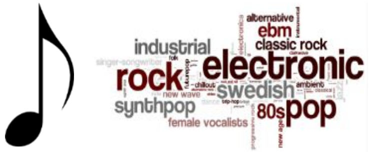
Collaborative filtering:

Recommend to target user items that other *similar users* liked in the past



Content-based filtering:

Recommend to target user *content similar* to what he or she liked in the past



Context-aware RS:

Recommend to target user items that he, she, or other users liked in a given *context or situation*

Hybrid RS: Any *combination* of the above

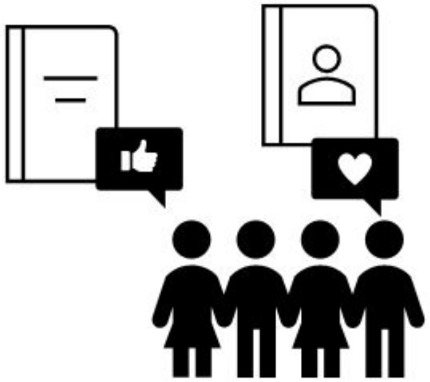
PIRSs

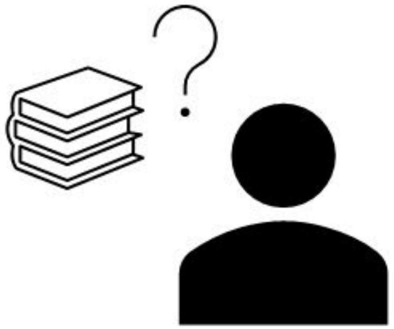
JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Evolution of Decision Making/Psychology and CS/RS Research

- RecSys motivated by observations that humans base their decisions on recommendations from other people
- Early RecSys aimed to mimic that behavior and were based on findings from psychology
  - Emotion & attention
  - User satisfaction / mood
  - Decision making
  - …
- Now: vast amounts of behavioral data available
  - Combine data-driven approaches with psychological models to improve the recommendation process



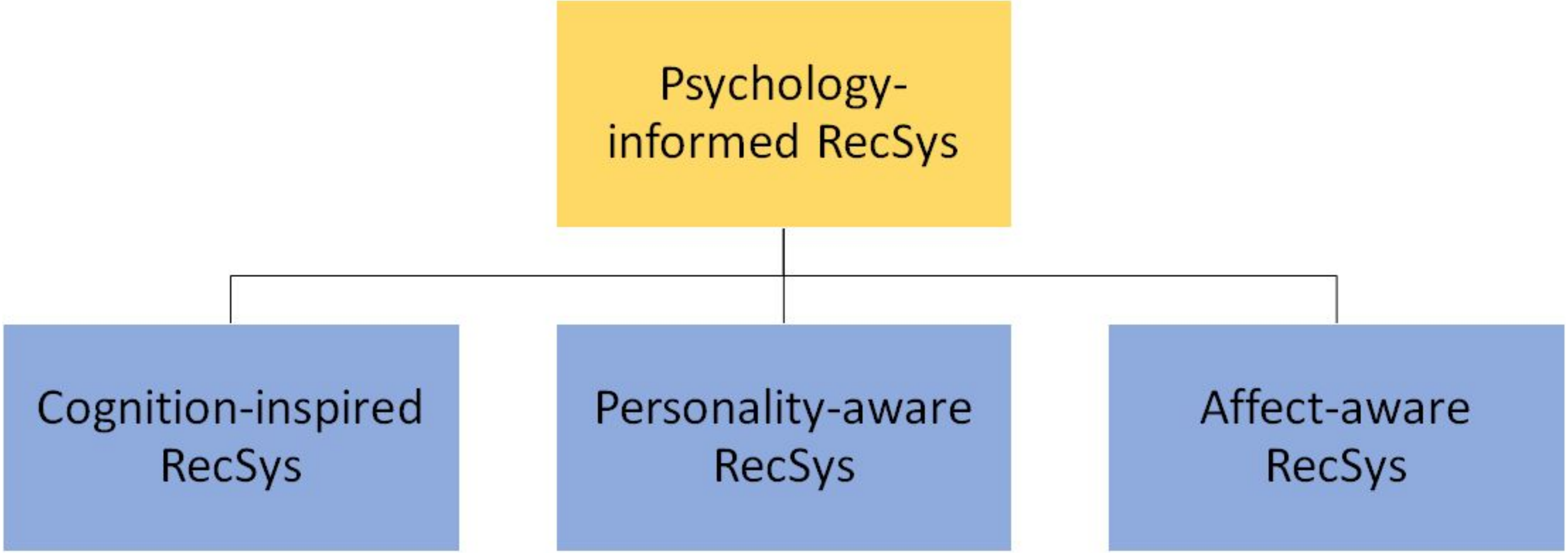


JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Part II: Taxonomy of Psychology-informed RecSys

Taxonomy of PIRS



JOHANNES KEPLER UNIVERSITY LINZ

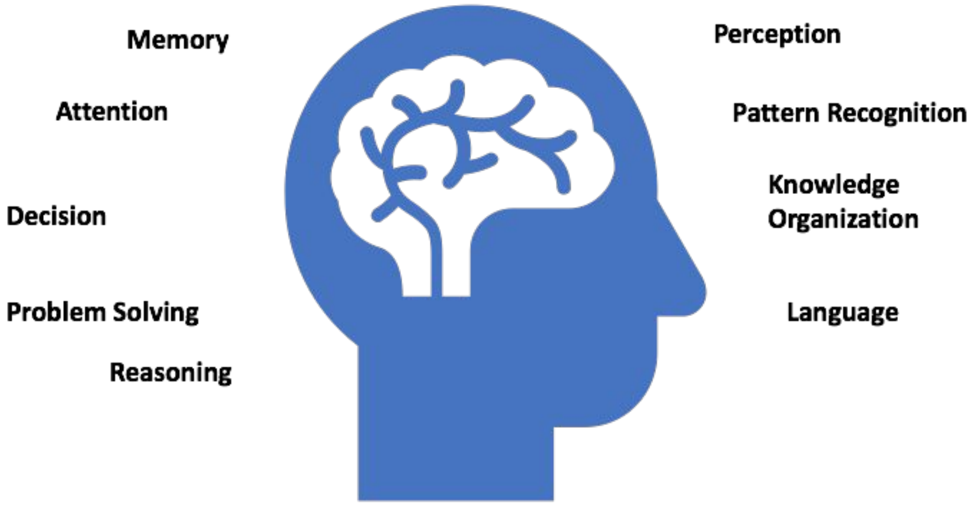
Cognition-inspired Recommender Systems

Cognition-inspired Recommender Systems

- Introduction
- Cognition-inspired models for recommender systems
  - **Stereotypes**
  - **Memory**
  - Case-based Reasoning
  - **Attention**
  - Competence

Cognition-inspired Recommender Systems

- Cognition-inspired recommender systems incorporate **models and theories of cognition**
  - to model user behavior and design recommender systems
  - to improve existing systems
- Cognition:
  - Accumulation of knowledge humans gain from learning and experience
  - Capability of processing information based on perception
  - Studied in **cognitive science**, psychology, sociology, computer science, neurology,...



The Link between Cognitive Science & RecSys

- Cognitive scientists aim to understand how the mind works
  - **describe and predict** people's behavior, and **explain** it
  - Ex.: forgetting a name - what cognitive process is responsible? Attention, memory?
- Approach: cognitive-computational modeling
  - experiments & behavioral data
  - statistical/probabilistic models from mathematical psychology
    - Ex: human mental processes: decision-making, memory, attention, perception,...
  - **Cognitive-computational metaphor:** simulate parts of human mind via computable models, complemented with data-driven approaches
    - test theories, interpret digital trails as manifestations of cognitive processes

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Cognition-inspired Models for Recommender Systems

Stereotypes

Human Memory Models

Case-based Reasoning

Competence Models

Attention

Stereotypes

- Collection of **frequently occurring characteristics of users**
  - “clusters of characteristics”
- Help reduce complexity via simplification & categorization [Hamilton, 1979]
  - Simplification: what characteristics of a person are attended to and remembered.
- Basis for early recommender systems, e.g., **Grundy System** [Rich, 1979]
  - Implemented for book recommendations to people that have been organized in categories according to stereotypes
  - Grundy acted like a librarian

Example: Grundy System

[Rich, E. 1979]

- 2 types of information:
  - Stereotypes: collections of traits
  - Collection of triggers: events that signal suitability of particular stereotypes

Advantage of stereotypes:

- simplistic, transparent
- often complemented with other RecSys approaches

NO-TV-TRIG (Besides asking for characteristic words, the other thing Grundy can do to find out about users is to ask them about TV. This trigger is activated if the user says he does not watch TV.)

FACET	*VALUE*	
Stereotype	NON-TV-PERSON	(this stereotype suggests that a person is likely to be educated and serious)
Rating	800	(maybe this person is not really a non-TV person. Maybe he just can't afford to buy one.)

SCI-ED-TRIG (This trigger is associated with the SCIENTIST stereotype and will be activated whenever the SCIENTIST stereotype is activated.)

FACET	*VALUE*
Stereotype	EDUCATED-PERSON
Rating	900
Reasons	SCIENTIST

Figure 2.1: Sample triggers by Rich [2]

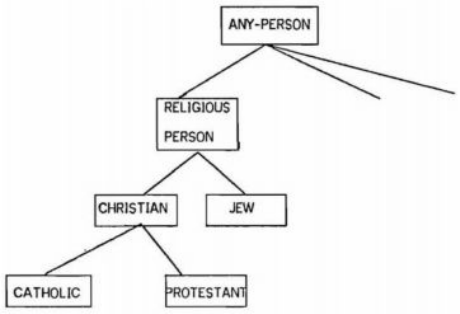


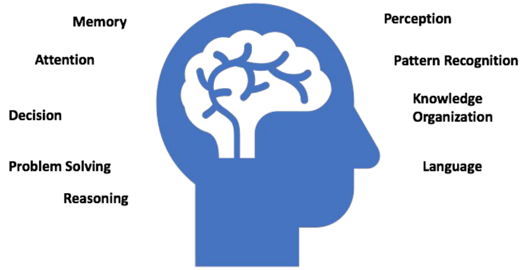
Figure 2.2: Stereotype hierarchy as developed in GRUNDY by Rich [2]

Source: Al-Rossais, N (2021). Intelligent, Item-Based Stereotype Recommender System. PhD thesis

JOHANNES KEPLER UNIVERSITY LINZ

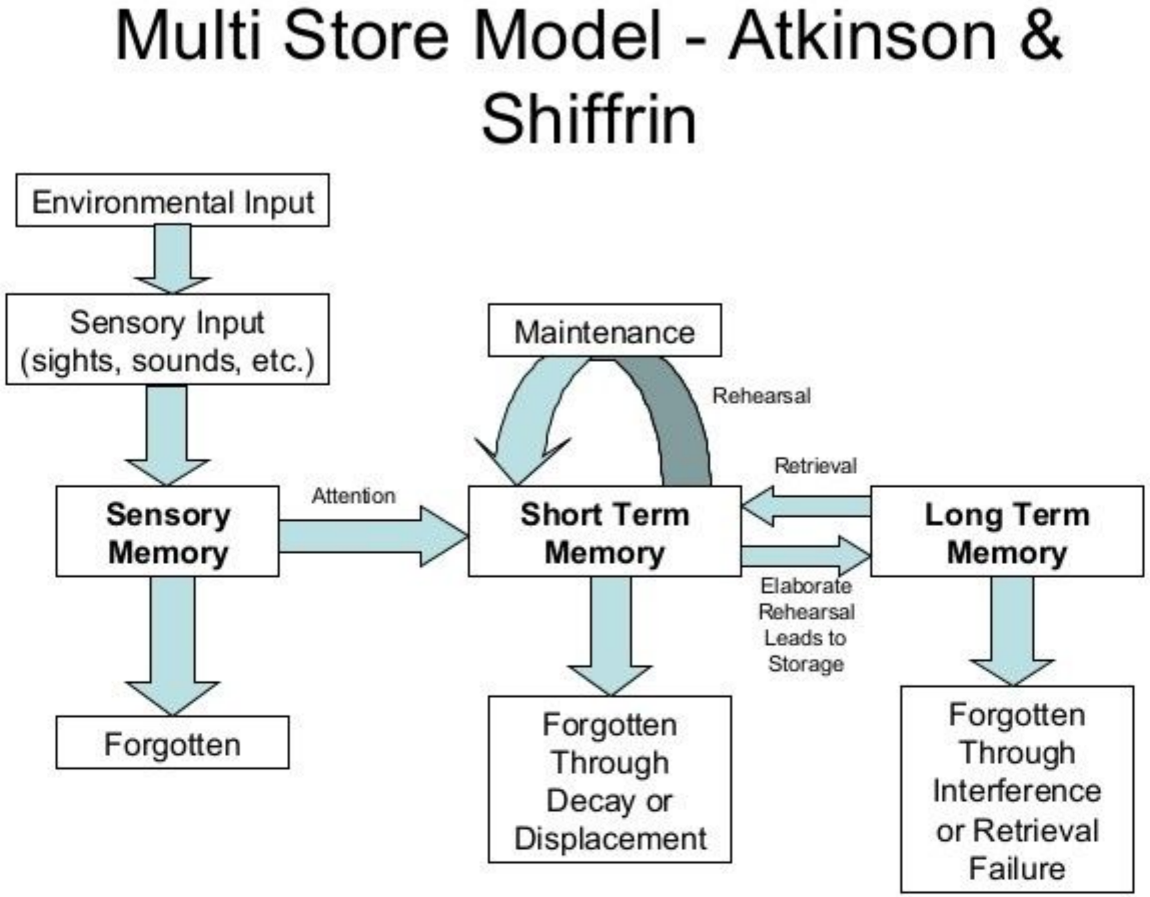
TU Graz

Memory



[Atkinson & Shiffrin 1968]

- Fundamental process of human cognition
- Supports goal-directed interactions with physical & social environment
- Central role in problem-solving, attention, decision-making, perception
- Consists of memory structures
  - sensory, short-term, long-term
- Many models of memory - e.g. Atkinson and Shiffrin model



The Atkinson and Shiffrin Model

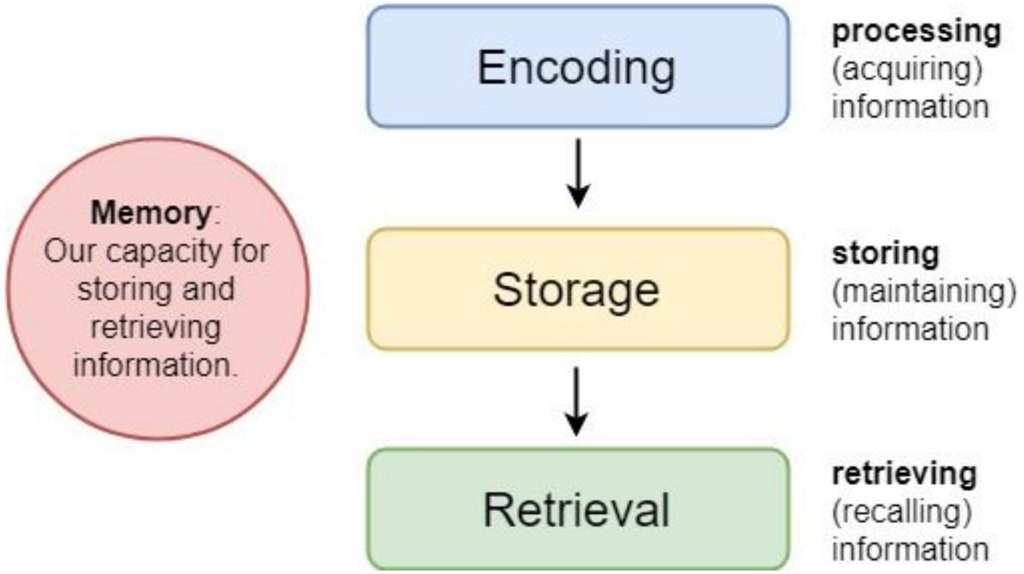
Source: https://www.wikiwand.com/en/Information_processing_theory

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Key Functions of Cognitive Process Memory

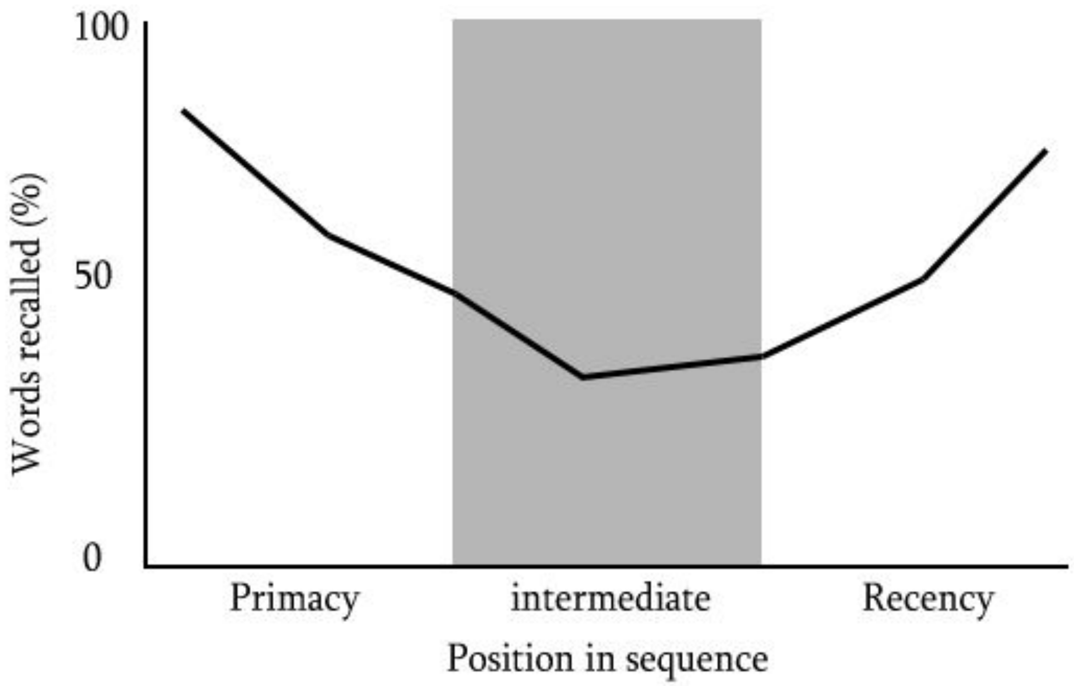
- Encoding: records information, so it becomes usable by memory system
  - bound to temporal & spatial context information: enables later context-guided search of memory
- Storage: encoded information retained and held over a period of time, so it can be used later
- Retrieval: stored information can be recovered from memory when the situation demands



Recalling Information - Memory Effects

- Serial Positioning Effect
  - we remember first and last items in lists much better than the ones in the middle!

