Group Recommender Systems

Judith Masthoff

University of Aberdeen

Scotland, UK

1495 UNIVERSITY OF ABERDEEN

Contents

- Introduction to group recommender systems
- Recommending sequences
- Modelling satisfaction
- Incorporating group attributes
- Explaining group recommendations
- Evaluating group recommender systems

Introduction to Group Recommender Systems

Introduction

- Group Recommender Systems: how to adapt to a group of people rather than to an individual

I know individual ratings of Peter, Jane, and Mary. What to recommend to the group?

How do groups make decisions?

- Split yourself into groups of 4-6 people
- Decide what your group would listen to
  if time for 1 song, for 2 songs, for 3 songs, for 4 songs

A

B

C

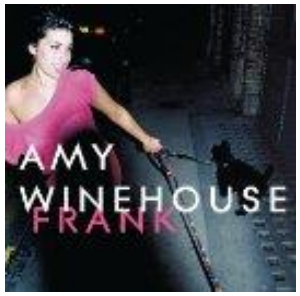
D

E

F

G

H



I

J

K

What would you recommend?

	A	B	C	D	E	F	G	H	I	J
Peter	10	4	3	6	10	9	6	8	10	8
Jane	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6

Average Strategy

Use average of the individual ratings

	A	B	C	D	E	F	G	H	I	J
Peter	10	4	3	6	10	9	6	8	10	8
Jane	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
	7	6	4.3	7.3	8.7	8.7	5.7	7.7	6.7	7.3

Group list: (E, F), H, (D, J), A, I, B, G, C

Least misery strategy

Use minimum of the individual ratings

	A	B	C	D	E	F	G	H	I	J
Peter	10	4	3	6	10	9	6	8	10	8
Jane	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
	1	4	2	6	7	8	5	6	3	6

Group list: F, E, (H, J, D), G, B, I, C, A

Average Without Misery Strategy

Use average of the individual ratings, excluding items rated below threshold

	A	B	C	D	E	F	G	H	I	J
Peter	10	4	**3**	6	10	9	6	8	10	8
Jane	**1**	9	8	9	7	9	6	9	**3**	8
Mary	10	5	**2**	7	9	8	5	6	7	6
		6		7.3	8.7	8.7	5.7	7.7		7.3

Group list: (E, F), H, (D, J), B, G

Many strategies exist

[Masthoff, 2004]

- Average
- Least misery
- Average without misery
- Multiplicative
- Plurality Voting
- Borda count
- Copeland rule
- Approval voting
- Most pleasure
- Fairness
- Most respected person

- Graph-based ranking [Kim et al, 2013]
- Spearman footrule rank [Baltrunas et al, 2010]
- Nash equilibrium [Carvalho & Macedo, 2013]
- Purity [Salamó et al, 2012]
- Completeness [Salamó et al, 2012]
- …….

Exp1: What do people do?

I know individual ratings of Peter, Mary, and Jane. What to recommend to the group? If time to watch 1-2-3-4-5-6-7 clips…

Why?

Compare what people do with what strategies do

Exp1: Results

- Participants do ‘use’ some of the strategies
- Care about Misery, Fairness, Preventing starvation

Exp2: What do people like?

You know the individual ratings of you and your two friends. I have decided to show you the following sequence. How satisfied would you be? And your friends?

Why?

Which strategy does best?
Which prediction function does best?

Exp2: Results

- Multiplicative strategy performs best (FEHJDI is the only sequence that has ratings ≥ 4 for all participants for all individuals)

- Prediction functions: Some evidence of normalization, Misery taken into account, Quadratic is better than linear

Groups matter

- Group size

- Homogeneity in opinions in group

- And many other attributes (*discussed later*)

Domains matter

Music

Restaurants

News

Tourist attractions

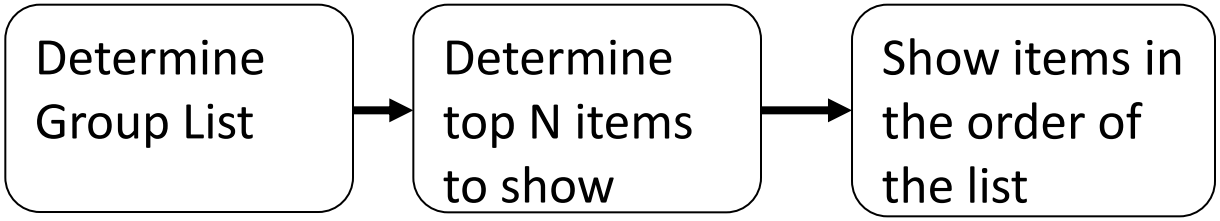
Why do they matter?

Recommending Sequences

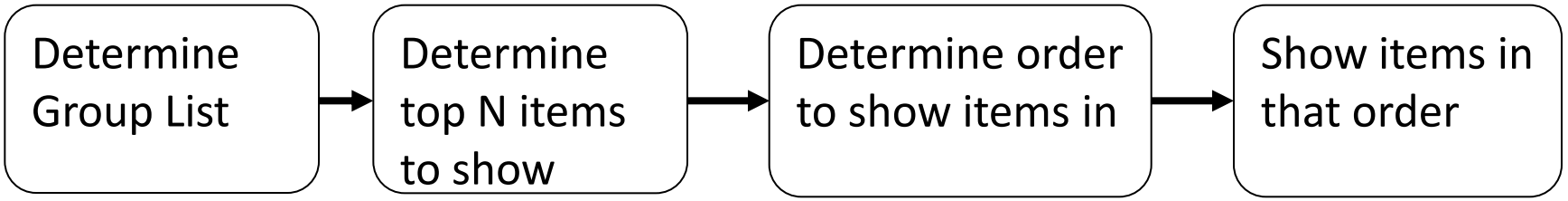
Why sequences?