→ Serial positioning effect detected by Ebbinghaus in the 1880ies!



Source: https://commons.wikimedia.org/wiki/File:Serial_position.png

JOHANNES KEPLER UNIVERSITY LINZ
TU Graz

Ebbinghaus Curve

[Ebbinghaus, 1885]

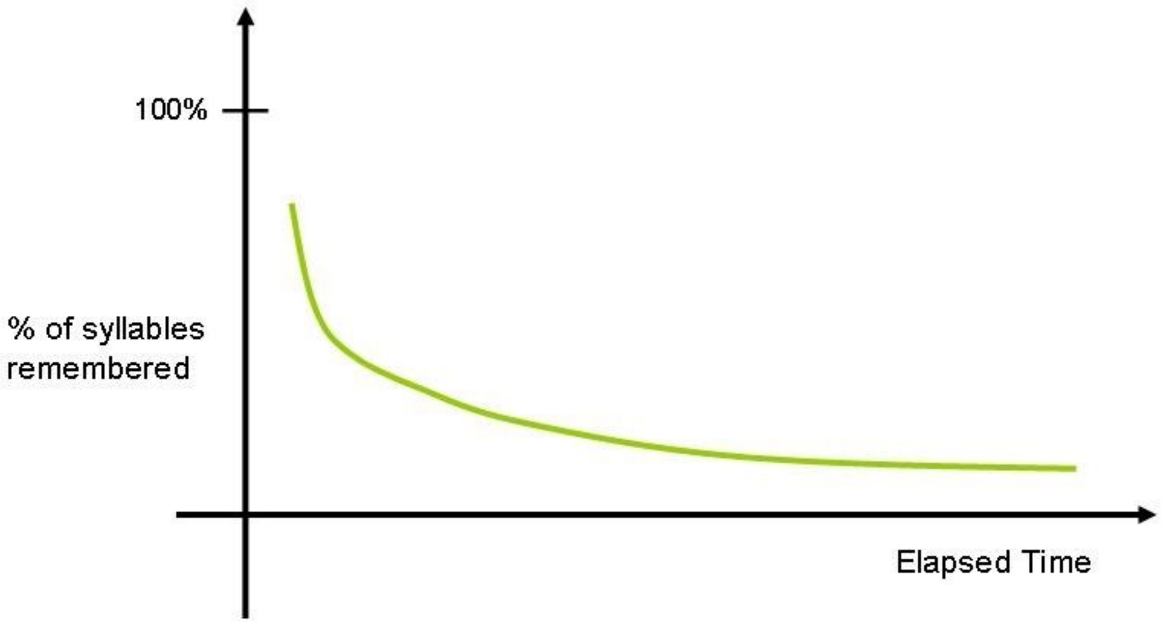
- Decline of memory retention in time
- Experiment by Ebbinghaus
  - Memorized nonsense syllables
  - Repeatedly tested his memorization
  - Aim: quantify rate of forgetting

time elapsed

memory retention

$$R = e^{-\frac{t}{s}}$$

relative strength of memory:
e.g. exposure event to item



Source: https://commons.wikimedia.org/wiki/File:Ebbinghaus_Forgetting_Curve.jpg

JOHANNES KEPLER UNIVERSITY LINZ
TU Graz

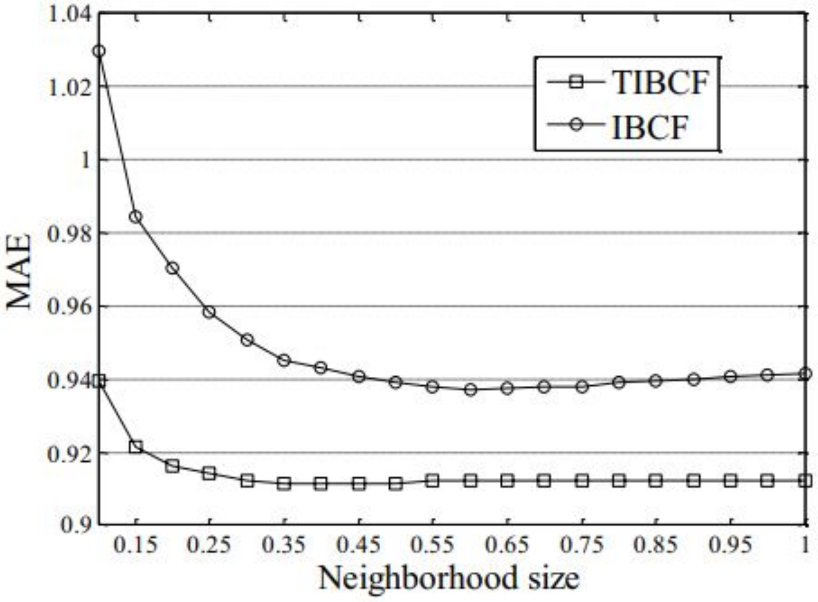
Example: Time-enhanced Collaborative Filtering Algorithm

- Problem: user interests change and in CF - time information about ratings ignored
- Idea: model user interest changes as a form of information forgetting
    - exploit Ebbinghaus curve
- Approach: time-based exponential decay weight based on the produced time of ratings
- Use weight for similarity computing and rating prediction

$$w(i,j) = e^{\frac{-RK(R_i,i,j)\times\log(2)}{\lambda / \log(|R_i|)}}$$

$$sim(i,j) = \frac{\sum_{u_c \in U_{ij}} w(c,i)\cdot(r_{ci}-\overline{r}_i)\cdot w(c,j)\cdot(r_{cj}-\overline{r}_j)}{\sqrt{\sum_{u_c \in U_{ij}} w(c,i)\cdot(r_{ci}-\overline{r}_i)^2}\sqrt{\sum_{u_c \in U_{ij}} w(c,j)\cdot(r_{cj}-\overline{r}_j)^2}}$$

$$\hat{r}_{ci} = \frac{\sum_{t_j \in T_{ci}} sim(i,j)\cdot w(c,j)\cdot r_{cj}}{\sum_{t_j \in T_{ci}} |sim(i,j)\cdot w(c,j)|}$$



Cognitive Architectures

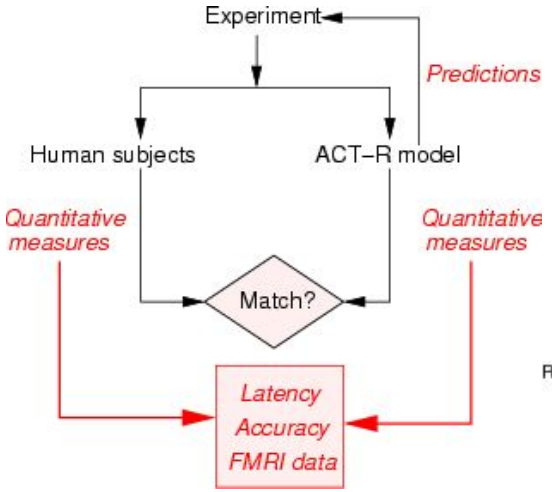
- Fundamentals of human cognition often organized in **cognitive architectures → aim is to provide a unified theory of the human mind**
- Cognitive architectures make theoretical assumptions about mechanisms underlying human cognition
  - Based on psychological findings
- Consist of modules that access and alter memories and representations
- Typically, programmatic implementations available

Cognitive Architectures

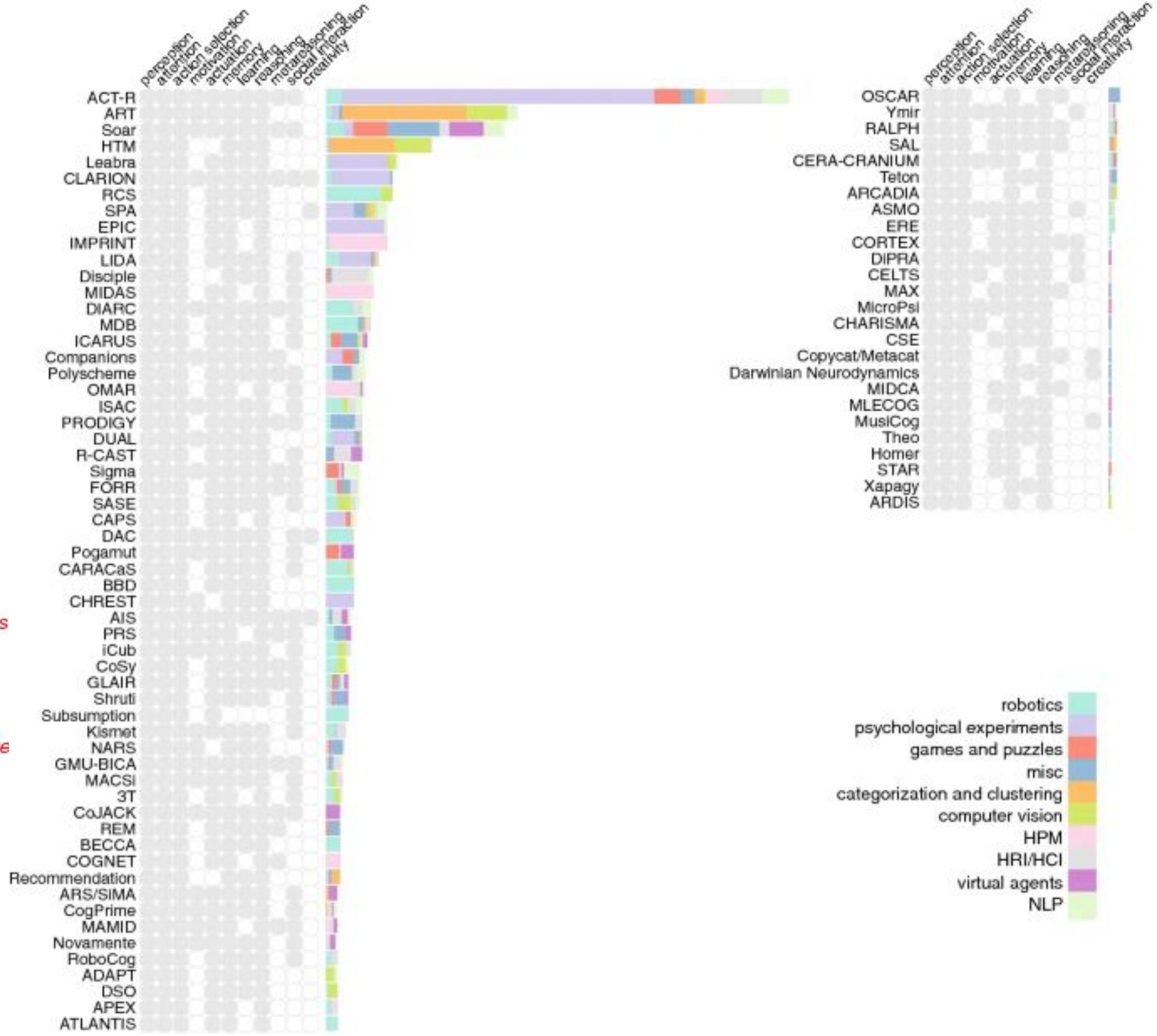
- Adaptive Control of Thought (ACT-R) highly prominent architecture (also in RecSys!)

Advantage of ACT-R:

- enables to collect quantitative measures that can be directly compared with quantitative measures obtained from human participants



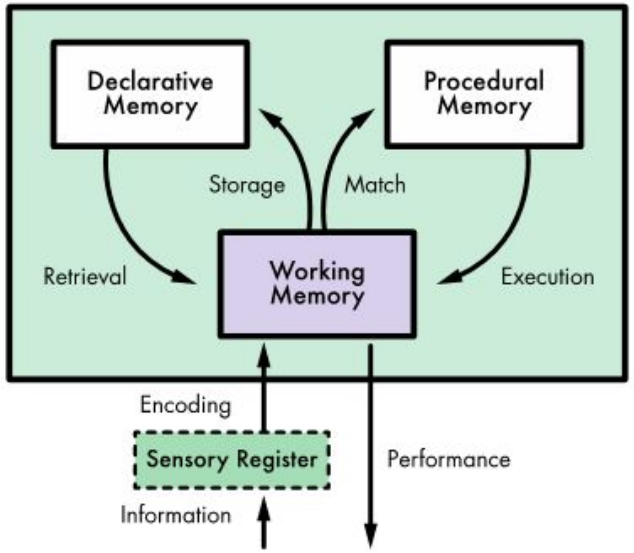
http://act-r.psy.cmu.edu/



Cognitive Architecture ACT-R

[Anderson et al., 2004]

- Describes activation processes in human memory in the declarative memory
  - Incorporates **base-level activation** and **associative activation**
  - Denotes relevance of memory unit in current context
- Information used **frequently** and **recently** easier extracted from human memory
  - Modeled in Base-Level-Learning equation Bi



Activation of memory unit

$$A_i = B_i + \sum_j (W_j \cdot S_{j,i})$$

base-level activation of *i*
(general usefulness)

associative activation of *i*
(relevance to **context cues** *j*)

$$B_i = ln(\sum_{j=1}^{n} t_j^{-d})$$

integrates **past usage frequency and recency of i**

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Example: Music Preferences and ACT-R

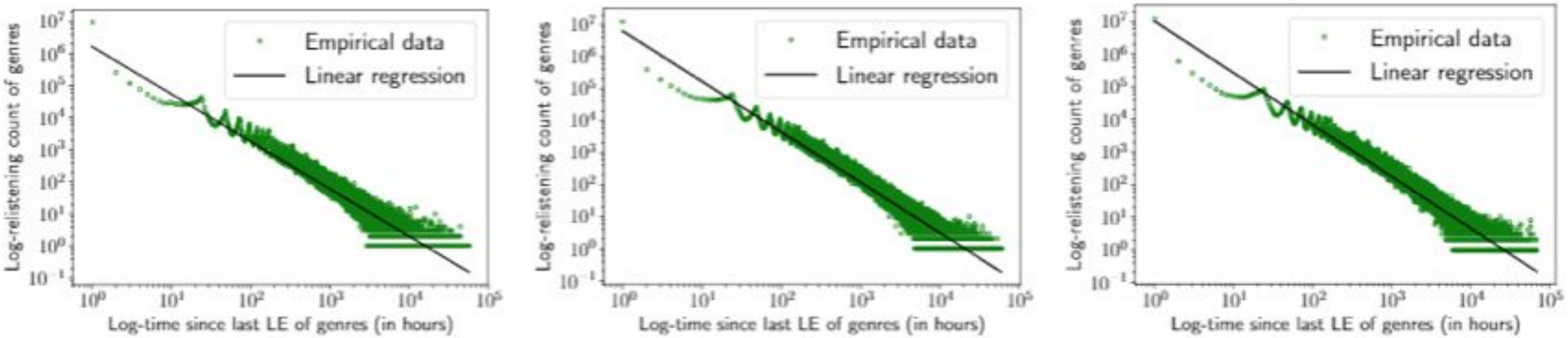
[Lex et al., 2020]

- Motivation: **Popularity bias** for consumers of low mainstream music
- Idea: Psychology-informed model of music consumption behavior
- Aim: **Predict music genre preferences for low, medium & high mainstream consumers**
- Approach
    - 1.1 billion listening events (LEs) from LFM-1b [Schedl, 2016]
        - Each LE contains a user identifier, artist, album, track name, and timestamp
        - Plus: mainstreaminess score:
            - Overlap between a user's personal listening history and the aggregated listening history of all Last.fm users in the dataset.
    - Created 3 equally sized groups based on mainstreaminess: low, medium, high mainstream

Temporal Dynamics of Music Consumption

[Lex et al., 2020]

- Re-listening count of genres over time plotted on log-log scale



(a) User group: LowMS
Linear regression: α = -1.480

(b) User group: MedMS
Linear regression: α = -1.574

(c) User group: HighMS
Linear regression: α = -1.587

--> the shorter the time since the last listening event of a genre the higher its relistening count!

$$B_i = ln\left(\sum_{j=1}^{n} t_j^{-d}\right)$$

Approach - BLL_U

[Lex et al., 2020]

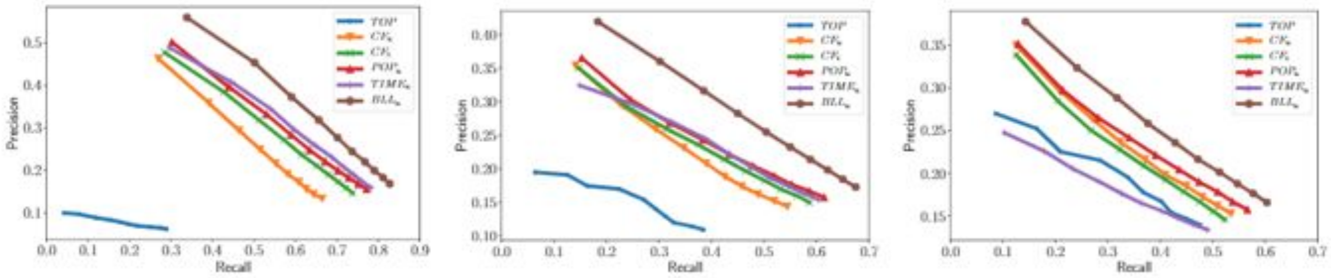
1. Compute base-level activation of a genre for a user
2. Normalize using soft max function
3. Predict top-k genres with highest activation

$$B_{u,g} = ln\left(\sum_{j=1}^{n} t_{u,g,j}^{-d}\right)$$

$$\widetilde{G_u^k} = \underset{g \in G_u}{\overset{k}{\arg\max}}(B'_{u,g})$$

$$B'_{u,g} = \frac{\exp(B_{u,g})}{\sum_{g' \in G_u} \exp(B_{u,g'})}$$

User group	Evaluation metric	TOP	CF_u	CF_i	POP_u	$TIME_u$	BLL_u
LowMS	F1@5	.108	.311	.341	.356	.368	.397***
	MRR@10	.101	.389	.425	.443	.445	.492***
	MAP@10	.112	.461	.505	.533	.550	.601***
	nDCG@10	.180	.541	.590	.618	.625	.679***
MedMS	F1@5	.196	.271	.284	.292	.293	.338***
	MRR@10	.146	.248	.264	.274	.272	.320***
	MAP@10	.187	.319	.336	.351	.365	.419***
	nDCG@10	.277	.419	.441	.460	.452	.523***
HighMS	F1@5	.247	.273	.266	.282	.228	.304***
	MRR@10	.188	.232	.229	.242	.201	.266***
	MAP@10	.246	.304	.298	.314	.267	.348***
	nDCG@10	.354	.413	.402	.429	.357	.462***



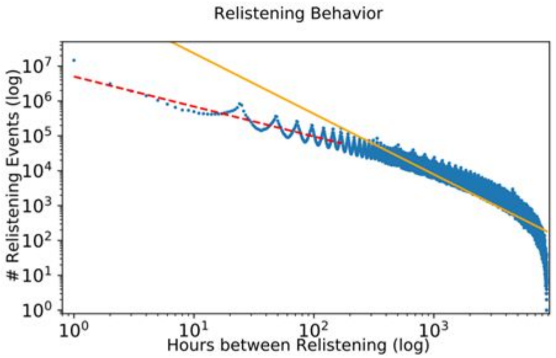
(a) User group: LowMS (b) User group: MedMS (c) User group: HighMS

Other Useful Components of ACT-R's Declarative Memory

- Declarative Memory Components
  - Base-level
    - models recency + frequency of exposure to items
  - Spreading
    - models co-occurrence with other items
  - Partial Matching
    - models similarity between items
  - Valuation
    - models familiarity with items
  - Noise
    - accounts for randomness in behavior

Music Cognition and Memory

[Reiter-Haas et al., 2021]



Algorithm	R-prec	Next-HR
TransProb	.03839	.15907
Partial Matching	.03895	.01320
Noise	.03996	.00289
Valuation(discrete)	.04751	.00533
Valuation(ratio)	.05987	.01042
Valuation(MP)	.08436	.01477
Spreading	.09235	.02117
Base-level(2019)	.09903	.03200
ACT-R(B,V)	.10069	.02416
MostRecent	.10167	.05189
Base-level(default)	.10380	.02451
Base-level(week)	.10489	.02883
ACT-R(S,V)	.11009	.02998
ACT-R(B,S)	.11042	.02972
ACT-R(B,S,V)	.11119	.02961

- Motivation
  - Tendency to **relisten to songs** (Frederick et al., 2019)
  - Repeated exposure **increase recognition and positive attitude** (Peretz et al., 1998)
- Aim: **Predict relistening behavior**
  - Sequential evaluation
  - ~1.7 Mio. LEs from LFM-2b (Melchiorre et al., 2021)
  - Listening sessions (30min)
  - Sliding window = 1 week
  - Predict tracks in session: Next (Hitrate) & Remaining (R-precision)

- Conclusion
  - **Recency & frequency of prior exposure** effective predictor
  - Adding **co-occurrence & familiarity** improves prediction

JOHANNES KEPLER UNIVERSITY LINZ

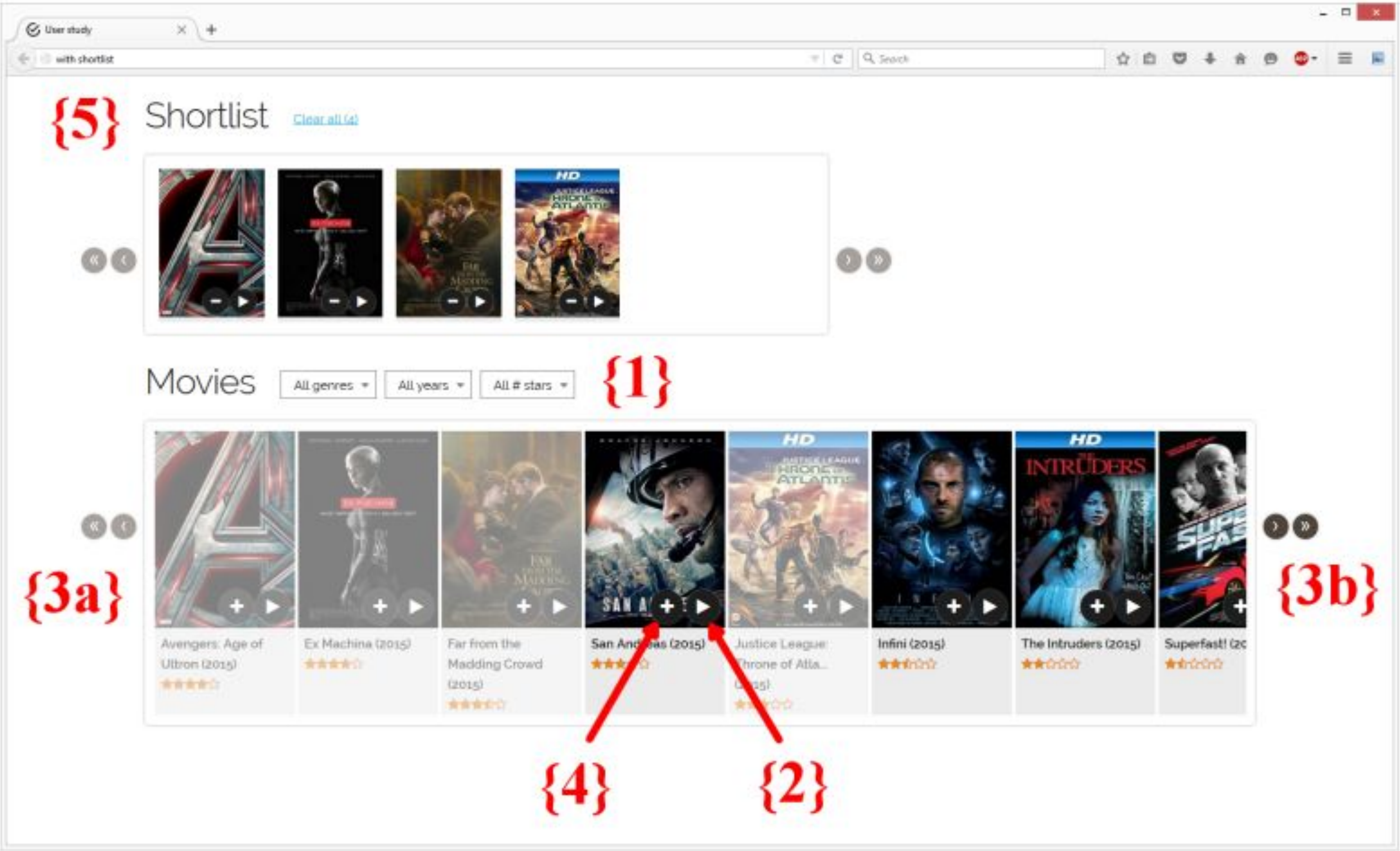
TU Graz

Supporting Human Memory with RecSys

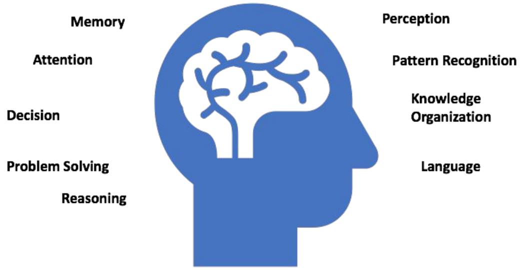
[Schnabel et al., 2016]

- Creating shortlists: [Schnabel et al., 2016] propose to support a user's short-term memory by creating a digital short-term memory in the form of shortlists
    - Contain items user currently considers → implicit feedback & additional training data

→ increased user satisfaction in terms of decision quality, engagement



Case-based Reasoning



[Kolodner 1992]

- Memory-based **problem-solving**
- A RecSys type of its own!
- Idea: **reasoner remembers previous cases** that are similar to the current case and uses them to solve new problems
  - analogous to an expert decision maker: mimic how humans draw on previous learning episodes when solving new problems.
- Technique pioneered by cognitive scientist Janet Kolodner

Some definitions:

Case-based reasoning is […] reasoning by remembering - Leake, 1996

A case-based reasoner solves new problems by adapting solutions that were used to solve old problems - Riesbeck & Schank, 1989

Case-based Reasoning RecSys

- CBR RecSys constitute early examples of psychology-informed recommender systems
  - Use problem solving architecture designed by psychologists
  - Similarity metrics used by CBR systems inspired by works in psychology on basic features of similarity
    - Similarity between two items is determined based on their common and distinctive features (see [Tversky, 1977])
- Requires a knowledge base!
- Advantage: transparent & explainable

CBR cycle according to Ian Wasten

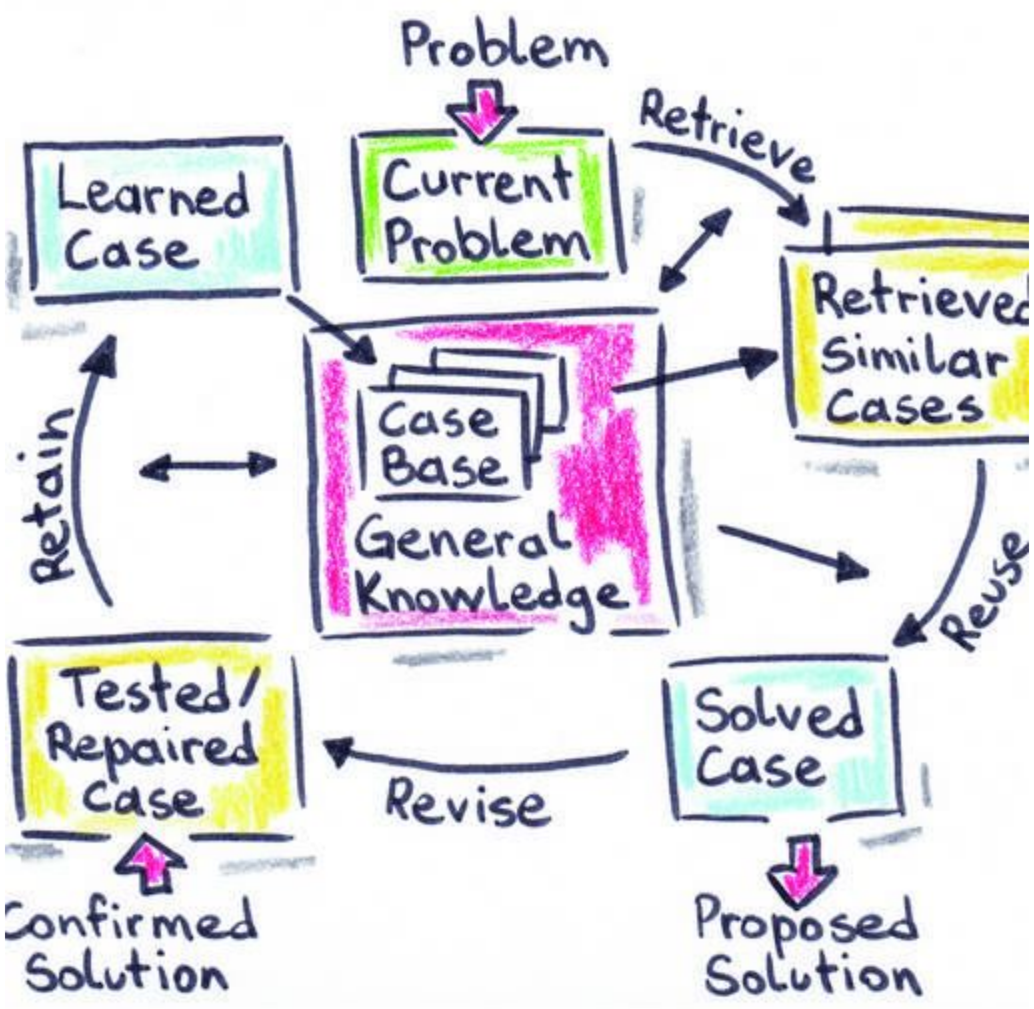
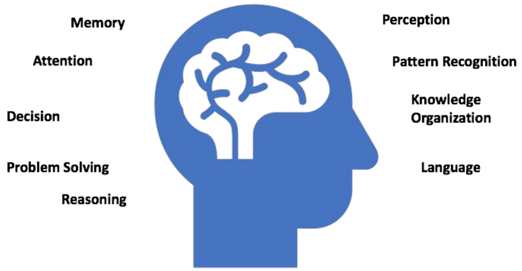