- Sequences for groups are a lot more interesting than individual items
- With a sequence, it is harder to please everybody
- Fairness has a larger role
- Example domains:
  tourist attractions, music in shop, TV news

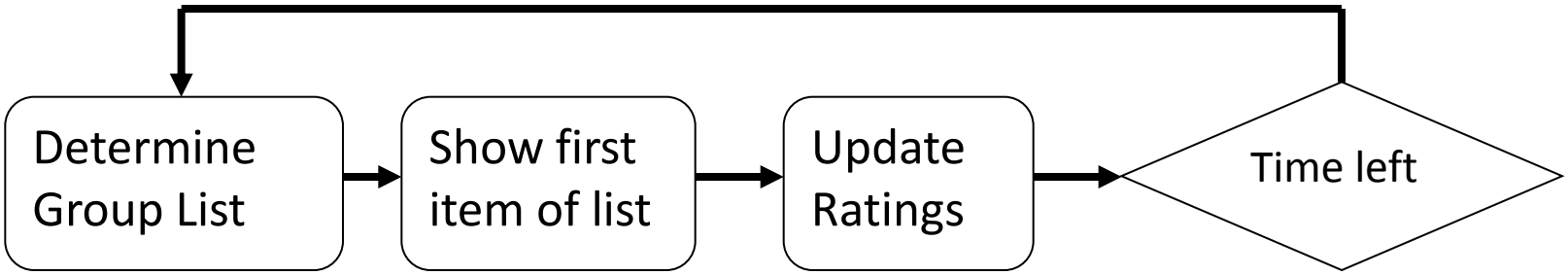
How to deal with order?



But: mood consistency, strong ending, narrative flow,..



But: given all this, perhaps other items are more suitable..



Exp3: Effect of mood, topic

[Insert name of your favorite sport's club] wins important game,
Fleet of limos for Jennifer Lopez 100-metre trip,
Heart disease could be halved, Is there room for God in Europe?,
Earthquake hits Bulgaria, UK fire strike continues,
Main three Bulgarian players injured after Bulgaria-Spain football match

How much would you want to watch these 7 news items?
How would they make you feel?

The first item on the news is "England football team has to play Bulgaria". Rate interest, resulting mood.

Rate interest in 7 news items again

Exp3: Results

- Mood can influence ratings
- Topical relatedness can influence ratings
- Effect of topical relatedness can depend on rating for first item (if interested then more likely to increase)
- Importance dimension

Domain specific aspects of sequences

For example, in tourist guide domain:

- Mutually exclusive / hard to combine items
- Physical proximity
- Diversity concerns

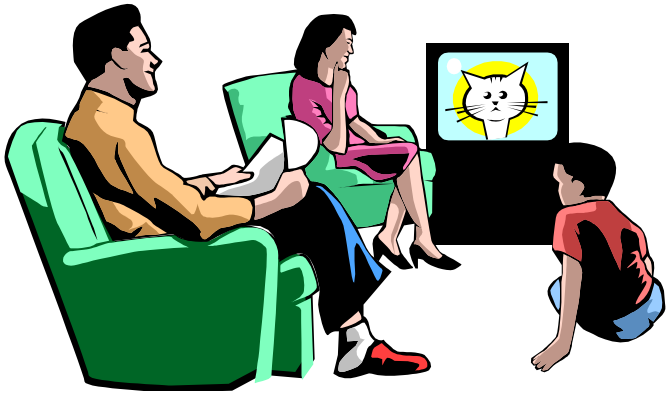
In news domain:

- Novelty concerns
- Topical relatedness

How about music?

Modelling Satisfaction

Why model satisfaction?



- When adapting to a *group* of people, you cannot give everyone what they like all the time

- But you don't want somebody to get too dissatisfied…

- When adapting a *sequence* to an individual, the order may impact satisfaction

Strategies that use satisfaction

Know how satisfied each user is with the items so far

And their profile

	B	C	H	I	J
	9	8	9	3	8
	5	2	6	7	6
	4	3	8	10	8

Decide which item to present next, trying to please the least satisfied user

Strongly support grumpiest strategy

- Pick item most liked by the least satisfied person
- If multiple items most liked, use existing strategy (e.g. Multiplicative) to choose between them

Problem: Suppose Mary least satisfied so far

	A	B	D	E
Peter	10	4	6	10
Jane	1	9	9	7
Mary	10	5	7	9

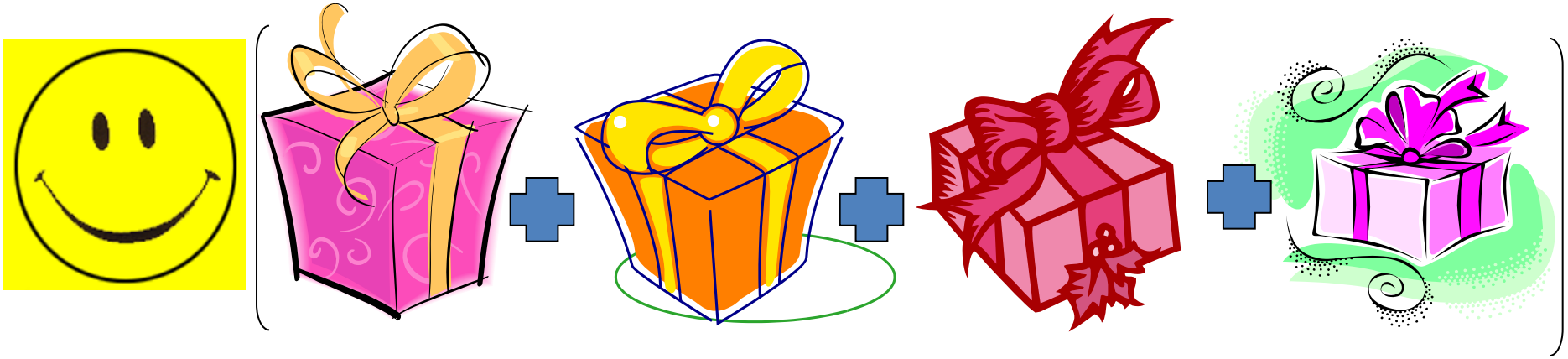
- Strategy would pick A
- Very bad for Jane
- Better to show E?