Image source: https://www.ask-flip.com/method/75

Some Examples of Case-based Reasoning RecSys

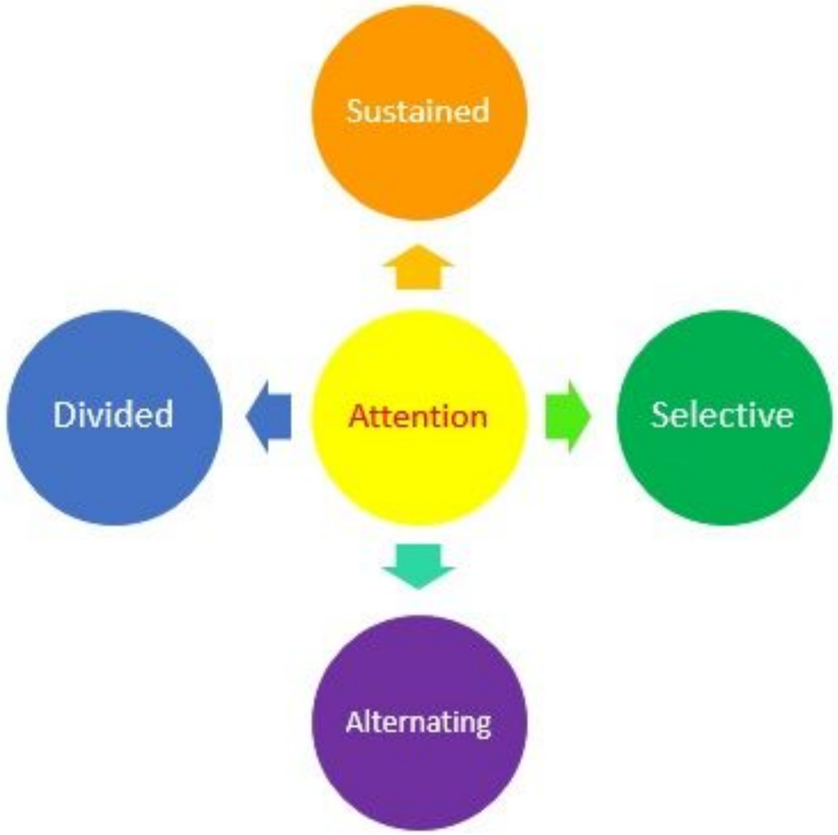
- Wasabi System - CBR to generate recommendations in an e-commerce setting [Burke, 1999] or to produce restaurant recommendations [Burke, 1996]
- CBR for travel recommendations [Ricchi et al., 2001, 2002, 2006]
- CBR for music recommendations [Aguzzoli 2002; Gong 2009], in combination with CF
- CBR to recommend personalized investment portfolios [Musto et al., 2015] to assist financial advisors
- CBR in educational settings - e.g., [Boushbahi et al., 2015] CBR-based recommendation approach to assist learners in finding massive open online courses (MOOCs) that meet their personal interests

Attention



- Mechanism to **selectively process information** in an environment in the face of distraction

- Psychologists describe 4 types
  - Selective: focus on a particular object
  - Divided: focus on several stimuli simultaneously - multitasking
  - Alternating: switch between tasks
  - Sustained: intensive focus on a specific task



JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

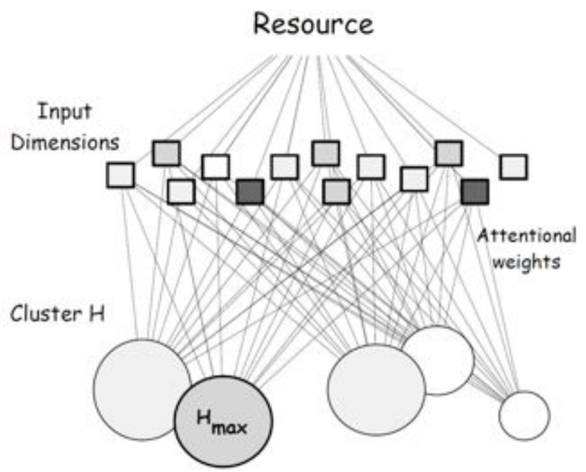
Modeling Attention

- Attention is dynamic → psychologists typically model attention using connectionist models
- Connectionism is a research strand in cognitive science, which uses artificial neural networks to study cognition and to model cognitive processes
  - Aim: model connections and dynamic aspects of cognition like in the brain
  - Networks of interconnected neurons

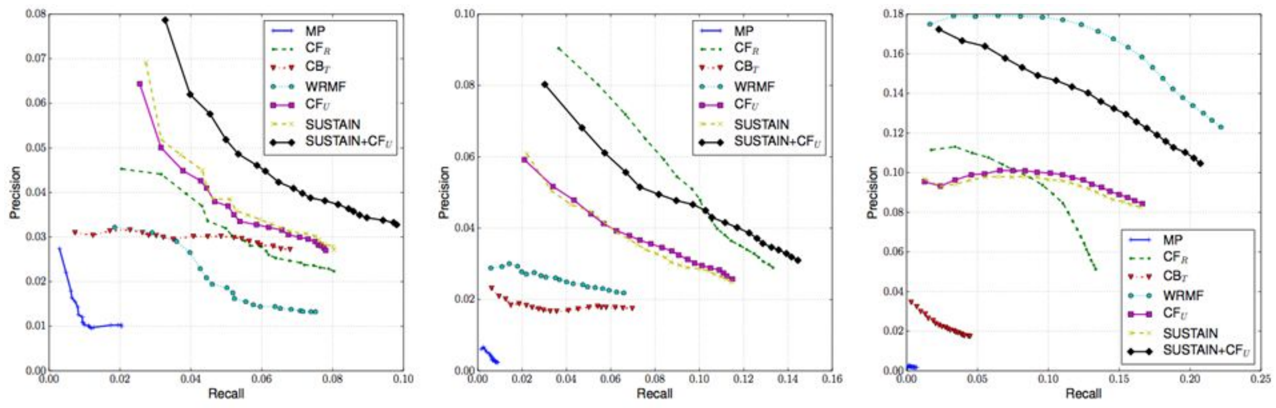
- Example of connectionist model: SUSTAIN [Love et al., 2004]
  - Cognitive model of human category learning
  - Input, hidden and output units interconnected within a multi-layer network

Examples - SUSTAIN

[Kopeinik et al. 2017]



- Idea: model attention dynamics with SUSTAIN to recommend resources that fit user's current attentional focus while interacting with learning resources + improve CF by re-ranking
- Approach
  - Input: topics describing resources
  - Output: decision to take or leave resource
  - Attentional weights of input units and interconnections evolve as network encounters new resources
- Results: SUSTAIN+CF improves prediction



(a) BibSonomy (b) CiteULike (c) Delicious

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Take Away

- Cognitive models of human cognition helps design and improve recommender systems
  - Underlying psychological models contribute to deeper understanding of user behavior.
  - Use RecSys to support / augment human memory

- Attention & shifts in user interests crucial issues, which can be tackled with RecSys
  - Success of deep learning has resulted in many attention-based approaches
  - However: scarce work on underlying psychological mechanisms

    → Potential for future research - also to foster transparency / interpretability

Personality-aware Recommender Systems

Personality-aware Recommender Systems

- Motivation
- Modeling personality (OCEAN five factor model)
- Acquiring personality traits (surveys vs. automatically from digital footprint)
- Personality and item preferences
- Using personality traits for recommendation

Motivation

- Alleviate *cold start* problem for new users, e.g.:
  - Extract personality of users from their user-generated content
  - Match users with items based on (1) items' "personality" or (2) models that correlate personality with item preferences (e.g., genre)

- Tailoring level of *diversity* in recommendation lists, e.g.:
  - Extract personality of users from their user-generated content
  - Use standard CF approach to create candidate recommendation list
  - Re-rank list based on models/studies that correlate personality traits with desired level of diversity in result lists

Modeling Personality (OCEAN/Five Factor Model)

- **O**penness to experience (inventive/curious vs. consistent/cautious)
- **C**onscientiousness (efficient/organized vs. extravagant/careless)
- **E**xtraversion (outgoing/energetic vs. solitary/reserved)
- **A**greeableness (friendly/compassionate vs. critical/rational)
- **N**euroticism (sensitive/nervous vs. resilient/confident)

A person is described on a numeric scale (e.g., between 1 and 7) for each trait.

Resources (measures and scales):
International Personality Item Pool (IPIP): https://ipip.ori.org [Goldberg et al., 2006]

Acquiring Personality Traits

- Either through *questionnaires* or automatically *from user-generated data through ML*
  Questionnaires: more accurate, more labor-intensive/expensive
  Machine learning: less accurate, less expensive, possible to train on small amount of data and apply to large-scale data (e.g., microblogs, Likes, sensor data)

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Acquiring Personality Traits: Questionnaires

- Either through *questionnaires* or automatically *from user-generated data through ML*
  Questionnaires: more accurate, more labor-intensive/expensive
  Machine learning: less accurate, less expensive, possible to train on small amount of data and apply to large-scale data (e.g., microblogs, Likes, sensor data)

- Common instruments/questionnaires:
  Ten Item Personality Inventory (TIPI):
  Questionnaire: https://gosling.psy.utexas.edu/wp-content/uploads/2014/09/tipi.pdf
  Questions like: "I see myself as disorganized, careless." rated from strongly disagree to strongly agree.
  Final score for each OCEAN trait computed as linear combination of answers

Acquiring Personality Traits: Questionnaires

- Either through *questionnaires* or automatically *from user-generated data through ML*
  Questionnaires: more accurate, more labor-intensive/expensive
  Machine learning: less accurate, less expensive, possible to train on small amount of data and apply to large-scale data (e.g., microblogs, Likes, sensor data)

- Common instruments/questionnaires:
  Big Five Inventory (BFI-44):
  Questionnaire:
  https://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Personality-BigFiveInventory.pdf
  Questions like: “I see myself as someone who is curious about many different things.” rated from strongly disagree to strongly agree.
  Final score for each OCEAN trait computed as linear combination of answers

Acquiring Personality Traits: Machine Learning

- Nowadays, usually learned/predicted from user-generated data shared on social media
- Good survey: [Azucar et al., 2018]
- Common **data sources**:
  - Text: e.g., microblogs shared on Twitter or Sina Weibo; word embeddings
  - Images: e.g., shared on Instagram; color hue, saturation, objects, faces
  - Music: e.g., genre, mood, listening intensity, diversity
  - Interactions: e.g., Liked content on Facebook
  - Sensor data, e.g., created by smartphones; app usage, motion sensors, time, location
  - Metadata: e.g., amount of shared content, properties of friendship network
- Common **machine learning techniques**:
  - Support vector machines, random forests, neural networks

Acquiring Personality Traits: Machine Learning

- Best performing approaches in terms of correlations betw. predictions and ground truth:

  0.77 for Openness
  0.76 for Conscientiousness
  0.68 for Extraversion
  0.70 for Agreeableness
  0.71 for Neuroticism

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

[Azucar et al., 2018]

Personality and Item Preferences

- Many studies have shown correlations between personality traits and item/consumption preferences → makes personality a valuable attribute for recommender systems

  Examples:
- **Personality and *genre preferences* (movies, music, books) [Cantador et al., 2013]**
  - Study based on explicit Likes of 53K Facebook users on 16 genres in each domain; average personality scores of people who liked a given genre
- Personality and preferences for *diverse content* [Chen et al., 2013]
  - User study with 181 Chinese participants; diversity based on Gini index over movie genres, directors, countries, etc.; correlations between OCEAN and diversity scores
- Personality and *perception of affect* in music [Schedl et al., 2018]
  - Study of emotions in classical music (Beethoven's 3rd "Eroica"); 241 European participants; correlations between personality scores (TIPI) and perceived emotions

Personality and Genre Preferences

[Cantador et al., 2013]

	All users					
MOVIE GENRE	OPE	CON	EXT	AGR	NEU	#users
action	3.87	3.45	3.57	3.58	2.72	2488
adventure	3.91	3.56	3.54	3.68	2.61	179
animation	4.04	3.22	3.26	3.35	3.02	85
cartoon	3.95	3.33	3.49	3.57	2.81	957
comedy	3.88	3.44	3.58	3.60	2.75	3969
cult	4.27	3.10	3.45	3.40	3.16	38
drama	3.99	3.43	3.66	3.60	2.86	905
foreign	4.15	3.46	3.47	3.54	2.81	112
horror	3.90	3.38	3.52	3.47	2.91	2284
independent	4.31	3.59	3.51	3.55	2.69	104
neo-noir	4.34	3.35	3.33	3.37	2.97	92
parody	4.13	3.36	3.35	3.28	2.73	25
romance	3.84	3.48	3.62	3.62	2.85	776
science fiction	3.99	3.55	3.33	3.57	2.73	215
tragedy	4.40	3.34	3.27	3.52	3.11	26
war	3.82	3.51	3.49	3.50	2.71	148
	4.05	**3.41**	**3.46**	**3.51**	**2.84**	

Average personality scores

	All users					
BOOK GENRE	OPE	CON	EXT	AGR	NEU	#users
comic	4.06	3.28	3.38	3.47	2.86	1107
crime	3.83	3.44	3.43	3.47	2.99	191
drama	3.81	3.36	3.53	3.67	2.84	66
educational	4.02	3.66	3.57	3.66	2.74	977
fantasy	4.04	3.34	3.27	3.54	2.87	994
fiction	4.00	3.41	3.42	3.55	2.82	339
humor	3.90	3.40	3.62	3.56	2.78	743
mystery	3.91	3.53	3.51	3.61	2.76	302
non fiction	4.01	3.51	3.43	3.62	2.76	319
poetry	4.16	3.34	3.38	3.54	2.94	160
romance	3.89	3.52	3.49	3.60	2.85	1132
scary	3.81	3.41	3.68	3.55	2.83	1084
science fiction	4.13	3.42	3.25	3.51	2.81	1191
self help	4.03	3.50	3.42	3.62	2.83	196
thriller	3.85	3.54	3.51	3.59	2.76	639
war	3.87	3.44	3.33	3.23	2.80	108
	3.96	**3.44**	**3.45**	**3.55**	**2.83**	

Personality and Item Preferences

- Many studies have shown correlations between personality traits and item/consumption preferences → makes personality a valuable attribute for recommender systems

  Examples:
- Personality and *genre preferences* (movies, music, books) [Cantador et al., 2013]
  - Study based on explicit Likes of 53K Facebook users on 16 genres in each domain; average personality scores of people who liked a given genre
- **Personality and preferences for *diverse content* [Chen et al., 2013]**
  - User study with 181 Chinese participants; diversity based on Gini index over movie genres, directors, countries, etc.; correlations between OCEAN and diversity scores
- Personality and *perception of affect* in music [Schedl et al., 2018]
  - Study of emotions in classical music (Beethoven's 3rd "Eroica"); 241 European participants; correlations between personality scores (TIPI) and perceived emotions

Personality and Preferences for Diversity

[Chen et al., 2013]

Correlation coefficients between diversity (Gini index) and personality scores (OCEAN) as well as demographics (* p<0.05; ** p<0.01)

	Div(genre)	Div(director)	Div(country)	Div(release_time)	Div(actor/actress)
Neuroticism (N)	-0.04	**0.17***	0.06	-0.08	0.09
Extraversion (E)	0.02	**-0.15***	-0.15	-0.14	-0.07
Openness (O)	0.10	0.07	0.07	-0.07	**0.20***
Agreeableness (A)	-0.04	-0.17	**-0.18***	-0.04	-0.10
Conscientiousness (C)	-0.12	-0.16	**-0.15***	**0.15***	-0.10
Age	**-0.18***	0.13	-0.14	-0.05	-0.01
Gender	-0.13	**0.24****	**0.23****	-0.12	0.10
Education	-0.10	**-0.20****	**-0.20****	0.06	-0.04

JOHANNES KEPLER UNIVERSITY LINZ

Personality and Item Preferences

- Many studies have shown correlations between personality traits and item/consumption preferences → makes personality a valuable attribute for recommender systems

  Examples:
- Personality and *genre preferences* (movies, music, books) [Cantador et al., 2013]
  - Study based on explicit Likes of 53K Facebook users on 16 genres in each domain; average personality scores of people who liked a given genre
- Personality and preferences for *diverse content* [Chen et al., 2013]
  - User study with 181 Chinese participants; diversity based on Gini index over movie genres, directors, countries, etc.; correlations between OCEAN and diversity scores
- **Personality and *perception of affect* in music [Schedl et al., 2018]**
  - Study of emotions in classical music (Beethoven's 3rd "Eroica"); 241 European participants; correlations between personality scores (TIPI) and perceived emotions

Personality and Perception of Affect in Music

[Schedl et al., 2018]

Correlations between personality scores (TIPI) and emotions perceived in classical music (* p<0.05; ** p<0.01)

	Trans.	Peace.	Power	Joyful.	Tension	Sadness	Anger	Disgust	Fear	Surprise	Tender
Extraverted	0.045	0.024	0.120	0.065	0.022	0.031	-0.014	-0.027	0.007	0.041	**0.166****
Critical	0.010	0.031	0.094	0.081	0.049	0.037	-0.035	-0.041	-0.011	**-0.141***	0.043
Dependable	0.054	-0.098	-0.074	-0.098	0.009	-0.049	-0.065	-0.035	0.011	-0.018	0.007
Anxious	-0.084	-0.054	-0.108	-0.114	-0.108	-0.003	0.017	0.064	0.055	0.023	-0.089
Open to new experiences	**0.159***	**0.139***	0.108	**0.181****	0.054	0.053	0.010	0.005	-0.003	0.009	**0.222****
Reserved	-0.049	0.033	-0.112	-0.057	-0.095	-0.038	-0.033	-0.014	-0.045	-0.042	-0.084
Sympathetic	0.077	**0.147***	0.098	0.107	0.059	-0.031	-0.012	0.020	0.026	0.078	**0.166****
Disorganized	0.076	0.120	0.032	0.083	0.114	**0.167****	**0.157***	**0.146***	0.116	0.111	**0.129***
Calm	0.076	**0.142***	-0.002	**0.153***	-0.032	-0.023	-0.044	-0.060	0.031	-0.063	**0.132***
Conventional	**-0.145***	0.099	-0.048	0.012	**-0.135***	0.050	0.087	0.070	0.102	0.008	-0.058

Using Personality Traits for Recommendation: Domains

- Personality-based RSs have been proposed for different domains:
  - Movies [Nalmpantis and Tjortjis, 2017; Fernandez-Tobias et al., 2016]
  - Music [Lu and Tintarev, 2018; Fernandez-Tobias et al., 2016]
  - Images [Gelli et al., 2017]
  - Books [Fernandez-Tobias et al., 2016]
  - Computer games [Yang and Huang, 2019]
  - Recipes [Adaji et al., 2018]
  - Interest groups to join on social platforms [Wu et al., 2018]
  - Conference attendees [Asabere et al., 2018]

Using Personality Traits for Recommendation: Remarks

- Most approaches that integrate personality into RSs are (still) quite simple
- *Stand-alone (only personality) approaches* treat personality as a content descriptor of items and use memory-based CBF / direct user-item matching [Yang and Huang, 2019]
- Commonly, *hybrid approaches* that integrate personality into CF or CBF are used, e.g.:
  - Often linearly combine collaborative similarity [Nalmpantis and Tjortjis, 2017] or content-based similarity [Wu et al., 2018] with similarity based on personality
  - Integrate personality into context-aware systems (e.g., CA-FMs) [Gelli et al., 2017]
  - Extending matrix factorization with personality factors [Fernandez-Tobias et al., 2016]
  - Adopt graph-based techniques, personality-based subgraph extraction [Adaji et al., 2018]
- For user-item matching, "personality" of items is commonly modeled via OCEAN scores extracted from user-generated text (reviews, microblogs, etc.); seems disputable

Using Personality Traits for Recommendation: Examples

[Yang and Huang, 2019]

- Recommender for computer games, purely based on personality traits
- **User** modeling: 5-dim. vector of OCEAN scores (*UP*), predicted from their social media posts
- **Item** modeling: 5-dim. vector of OCEAN scores (*GP*), predicted from
  - OCEAN scores of the users playing the game
  - OCEAN scores extracted from game reviews

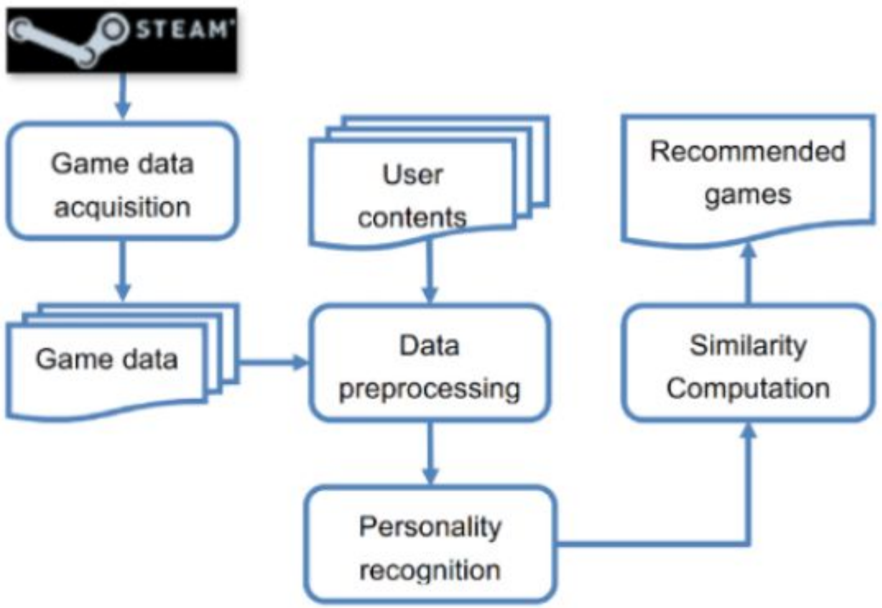
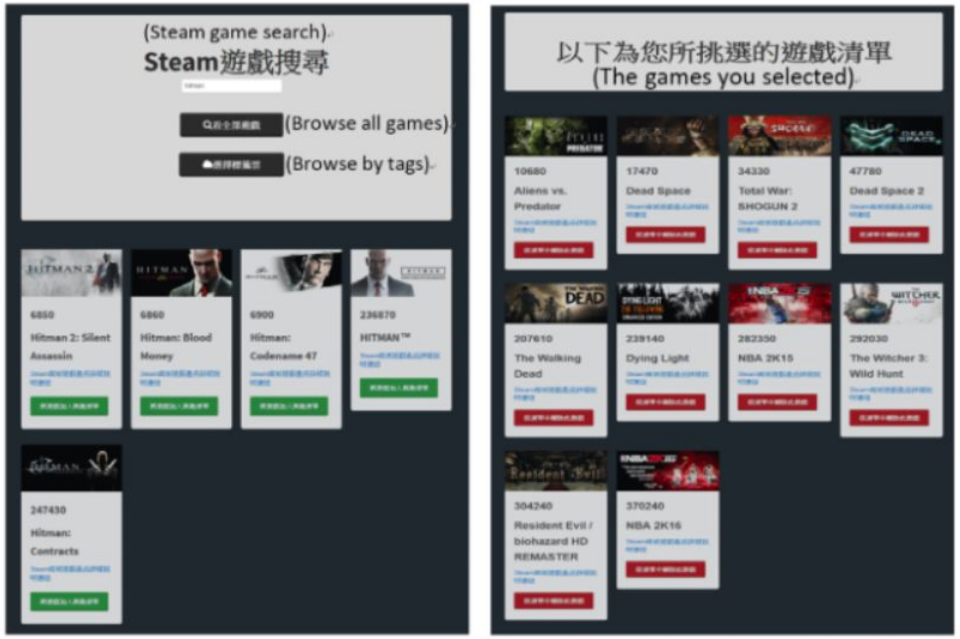


Fig. 2. An example of the user reviews on the list sorted by 'MOST HELPFUL' option. Most (93%) of the rating users approved this review as shown in the figure.



(a) Favorite game submission interface (b) List of favorite games submitted by the reviewer

Using Personality Traits for Recommendation: Examples

[Yang and Huang, 2019]

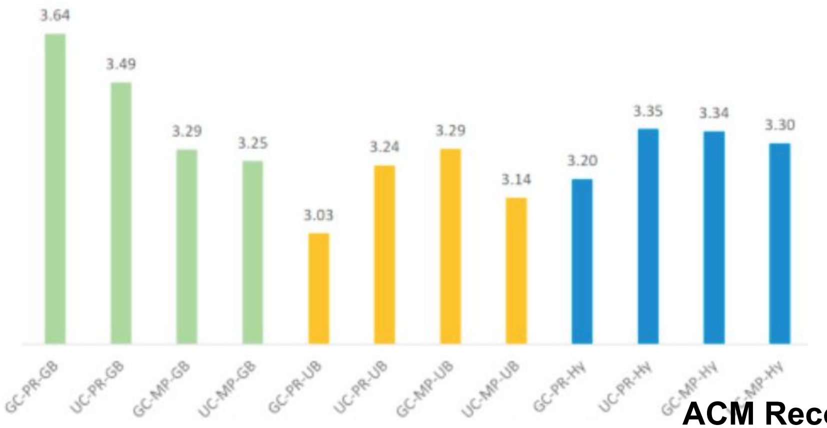
- Recommender for computer games, purely based on personality traits
- **Recommendation approaches:**
  - Direct user-game matching:
    cosine sim. between *UP* and *GP* (S_{user})
  - CBF variant based on *GP* of games
    the target user interacted with (S_{game})
  - Linear combination of both (S_{hybrid})

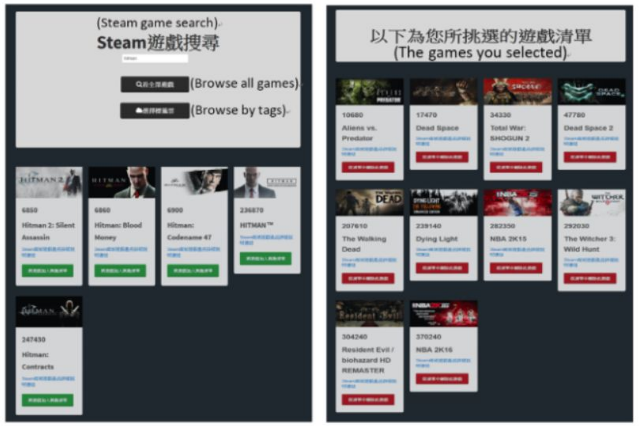
$$S_{user}(G_i, U_j) = \frac{\mathbf{P}_{G_i} \cdot \mathbf{P}_{U_j}}{\|\mathbf{P}_{G_i}\| \|\mathbf{P}_{U_j}\|}$$

$$S_{game}(G_i, U_j) = \frac{1}{|C_{U_j}|} \sum_{g \in C_{U_j}} \frac{\mathbf{P}_{G_i} \cdot \mathbf{P}_g}{\|\mathbf{P}_{G_i}\| \|\mathbf{P}_g\|}$$

$$S_{hybrid}(G_i, U_j) = w_u S_{user}(G_i, U_j) + w_g S_{game}(G_i, U_j)$$

- **Evaluation**:
  - User study among 63 players
  - Users scored recommendations of approaches on 5-point scale
  - CBF approach scored best





(a) Favorite game submission interface (b) List of favorite games submitted by the reviewer

Using Personality Traits for Recommendation: Examples

[Lu and Tintarev, 2018]

- Recommender for music items (songs)
- Hybrid re-ranking approach based on CF via factorization machine (FM)
- Personalize/re-rank results of FM by tailoring level of *diversity* in recommendation lists to user's preference for diversity
- **User** modeling: OCEAN scores, gathered explicitly via Ten Item Personality Inventory (TIPI)
- **Item** modeling: release year, artist, genre, tempo, key; correlations between OCEAN and diversity needs
- **Diversity** definition: intra-list diversity (avg. pairwise dissimilarity between items in rec. list) of genre, artist, and key

Using Personality Traits for Recommendation: Examples

[Lu and Tintarev, 2018]

- **Recommendation approach:**
  - Create initial/original recommendation list *O* via FM (trained on MSD Taste Profile data)
  - Create re-ranked list *R* by minimizing objective function when selecting next item from *O*

$$argmin_{p \in O \setminus R} \ (1-\lambda) \cdot rank(p, O) + \lambda \cdot div_{overall}(p, R) \qquad div_{overall}(p, R) = \sum_{i=1...n} \theta_i \cdot div_i(p, R)$$

R…re-ranked list so far (initialized with top-ranked item of *O*)

rank(p,O)...rank of item *p* in original list *O*

$div_i(p,R)$...average diversity of *R* w.r.t. item *p*

weights λ and θ_i computed from *u*'s OCEAN scores and correlation with diversity preferences

	E	A	C	ES	O
Div(Release times)	-0.03	-0.12	0.01	0.11	-0.15
Div(Artists)	0.10	0.09	0.11	**0.22****	-0.04
Div(Artists number)	0.00	**0.25****	0.13	0.15	0.07
Div(Genres)	0.07	0.00	-0.01	**0.25****	0.06
Div(Tempo)	0.11	0.09	0.11	**0.24****	0.08
Div(Key)	**0.21****	0.05	0.06	0.17*	0.08

Correlations between diversity preferences and OCEAN scores (* p<0.05; ** p<0.01)

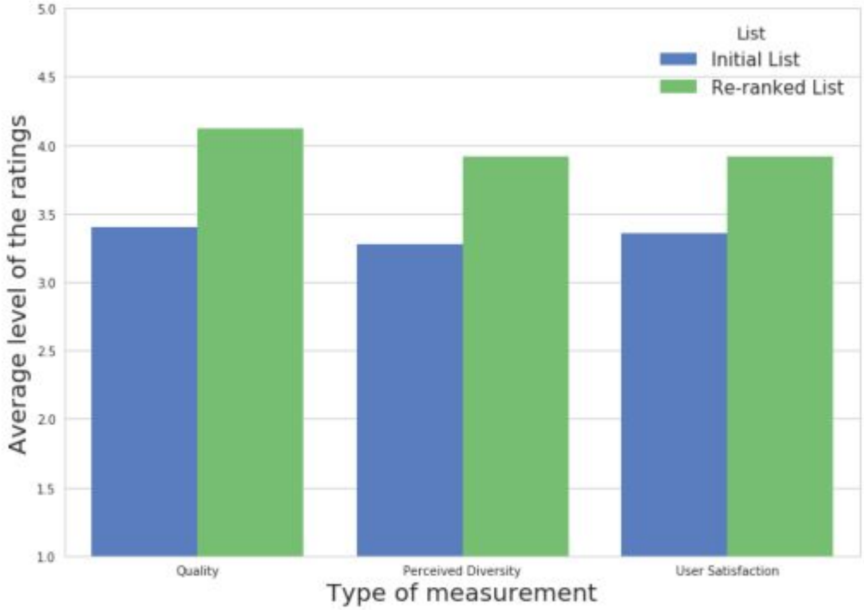
JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Using Personality Traits for Recommendation: Examples

[Lu and Tintarev, 2018]

- **Evaluation:**
  - User study among 25 participants
  - Participants judged original recommendations and re-ranked recommendations on 5-point scale w.r.t. quality, diversity, overall satisfaction



Affect-aware Recommender Systems

Affect-aware Recommender Systems

- Definition/Motivation
- Modeling mood/emotion (categorical vs. dimensional models)
- Acquiring affective cues
- Using affective cues for recommendation

Definition/Motivation

- **Emotion:**
  - High-intensity affective experience
  - Response to a stimulus
  - Short duration (seconds to minutes)
- **Mood:**
  - Low-intensity affective experience
  - Long duration (minutes to hours)

- **Motivation:**
  - Increase level of *personalization* of RSs
  - Regulate user’s mood
  - Exploit interdependence between item preferences, personality, and mood

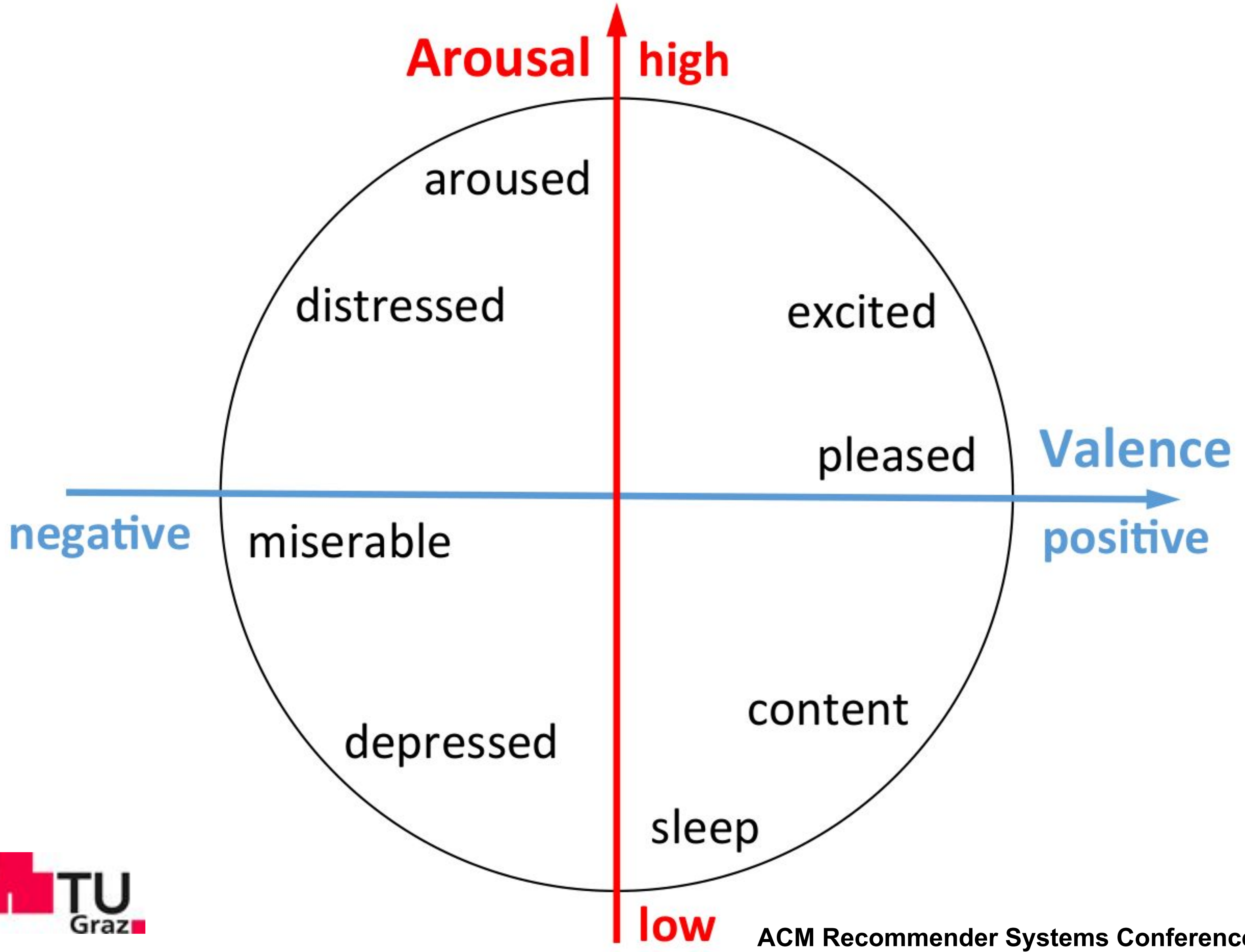
Modeling Affect

- **Categorical models:**
  - Affect is described via distinct categories
  - E.g., Ekman’s six basic emotions: happiness, sadness, disgust, fear, surprise, anger
- **Dimensional models:**
  - Affect is described on a continuous scale along 2 (or 3) dimensions
  - Valence: level of pleasantness (positive vs. negative)
  - Arousal: level of intensity (high vs. low)
  - (Dominance): How much is one in control of their emotion?
- **Hybrid models:**
  - Continuous or ordinal scale within each of a set of categories