Alternative strategies using satisfaction

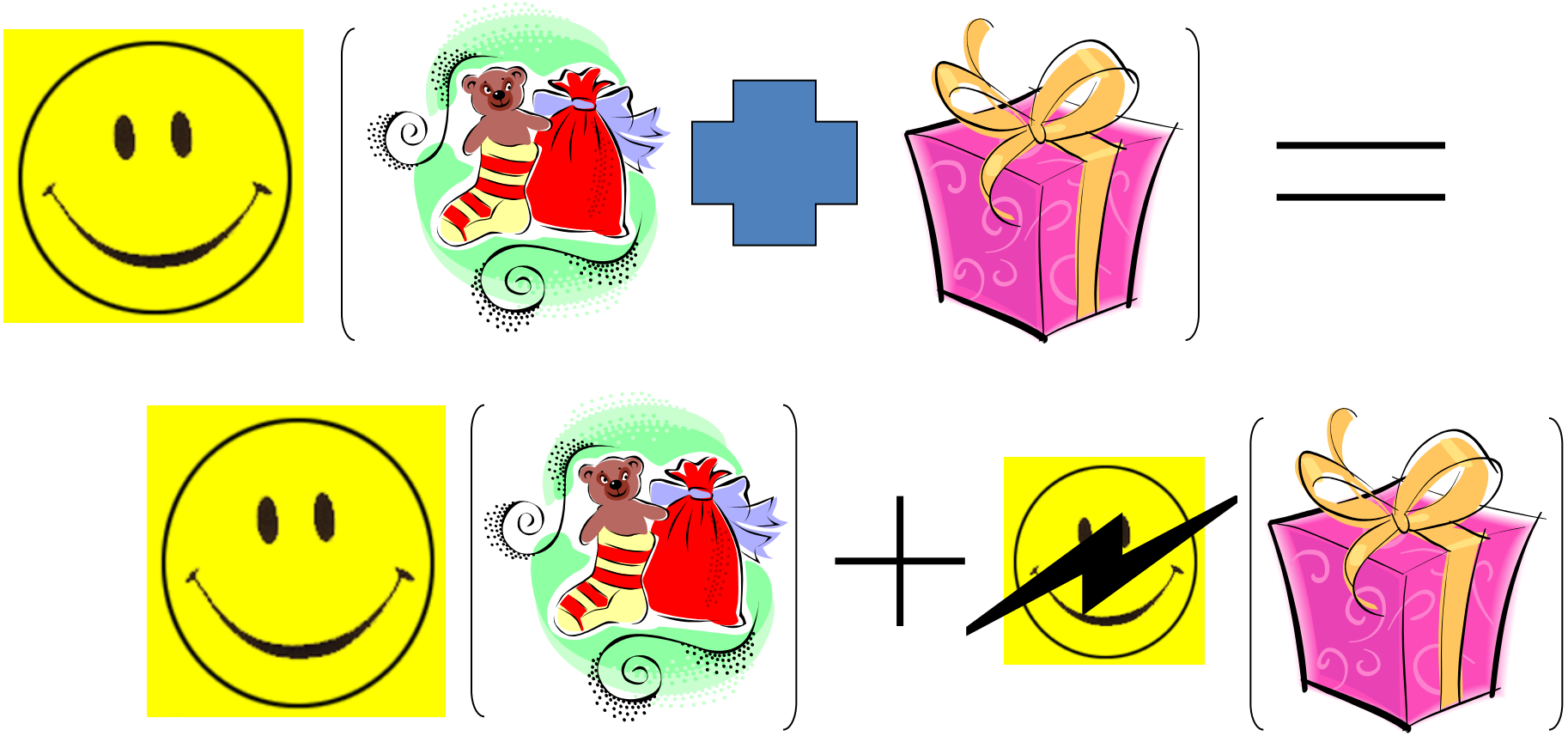
- Weakly support grumpiest strategy
  - Consider all items quite liked (say rating>7) by the least satisfied person
  - Use existing strategy to choose between them
- Strategies using weights
  - Assign weights to users depending on satisfaction
  - Use weighted form of existing strategy, e.g. weighted Average
  - Cannot be done with some strategies, such as Least Misery

Challenge is to model satisfaction

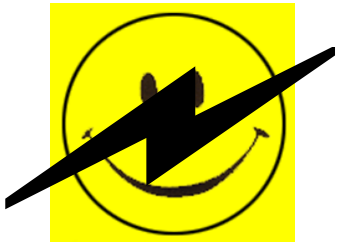
- Would like a model that *predicts* satisfaction of an individual user after a sequence of items



Basic model



Impact



=

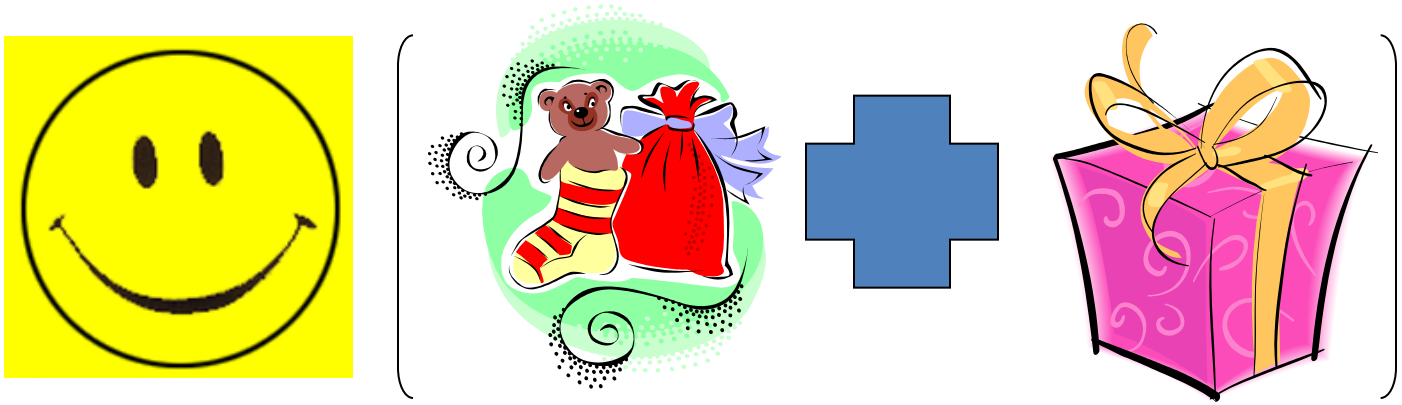
Quadratic(

Rebalanced(

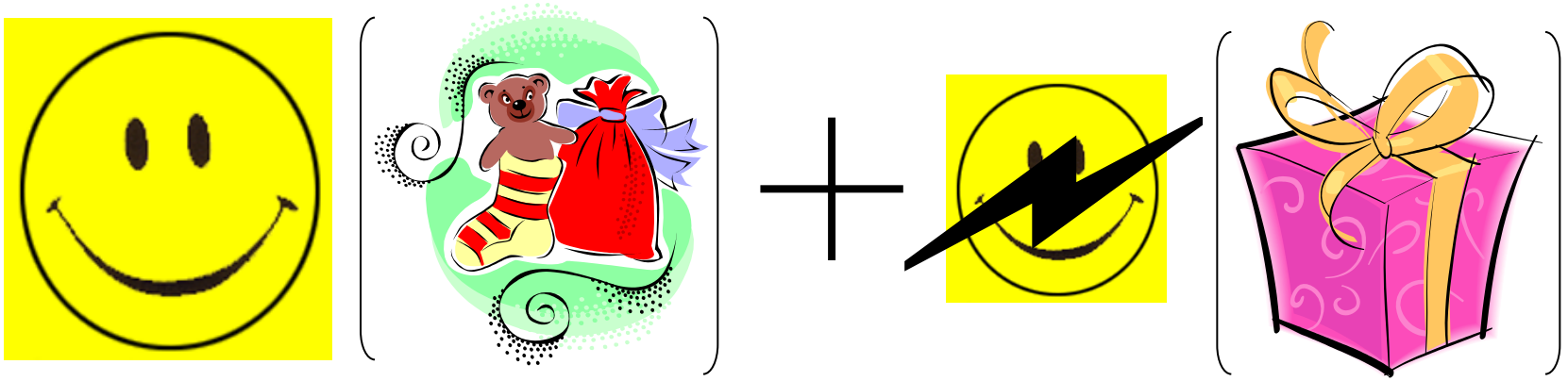
Normalized(

Rating( )))