Dimensional Affect Model: Valence-Arousal Plane

Russel's two-dimensional circumplex model
(with emotions integrated)

[Russel, 1980]



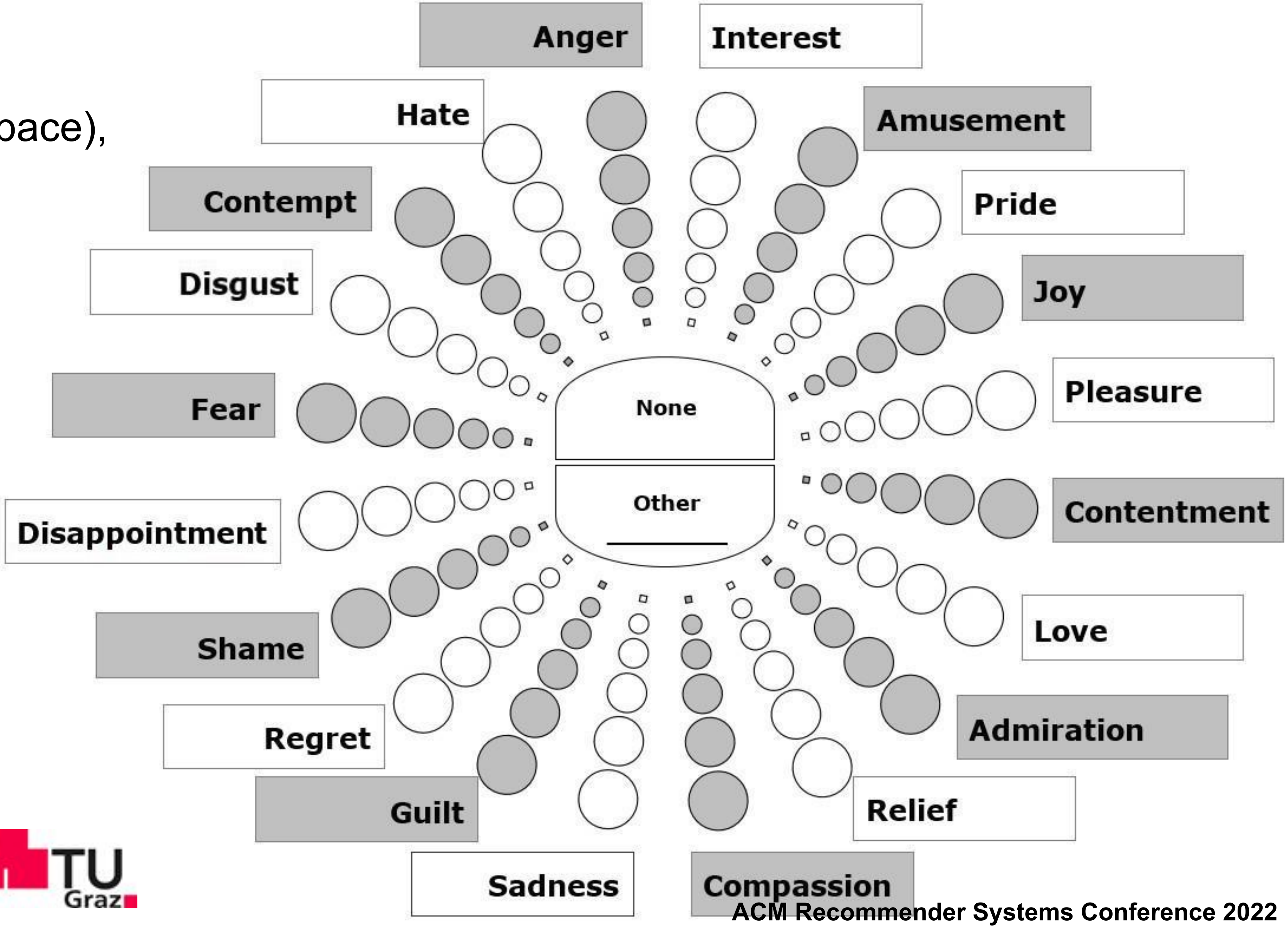
JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Hybrid Affect Model: Geneva Emotion Wheel

Various emotion dimensions (roughly arranged w.r.t. V/A space), the intensity of each is described on an ordinal scale

[Scherer, 2005]



JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Acquiring Affective Cues

- Like for personality: *explicitly* (asking users) or *inferred via ML* from user-generated data

- **Explicit acquisition:**
  - More accurate, but more labor-intensive/expensive
  - Typically, user chooses one emotion from a set of emotion categories; Less often, user positions a cursor in a V/A space visualization

- **Machine learning:**
  - Less accurate, but less expensive
  - Typically, predicted from user-generated texts shared on social media or sensor data

Using Affective Cues for Recommendation: Remarks & Domains

- Overall, less works on emotion-aware RSs than on personality-aware RSs
- Mostly, quite simple extensions to CF or CBF approaches, or even just matching mood(item) ↔ mood(user)
- Affect-aware RSs have been proposed for the several domains, including:
  - Locations: [Ravi and Vairavasundaram, 2017]
  - Fashion: [Piazza et al., 2017]
  - Music: [Kaminskas et al., 2013; Deng et al., 2015; Ayata et al., 2018]
  - Music and Arts: https://ars.electronica.art/newdigitaldeal/en/music-tower-blocks
  - Generally, lots on music since music is known to evoke stronger emotions than most other stimuli

Using Affective Cues for Recommendation: Examples

[Ravi and Vairavasundaram, 2017]

- Recommender for locations / points-of-interest
- **User** modeling: lexicon-based emotion classification from posts shared on social media, using categorical model (happy, surprised, angry, sad, fear, ...) → emotion vector
- **Item** modeling: lexicon-based emotion classification from posts shared at a particular location, using categorical model (happy, surprised, angry, sad, fear, ...) → emotion vector
- **Recommendation approaches:**
  - User-based CF: similarity between users (*u*, *v*) are computed as product of their emotional sim. (between their emotion vectors) and sim. between the current emotion vector of target user *u* and *v*'s emotion at the location
  - Item-based CF: predicts emotionally most similar locations to those *u* already visited
  - Hybrid: linear combination of both

$$S_{user}(u,v) = S_{user}^{emo}(u,v) \cdot S_{user}^{loc}(u,v)$$

$$S_{user}^{emo}(u,v) = \frac{E_u \cdot E_v}{\|E_u\| \cdot \|E_v\|}$$

$$S_{user}^{loc}(u,v) = \frac{E_u(now) \cdot E_v(loc)}{\|E_u(now)\| \cdot \|E_v(loc)\|}$$

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

Using Affective Cues for Recommendation: Examples

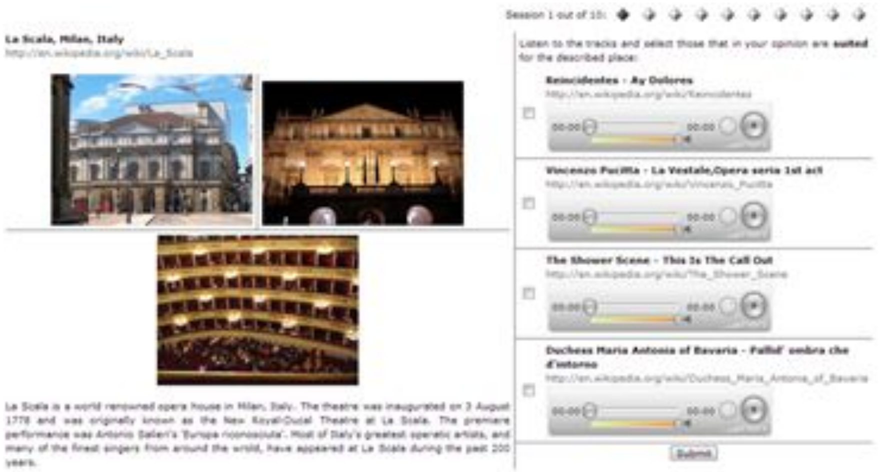
[Kaminskas et al., 2013]

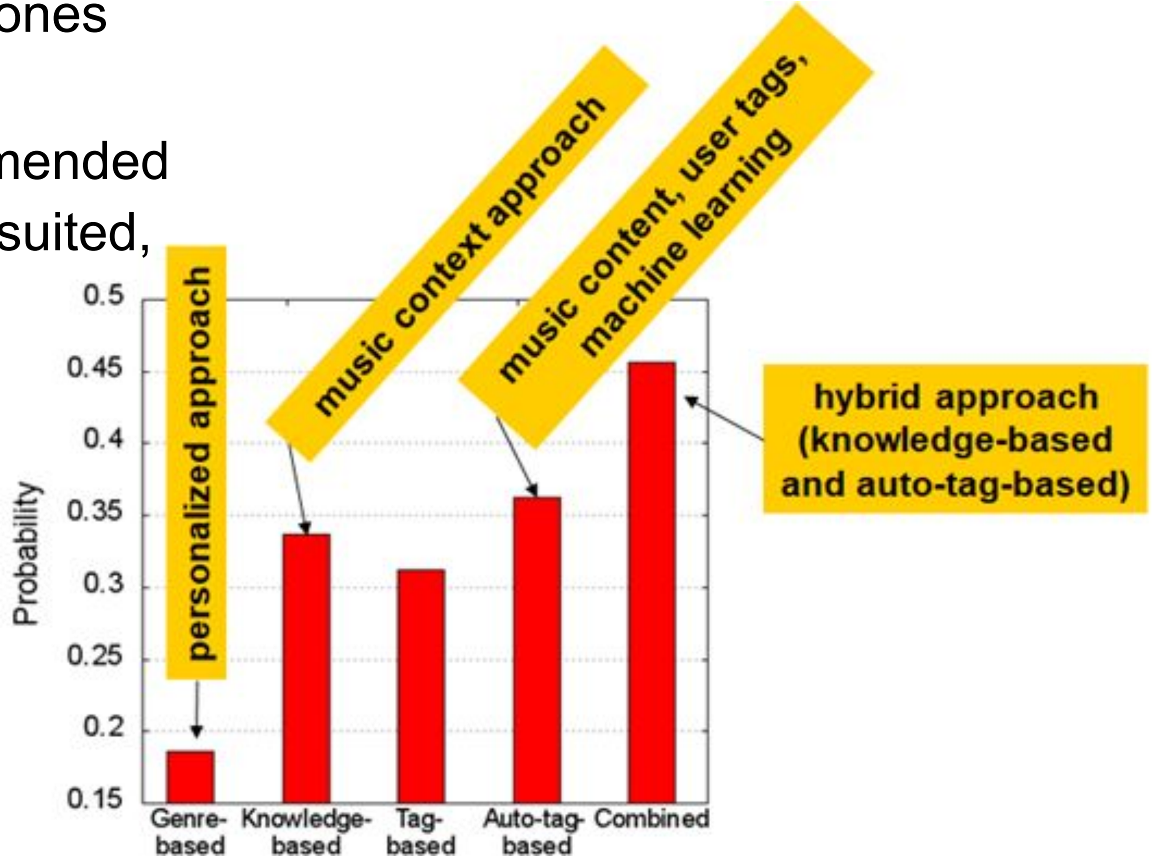
- Recommender for music pieces given a place-of-interest
- Given a place-of-interest, identify best-suited music, via matching emotional cues
- Modeling **place-of-interest**: bag-of-words (BoW) representation of 24 emotion categories (annotated via web survey)
- Modeling **music track**: BoW representation of 24 emotion categories (predicted via music auto-tagger, trained on user annotations)
- **Recommendation approaches:**
  - Auto-tag-based: Jaccard similarity between *t*rack's BoW and *p*lace's BoW: $S(t, p) = \frac{|E_t \cap E_p|}{|E_t \cup E_p|}$
  - Knowledge-based: Infer similarity between *t* and *p* from path statistics in DBpedia KG
  - Hybrid: Borda rank-aggregation of the two recommendation lists of above approaches

Using Affective Cues for Recommendation: Examples

[Kaminskas et al., 2013]

- **Evaluation:**
  - Web-based user study among 58 participants
  - Users were given the pooled and randomized recommendations, then had to indicate which ones matched of given PoI
  - Performance measure: share of tracks recommended by an approach *A* which were marked as well-suited, among all recommendations made by *A*





[https://ars.electronica.art/newdigitaldeal/en/music-tower-blocks]

Emotional Music Tower Blocks (EmoMTB)
Emotion-aware Music Recommendation and Exploration

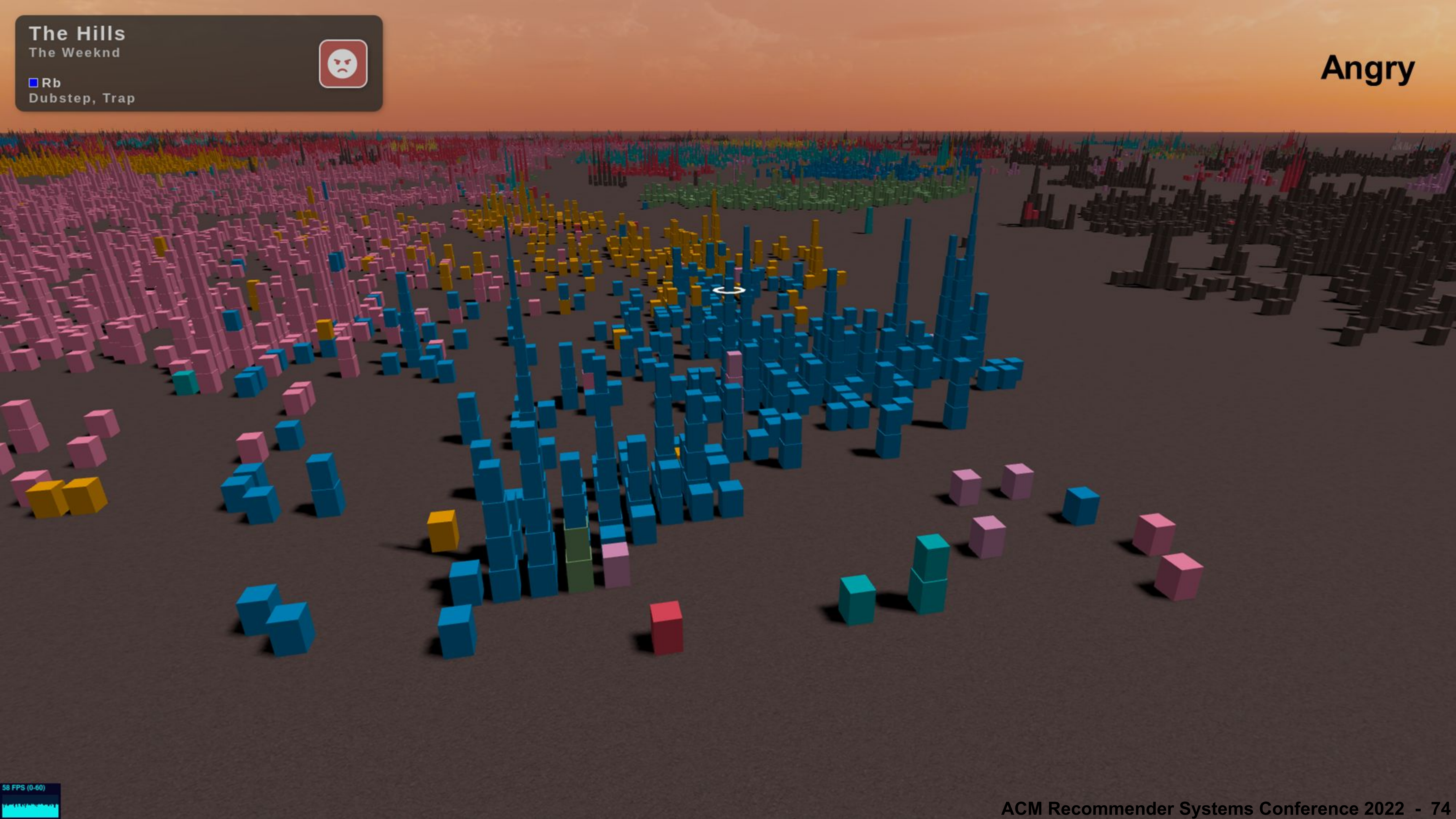
- Artistic/scientific project presented at Ars Electronica Festival of Media Arts 2021
- Audiovisual exploration of a music collection (~500K tracks) using metaphor of city
- Tracks are clustered based on (very fine-grained) genre information and audio features
- Visualized as blocks; very similar ones are stacked to form buildings
- Nearby buildings form neighborhoods of similar genres (genres are color-coded)
- Each track is assigned an emotion (predicted from Last.fm tags)
- User selects an emotion → recommendations and visualizations update accordingly
- Explanatory video: https://bit.ly/3hfVH1S

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

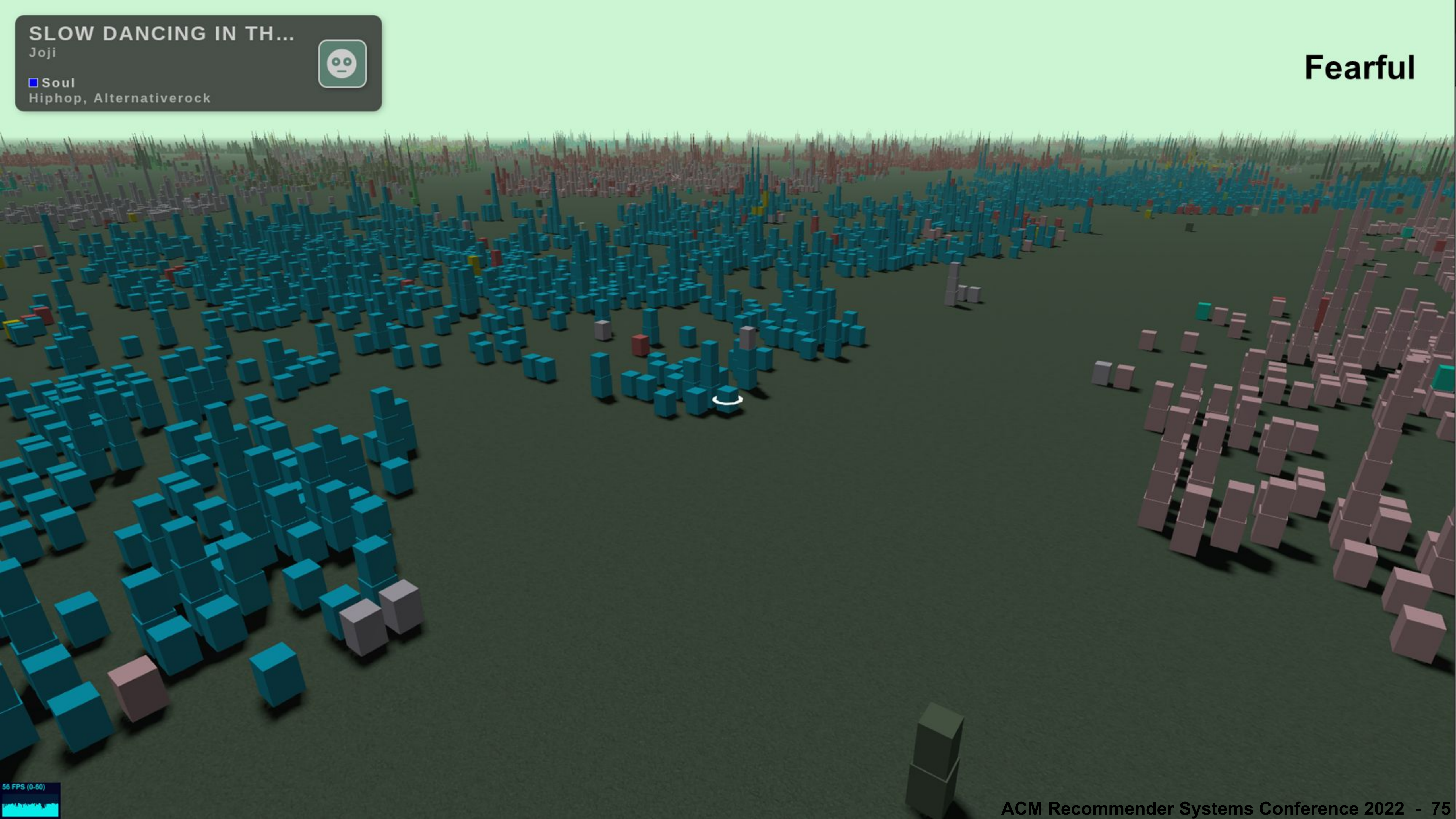
EmoMTB: User Controls

Hips Don't Lie (feat. W...
Shakira, Wyclef Jean
Pop
Latin, Reggaeton
Happy
55 FPS (0-60)



Angry

Fearful


SLOW DANCING IN TH...
Joji
Soul
Hiphop, Alternativerock
56 FPS (0-60)

Part III: User-centric Evaluation

A Quick Overview of RecSys Evaluation

- Traditionally: strong focus on algorithmic performance & recommendation accuracy
  - Overview of accuracy metrics: [Gunawardana and Shani, 2009]
- Nowadays: many metrics, e.g. diversity, novelty, serendipity, fairness
- Simulation
- Classic recommender systems evaluation techniques:
  - Offline evaluation: pre-collected datasets
  - Online evaluation: observe user behavior in real world (A/B tests)
  - User studies: smaller groups of users report their experience with the RecSys
  - or combinations of the above
- See Recommender Systems Handbook chapter on evaluation [Shani and Gunawardana, 2011]

User-Centric Evaluation

- User experience: delivery of recommender systems outputs to users & users' interactions with recommendations [Konstan and Riedl, 2012]
- User experience helps us understand many relevant issues
  - Use of RecSys
  - Perceived value
  - Factors that shape decision making processes [Xiao and Benbasat, 2007]
    - User attitude, motivation, perceived trust, perception of recommendations
- User-centric evaluation requires user experiments (user studies, randomized field trials)

→ RecSys handbook chapter on **Evaluating Recommender Systems with User Experiments** [Knijenburg and Willemsen, 2015]

Studies of User Experience

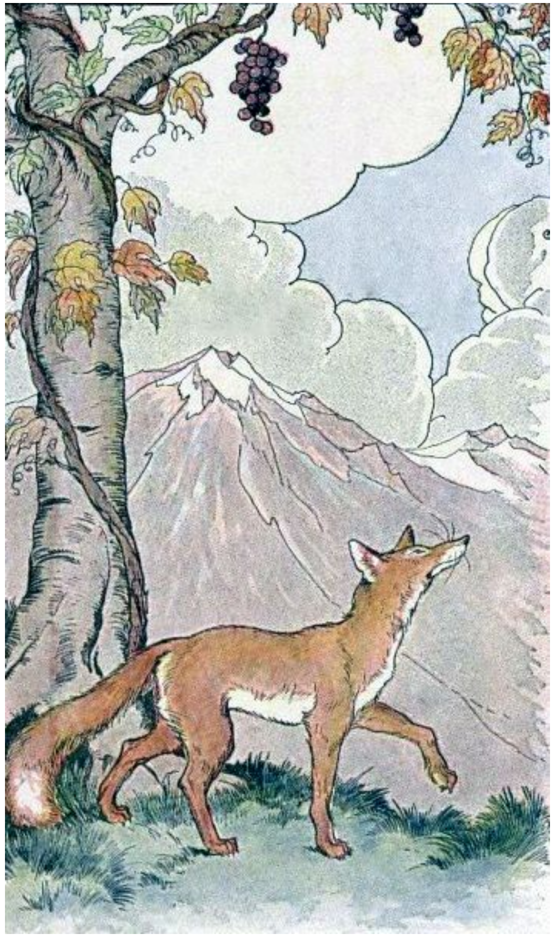
- Improving preference elicitation [McNee et al., 2003]
- Increase user satisfaction [Ziegler et al., 2005]
- User engagement [O’Brien and Toms, 2008]
- Trust in the system [Pu and Chen, 2006]
- Improving recommendation interfaces [Cosley et al., 2003]
- and many more

Next: psychological factors that influence user experience

Cognitive dissonance [Festinger, 1954]

- Aversive cognitive-affective response to exposure to contradicting information
    - can lead to users losing trust in the system

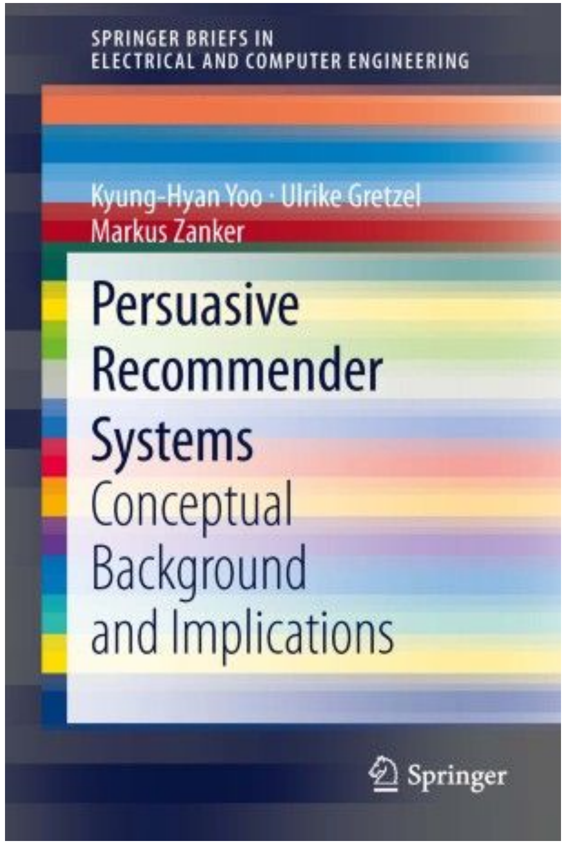
- Can happen when recommendations are *inconsistent* with user preferences [Schwind et al., 2011]
- Or, when users *reevaluate* a choice they made due to a RecSys [Surendren and Bhuvaneswari, 2014]



https://en.wikipedia.org/wiki/Cognitive_dissonance #/media/File:The_Fox_and_the_Grapes.jpg

Persuasion [Fogg, 2002]

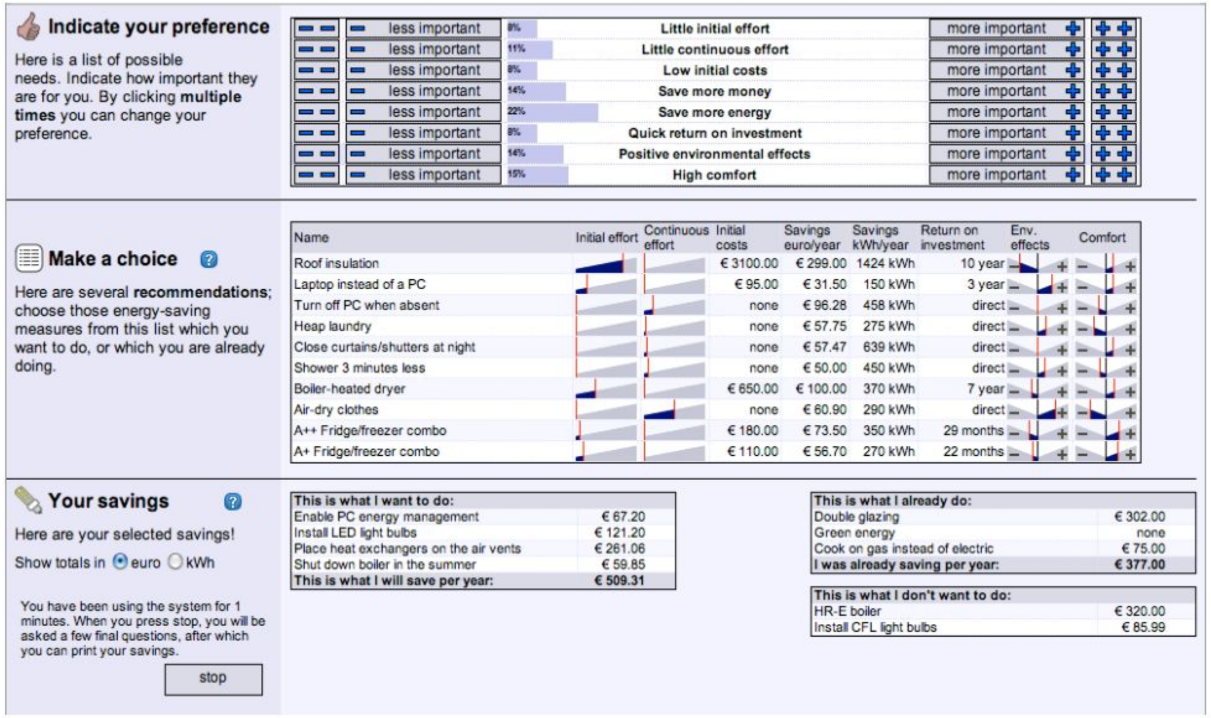
- Communication process to convince others to adapt their behavior and attitudes
- Persuasive recommender systems: users are influenced by
  - the RecSys itself (source)
  - the recommendation (message)
  - the user (target)
  - the context in which recommendation is delivered
- RecSys are persuasive when
  - credibility is high [Yoo and Gretzel, 2011]
  - perceived novelty of recommendations is high [Cremonesi et al., 2012]
  - items are attractive [Felfernig et al., 2008]
  - explanations are given [Tintarev and Masthoff, 2012]



JOHANNES KEPLER UNIVERSITY LINZ

Interaction methods and interfaces

- [Knijnenburg et al., 2011] studied 5 interaction methods
  - top-N recommendation list
  - sort method to sort recommendations according to users' preferences
  - explicit method to let users explicitly express their preferences via weights
  - implicit method to assign weights based on interaction history
  - hybrid combination of explicit and implicit
- User study showed that most users most satisfied with hybrid method
  - satisfaction, trust, perceived control, choice quality



Explicit interact method - top-N, sort, implicit look similar but do not show weights

JOHANNES KEPLER UNIVERSITY LINZ TU Graz

More details on user interaction in RecSys: [Jugovac and Jannach, 2017]

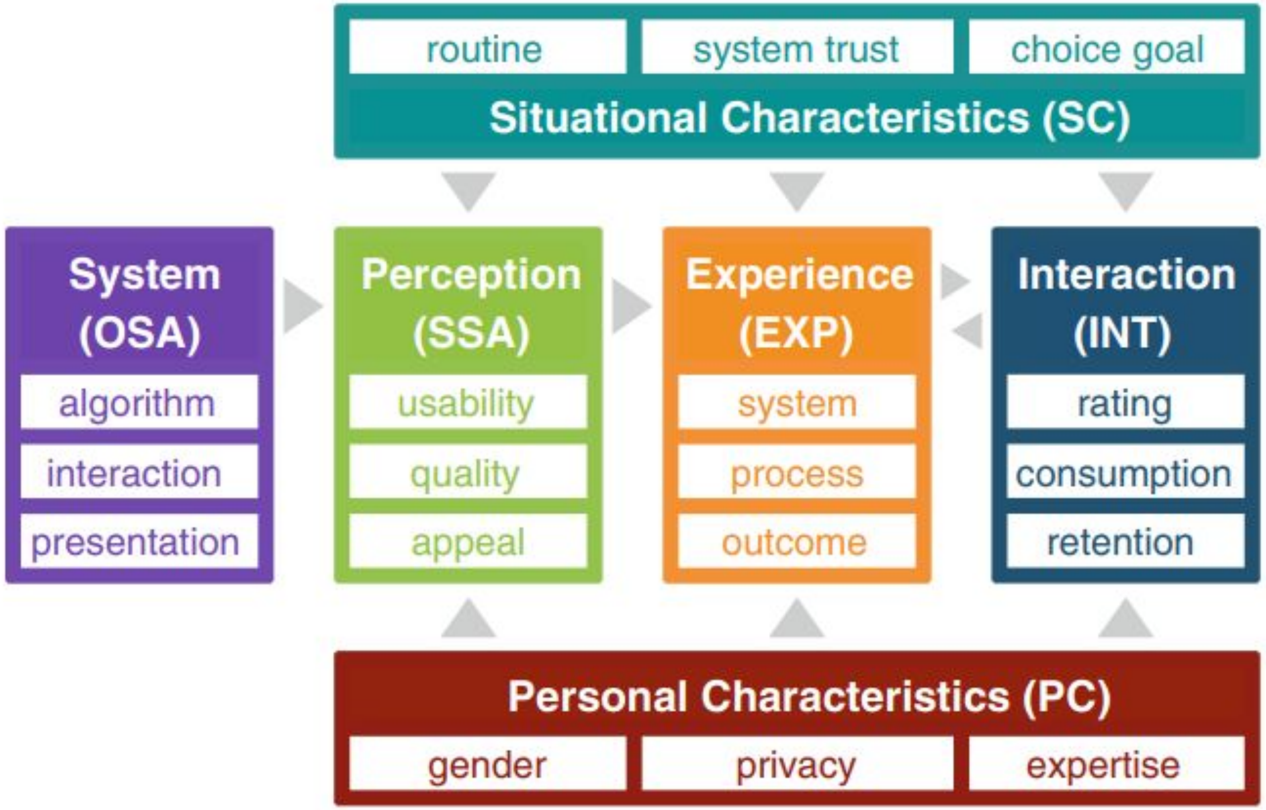
Designing User Studies for RecSys Evaluation