Variant 1: Satisfaction decreases over time

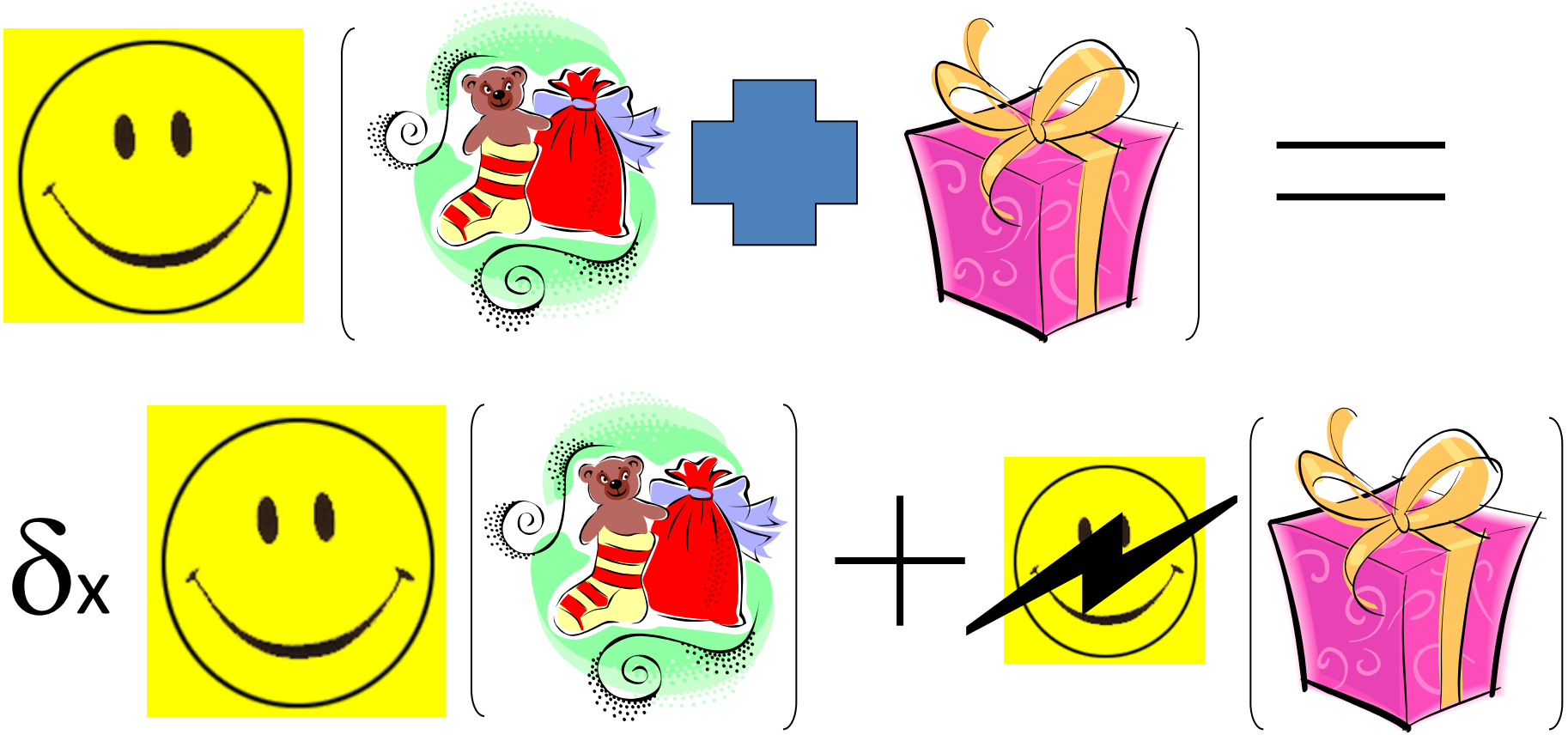


δx

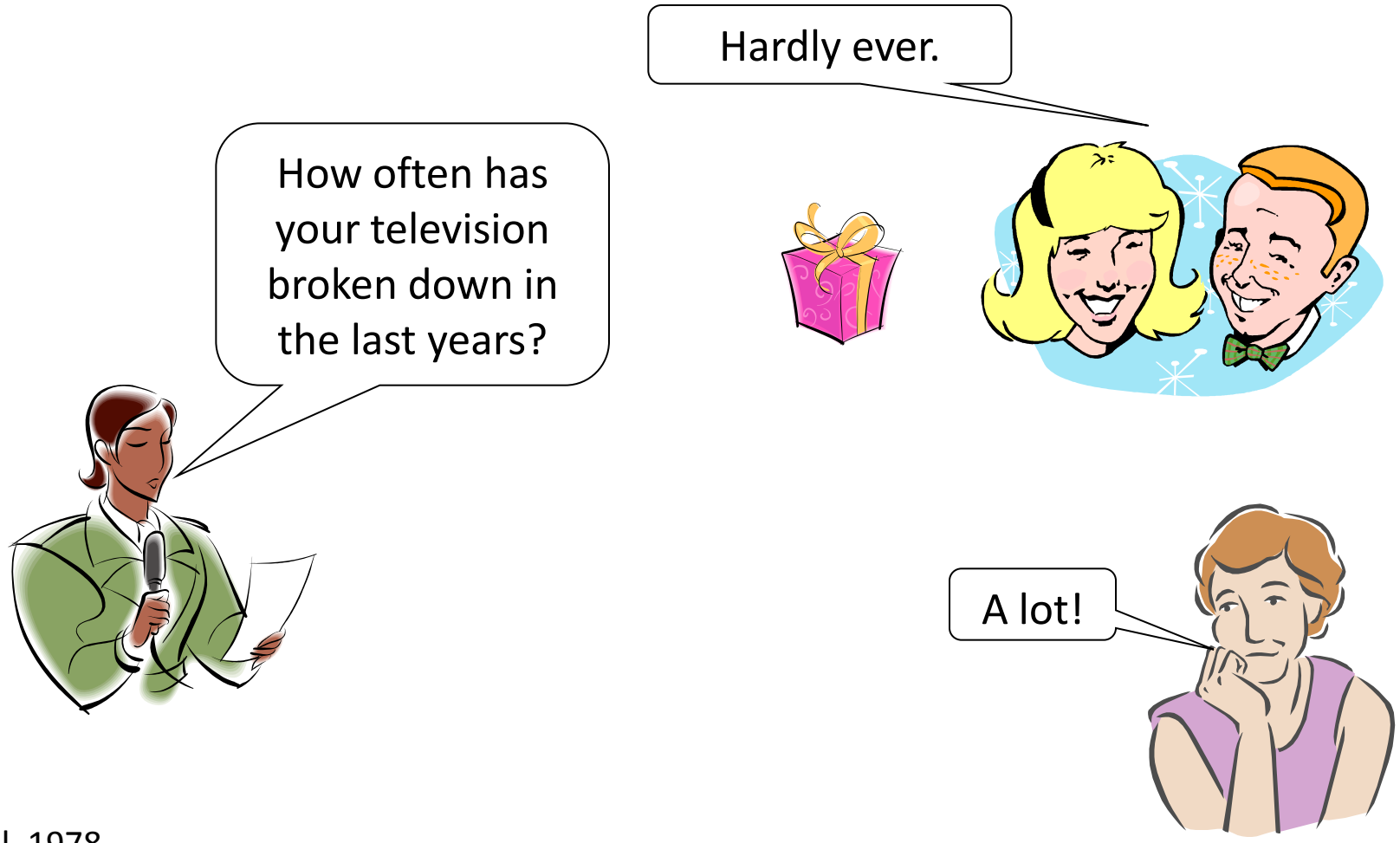


$0 \leq \delta \leq 1$ δ=0: No memory δ =1: Perfect memory

Variant 2: Satisfaction is bounded

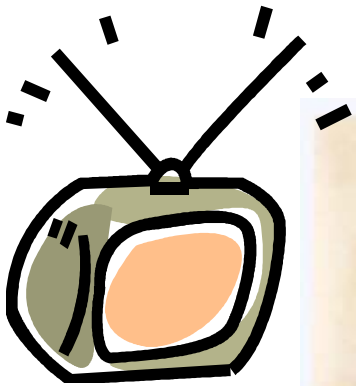