- User studies rely on self-reports [McCroskey et al., 1984]
  - tests and measures that users need when reporting on their behavior
- Self-reports require representative samples of participants
- Leverage latent factor analysis e.g., Exploratory Factor Analysis [O'Brien and Toms, 2010]
  - first: group inter-item correlations into distinct dimension
  - then: run a Confirmatory Factor Analysis on independent data set to validate factor structure

- Creating user studies for RecSys research is labor some - several frameworks exist to support researchers develop hypotheses and design user studies

Framework by Knijnenburg and Willemsen (2015)

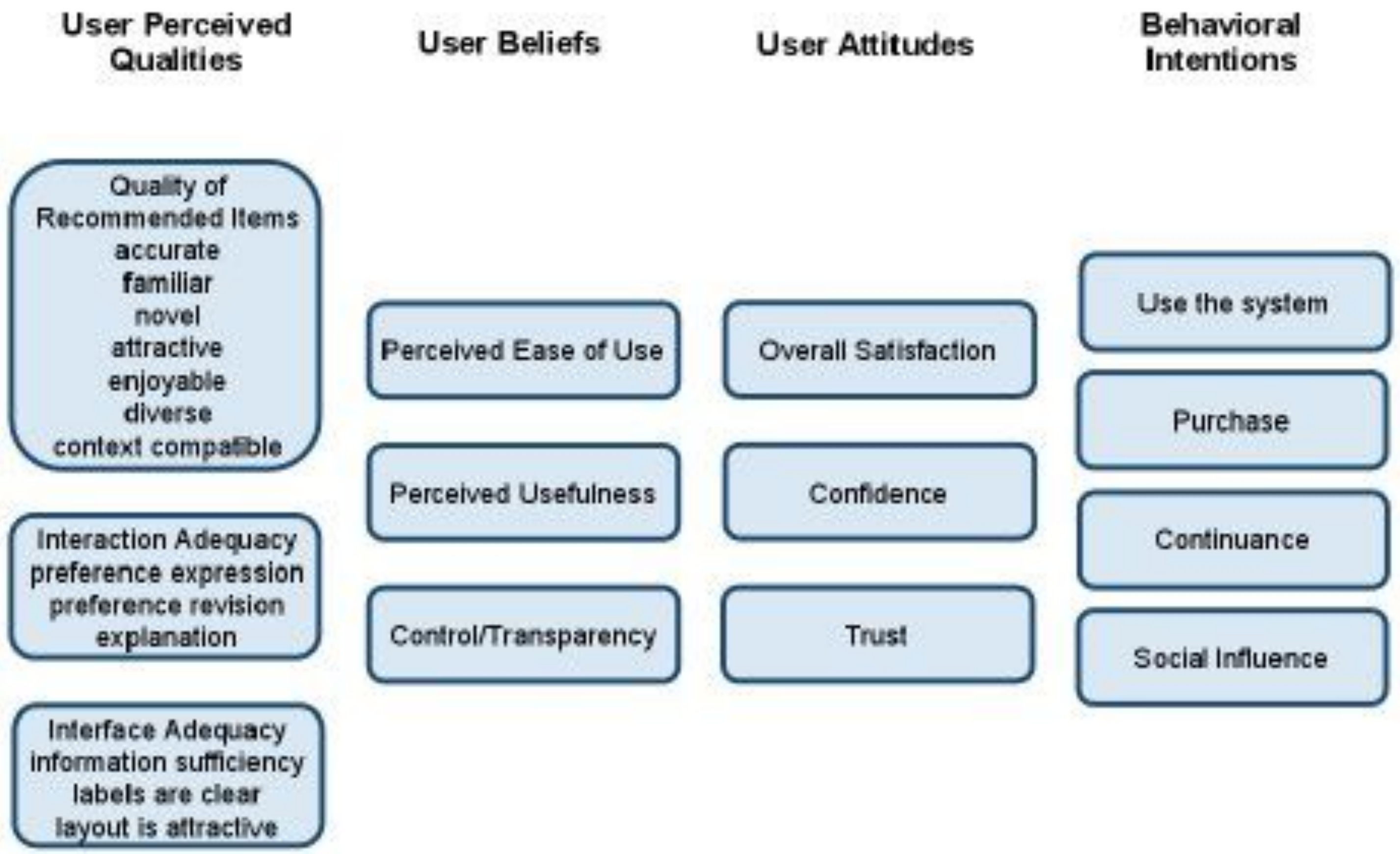
- Based on theories of attitude and behaviors, technology acceptance and user experience
- Enables to study how users' subjective perception (e.g., perceived quality), in combination with personal and situational characteristics, influence the user experience with the RecSys
- Situational and personal characteristics: help account for context-relevant information and individual variables



JOHANNES KEPLER UNIVERSITY LINZ
TU Graz

ResQue framework by Pu et al. (2011)

- Designed to assess perceived recommendation quality, usability, interface adequacy, interaction quality and overall user satisfaction with the RecSys and the user’s behavioral intentions



Part IV: Grand Challenges

JOHANNES KEPLER UNIVERSITY LINZ

Grand Challenges: Cognition-informed RecSys

- Related work shows strong link between human memory processes & user behavior
- Scarce work on using recommender systems to **support** human memory in retrieving objects
  - [Elsweiler, 2007]: design improved information management tools based on research on how humans recover from memory lapses
  - [Gemmel et al., 2002]: augmenting human memory - MyLifeBits - system that reminds users of their stored its
  - Both works highlight importance of context for memory retrieval!

    → Opportunities for future research
- Incorporating user’s attention crucial research topic in RecSys
  - Link to psychological models and theories of attention yet underexplored - could lead to more transparent and explainable models

Grand Challenges: Personality-based RecSys

- Still not well understood to which extent personality influences perceived recommendation quality
  - Variability between users & domains
  - Personality could be perceived as irrelevant, or invasive concerning privacy and ethics
- Using personality signals in a privacy-aware fashion needs more research!
- Current approaches integrate personality using quite simplistic ways
  - e.g.: extensions of standard CF, linear combinations between content-based similarity & personality/user-based similarity metric
  - Recent work by Beheshti et al. (2020): personality signals as features in a neural embedding framework → more research needed how to integrate personality into current DL methods!
- Personality traits on the item level still underresearched topic

Grand Challenges: Affect-aware RecSys

- Not well understood to which extent a user's mood or emotion influences perceived recommendation quality (like in the case of personality)
- More research needed on importance of mood or emotion changes during item consumption
  - Detecting such changes challenging
  - Integrating affect dynamics into recommender systems
- Again, mood and emotion are sensitive information
  - More research needed to make emotion detection and inclusion of emotion as a contextual factor in recommender systems privacy-aware.

Grand Challenges: Evaluation

- Many papers we discuss in our survey employ standard performance metrics from information retrieval and machine learning for evaluation
    - Future work: explore what metrics psychology-informed recommender systems can improve beyond accuracy, such as algorithmic fairness or transparency
    - Frameworks like the one presented by Deldjoo et al. (2021) could be applied to evaluate user and item fairness and to devise suitable metrics.
- More research is also needed on the online performance of psychology-informed recommender systems to better understand whether their recommendations result in higher user satisfaction

Grand Challenges: Evaluation

- Many psychological factors influence how users experience recommender systems
- Identifying and understanding such subjective factors requires user studies
- User study design largely influenced by psychological methods
  - e.g., questionnaires, factor analysis, etc.
- Conducting such studies with ecological validity in mind can be challenging
  - in particular, to gather a sufficiently large sample of participants that allows for drawing significant and meaningful conclusions
  - design studies that do not overburden users but still result in sufficient amounts of data
- One solution to facilitate the design and execution of user studies: evaluation frameworks
- Major challenge: access to real-world systems & ability to observe long-term user behavior

Our Vision

“Our vision for future recommender systems research is, therefore, to draw from the decent knowledge of these disciplines in the entire workflow of creating and evaluating recommender systems. Corresponding systems should, as a result, holistically consider extrinsic and intrinsic human factors; corresponding research should adopt a genuinely user-centric perspective.”

Lex, E., Kowald, D., Seitlinger, P., Tran, T.N.T., Felfernig, A., and Schedl, M. **Psychology-informed Recommender Systems**, Foundations and Trends in Information Retrieval, 15(2):134-242, 2021. http://dx.doi.org/10.1561/1500000090

References

[Adaji et al., 2018]: Adaji, I., Sharmaine, C., Debrowney, S., Oyibo, K., Vassileva, J. *Personality Based Recipe Recommendation Using Recipe Network Graphs*, Social Computing and Social Media: Technologies and Analytics. Ed. by G. Meiselwitz, Springer, 161-170, 2018.

[Anderson et al., 2004]: Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y. *An integrated theory of the mind*. Psychological review, 1(4):1036, 2004.

[Asabere et al., 2018]: Asabere, N.Y., Acakpovi, A., Michael, M.B. *Improving Socially-Aware Recommendation Accuracy Through Personality*, IEEE Transactions on Affective Computing 9(3):351-361, 2018.

[Atkinson & Shiffrin, 1968]: Atkinson RC, Shiffrin RM. *Human memory: A proposed system and its control processes*. In Psychology of learning and motivation, vol. 2, pp. 89-195, 1968.

[Ayata et al., 2018]: Ayata, D., Yaslan, Y., Kamasak, M.E. *Emotion Based Music Recommendation System Using Wearable Physiological Sensors*, IEEE Transactions on Consumer Electronics, 64(2):196-203, 2018.

[Azucar et al., 2018]: Azucar, D., Marengo, D., Settanni, M. *Predicting the Big 5 personality traits from digital footprints on social media*: *A meta-analysis*, Personality and Individual Differences 124:150-159, 2018.

[Cantador et al., 2013]: Cantador, I., Fernández-Tobías, I., Bellogín, A. *Relating Personality Types with User Preferences in Multiple Entertainment Domains*. Proc. UMAP Workshops 2013.

References

[Chen et al., 2013]: Chen, L., Wu, W., He, L. *How personality influences users' needs for recommendation diversity?*. Proc. CHI Extended Abstracts 2013.

[Conrad et al., 2019]: Conrad, F., Corey, J., Goldstein, S., Ostrow, J., & Sadowsky, M. (2019). *Extreme re-listening: Songs people love... and continue to love*. Psychology of Music, 47(2), 158-172.

[Deng et al., 2015]: Deng, S., Wang, D., Li, X., Xu, G. *Exploring User Emotion in Microblogs for Music Recommendation*, Expert Systems with Applications, 42(23):9284-9293, 2015.

[Ebbinghaus, 1885]: Ebbinghaus H. Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Duncker & Humblot; 1885.

[Fernandez-Tobias et al., 2016]: Fernandez-Tobias, I., Braunhofer, M., Elahi, M., Ricci, F., Cantador, I. *Alleviating the New User Problem in Collaborative Filtering by Exploiting Personality Information*, User Modeling and User-Adapted Interaction 26(2-3):221-255, 2016.

[Gelli et al., 2017]: Gelli, F., He, X., Chen, T., Chua, T.-S. *How Personality Affects Our Likes: Towards a Better Understanding of Actionable Images*, Proc. ACM Multimedia, 1828-1837, 2017.

References

[Goldberg et al., 2006]: Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., Gough, H.G. *The international personality item pool and the future of public-domain personality measures*, Journal of Research in Personality 40(1):84-96, 2006.

[Gunawardana and Shani, 2009]: Gunawardana, A. and G. Shani. 2009. *A survey of accuracy evaluation metrics of recommendation tasks.* Journal of Machine Learning Research. 10(12).

[Hamilton, 1979]: Hamilton, D. L. *A cognitive-attributional analysis of Stereotyping*. In Advances in experimental social psychology (Vol. 12, pp. 53-84). Academic Press. 1979

[Kaminskas et al., 2013]: Kaminskas, M., Ricci, F., Schedl, M. Location-aware music recommendation using auto-tagging and hybrid matching, Proc. ACM RecSys, 17-24, 2013.

[Lex et al., 2020]: Lex E, Kowald D, Schedl M. *Modeling popularity and temporal drift of music genre preferences.* Transactions of the International Society for Music Information Retrieval. 2020 Mar 25;3(1).

[Lu and Tintarev, 2018]: Lu, F., Tintarev, N. *A Diversity Adjusting Strategy with Personality for Music Recommendation*, Proc. IntRS@RecSys, 7-14, 2018.

[Melchiorre et al., 2021]: Melchiorre, A. B., Rekabsaz, N., Parada-Cabaleiro, E., Brandl, S., Lesota, O., & Schedl, M. (2021). *Investigating gender fairness of recommendation algorithms in the music domain.* Information Processing & Management, 58(5), 102666.

References

[Nalmpantis and Tjortjis, 2017]: Nalmpantis, O., Tjortjis, C. *The 50/50 Recommender: A Method Incorporating Personality into Movie Recommender Systems*, Engineering Applications of Neural Networks. Ed. by G. Boracchi, L. Iliadis, C. Jayne, and A. Likas, Springer, 498-507, 2017.

[Peretz et al., 1998] Peretz, I., Gaudreau, D., & Bonnel, A. M. (1998). *Exposure effects on music preference and recognition*. Memory & cognition, 26(5), 884-902.

[Piazza et al., 2017]: Piazza, A., Kröckel, P., Bodendorf F. Emotions and Fashion Recommendations: Evaluating the Predictive Power of Affective Information for the Prediction of Fashion Product Preferences in Cold-start Scenarios, Proc. ACM Web Intelligence, 1234-1240, 2017.

[Ravi and Subramaniyaswamy, 2017]: Ravi, L., Subramaniyaswamy, V. *Learning Recency and Inferring Associations in Location Based Social Network for Emotion Induced Point-of-Interest Recommendation*, Journal of Information Science and Engineering, 33(6):1629-1647, 2017.

[Ren, 2015]: Ren L. *A Time-Enhanced Collaborative Filtering Approach*. In 2015 4th International Conference on Next Generation Computer and Information Technology (NGCIT) (pp. 7-10). IEEE, 2015.

[Reiter-Haas et al., 2021]: Reiter-Haas M, Parada-Cabaleiro E, Schedl M, Motamedi E, Tkalcic M, Lex E. *Predicting Music Relistening Behavior Using the ACT-R Framework*. Accepted as LBR paper at RecSys’21. arXiv preprint arXiv:2108.02138. 2021

JOHANNES KEPLER UNIVERSITY LINZ

TU Graz

References

[Rich, 1979]: Rich E. User modeling via stereotypes. Cognitive science, 3(4):329-54, 1979.

[Russell, 1980]: Russell, J.A. *A Circumplex Model of Affect*, Journal of Personality and Social Psychology, 39(6):1161-1178, 1980.

[Schedl et al., 2018]: Schedl, M., Gómez, E., Trent, E.S., Tkalčič, M., Eghbal-Zadeh, H., Martorell, A. *On the Interrelation between Listener Characteristics and the Perception of Emotions in Classical Orchestra Music*, IEEE Transactions on Affective Computing, 9(4):507-525, 2018.

[Scherer, 2005]: Scherer, K. *What are emotions? And how can they be measured?.* Social Science Information, 44(4):693-727, 2005.

[Schnabel et al., 2016]: Schnabel T, Bennett PN, Dumais ST, Joachims T. *Using shortlists to support decision making and improve recommender system performance.* In Proceedings of the 25th International Conference on World Wide Web 2016 Apr 11 (pp. 987-997).

[Shani and Gunawardana, 2011] Shani, G. and A. Gunawardana. 2011. *Evaluating recommendation systems*. In: Recommender systems handbook. Springer. 257–297.

[Wu et al., 2018]: Wu, W., Chen, L., Zhao, Y. *Personalizing recommendation diversity based on user personality*, User Modeling and User-Adapted Interaction 28(3):237-276, 2018.

[Yang and Huang, 2019]: Yang, H.-C., Huang, Z.-R. *Mining personality traits from social messages for game recommender systems*, Knowledge-Based Systems 165:157-168, 2019.