Mood impacts evaluative judgement



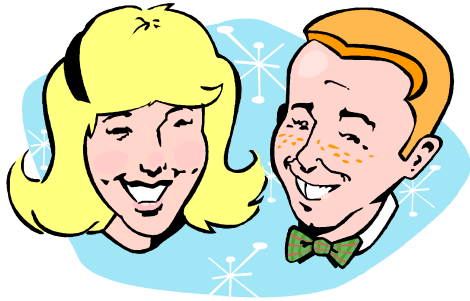
Isen et al, 1978

Mood impacts evaluative judgement



How much have you been persuaded?

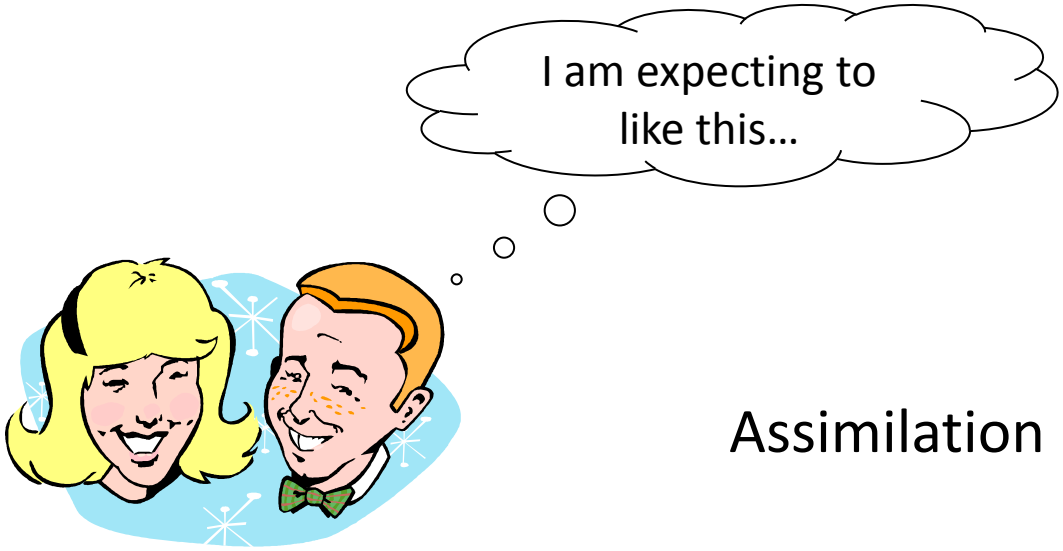
A lot.



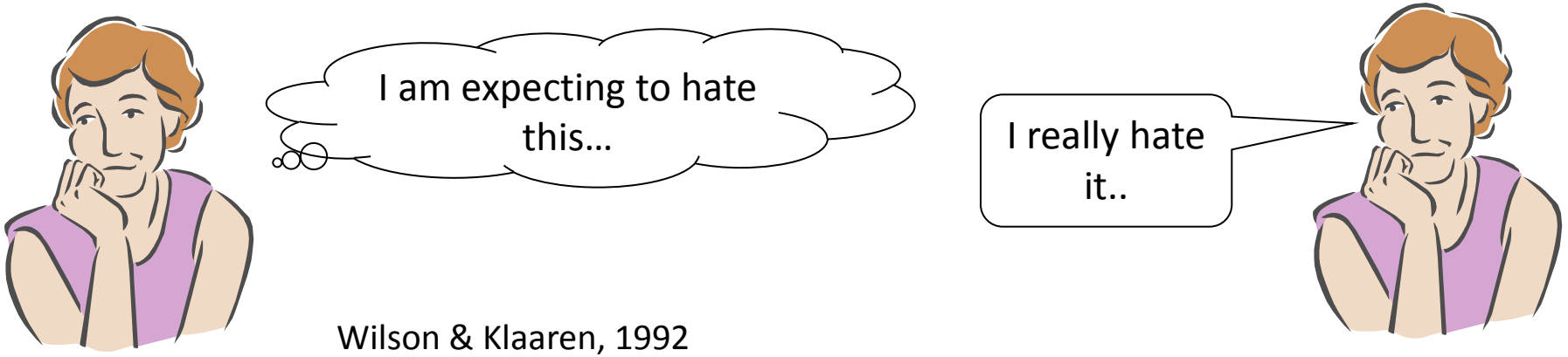
A little.

Mackie & Worth, 1989

Affective forecasting can change actual emotional experience

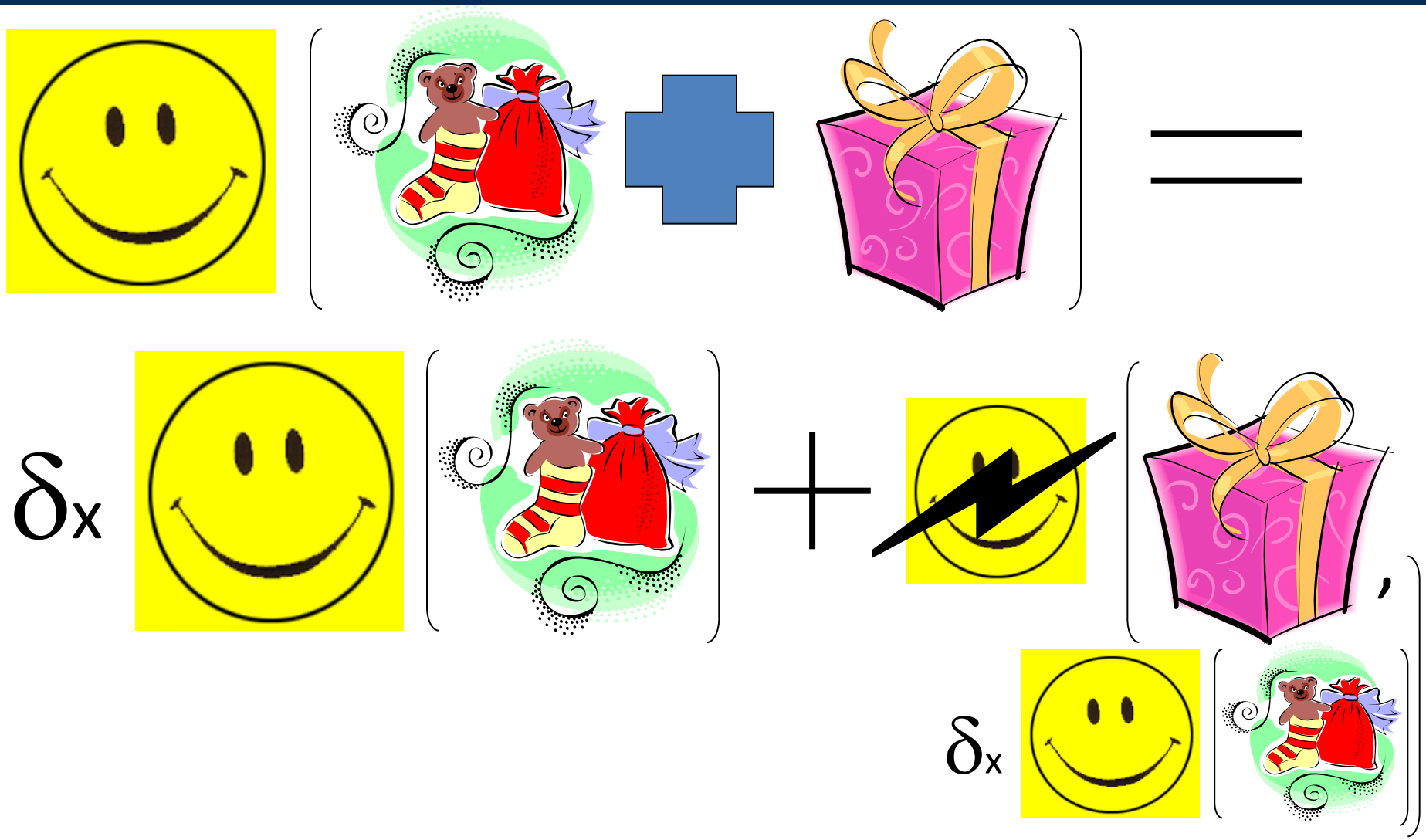


Assimilation

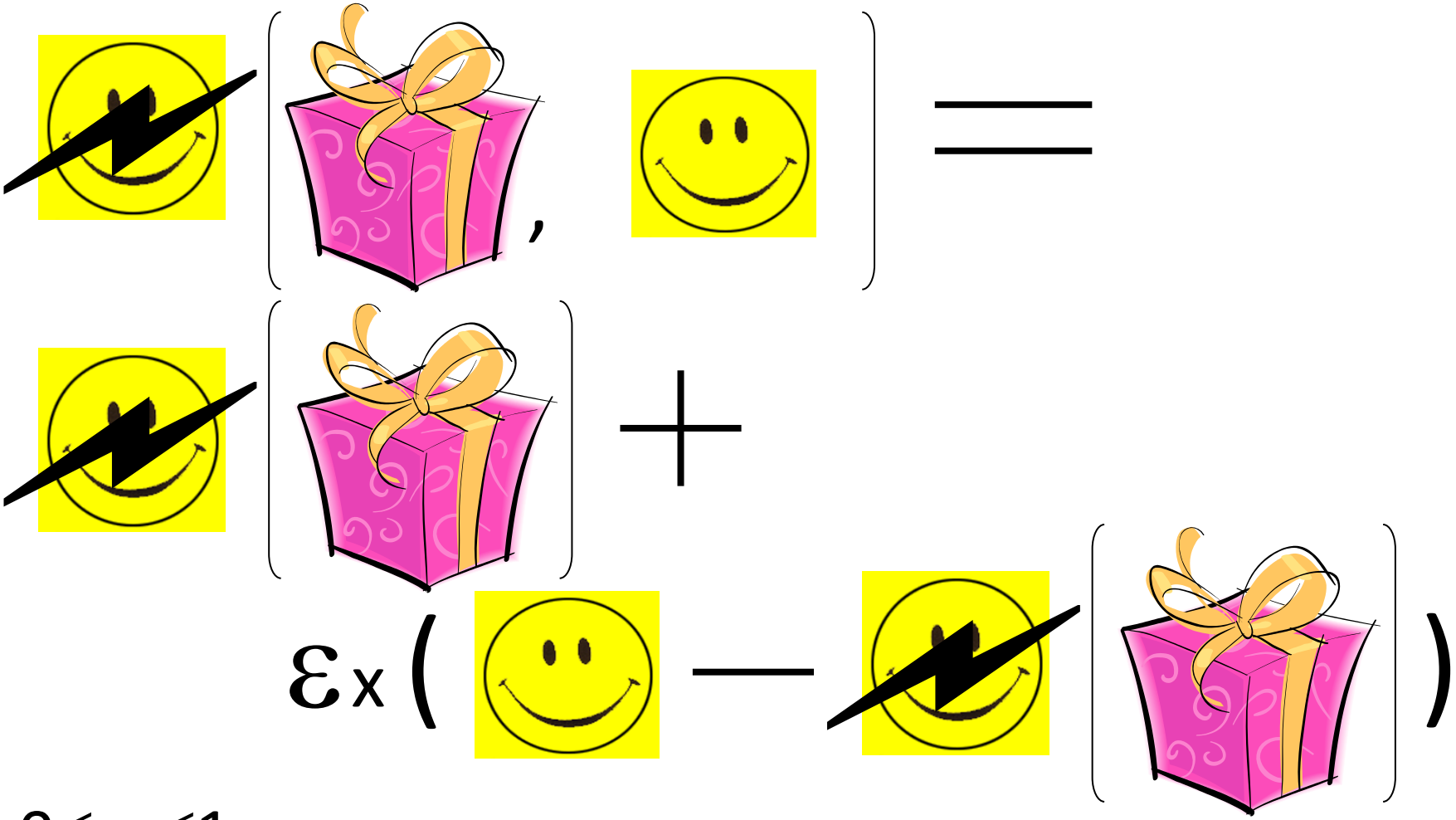


Wilson & Klaaren, 1992

Variant 3: Impact depends on mood

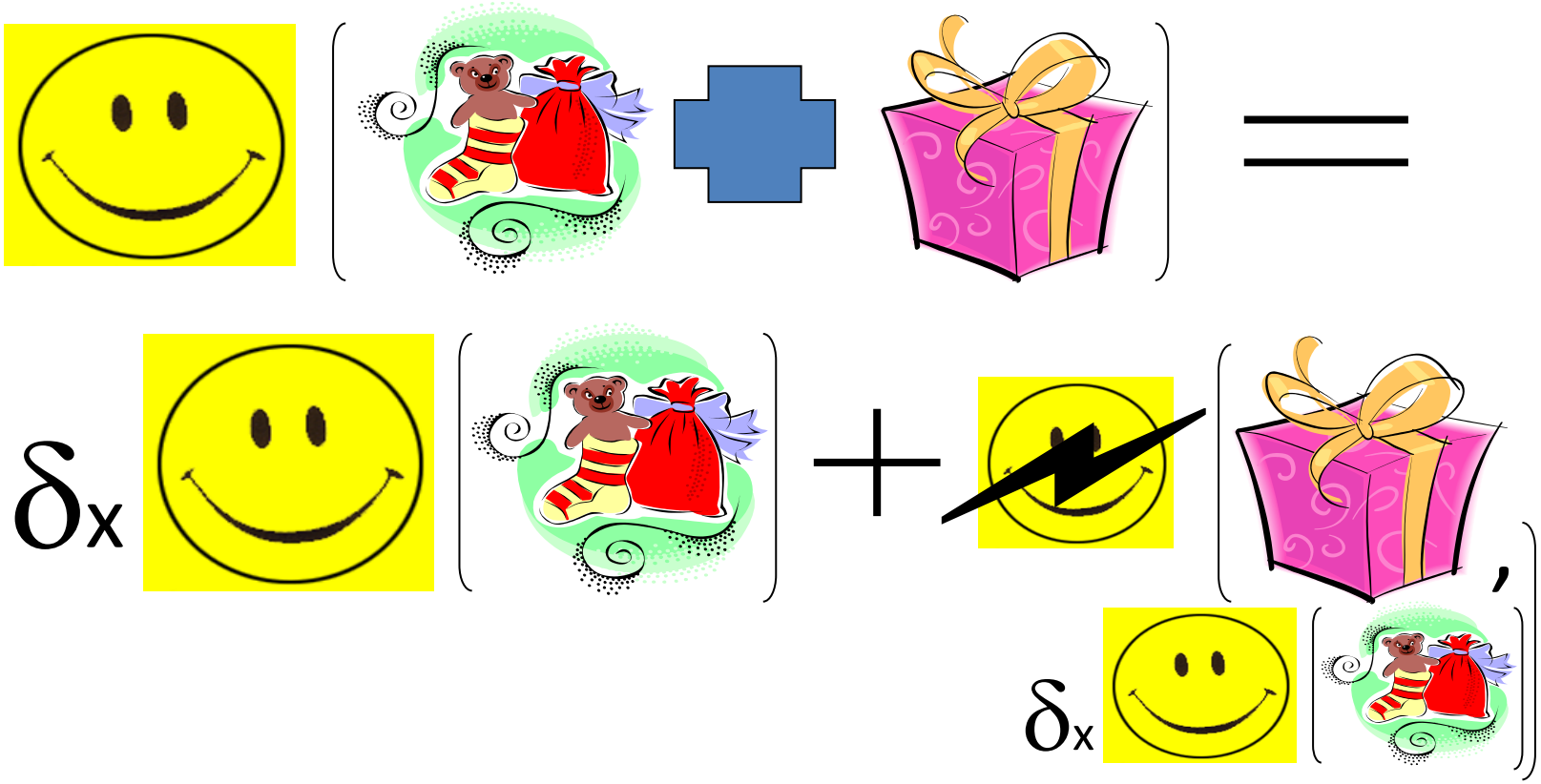


Impact depends on mood



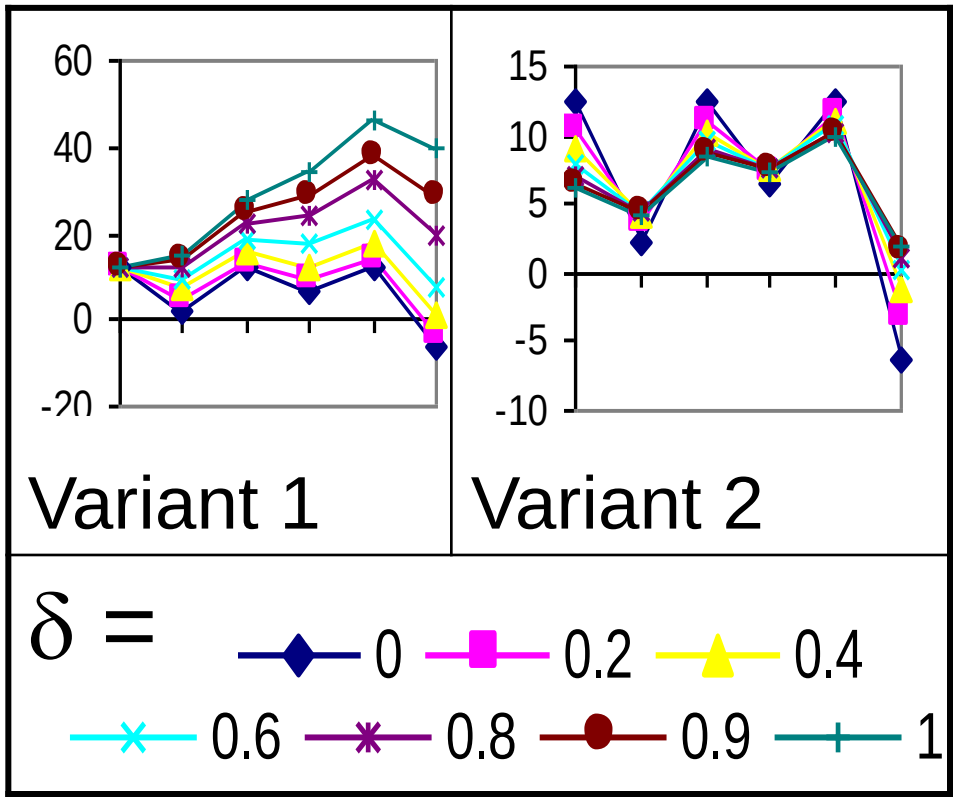
$0 \leq \varepsilon \leq 1$ ε=0: No impact mood ε =1: Mood determines all

Variant 4: Combination of Variants 2 and 3



Evaluation by simulation

- Models predict satisfaction of Peter, Jane, Mary with a sequence, given δ, ε
- Compare to human predictions (from Exp2)
- Some strategies bad for any δ
- δ should be high (>0.5), ε low
- Multiplicative best for high δ



Jane, sequence from Multiplicative strategy

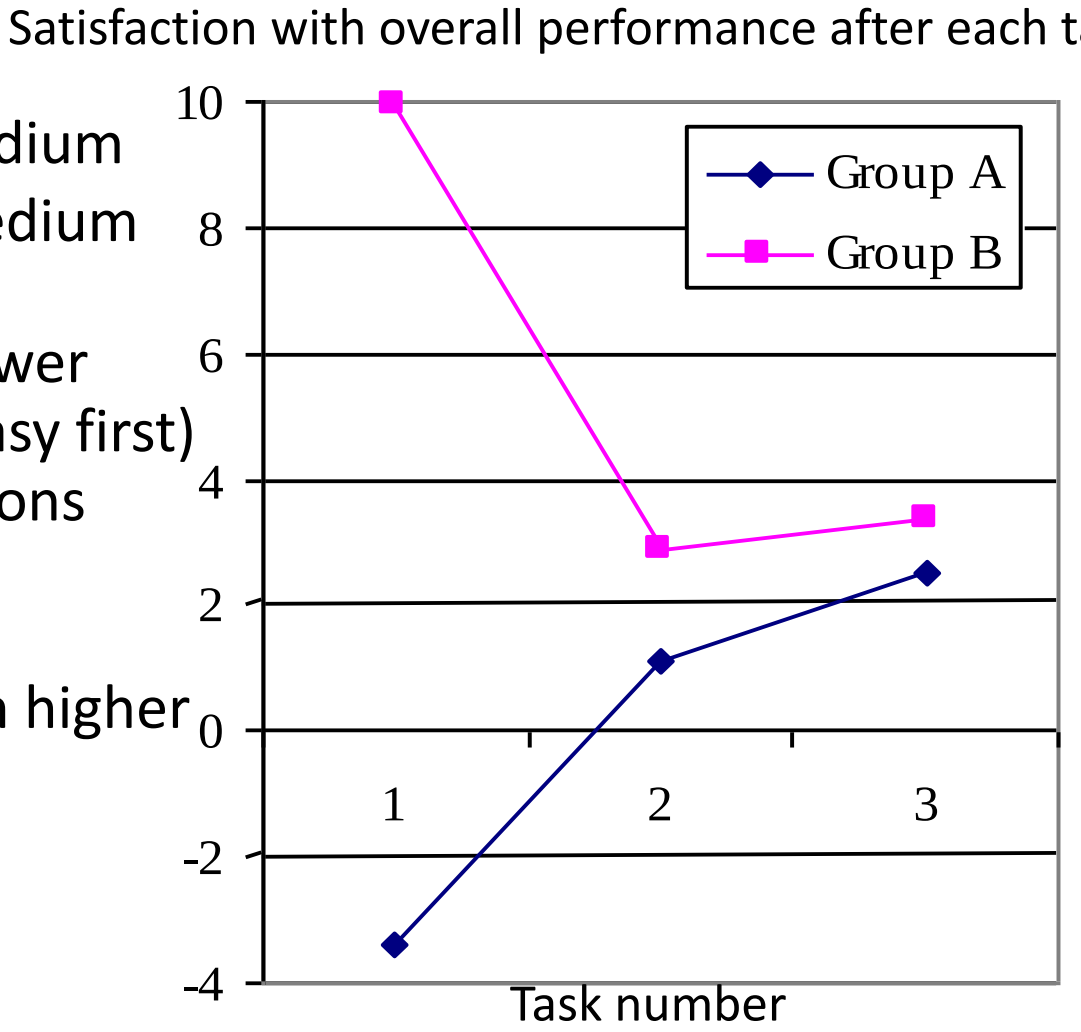
Evaluation by study (Exp4)

Group A: Hard – Easy –Medium
Group B: Easy – Hard – Medium

Variants 1 and 2 predict lower satisfaction for group B (easy first) after 2 tasks, due to emotions wearing off.

Assimilation could result in higher satisfaction for B.

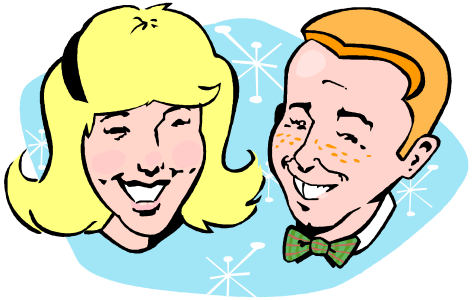
Variant 4 seems best



Emotional contagion

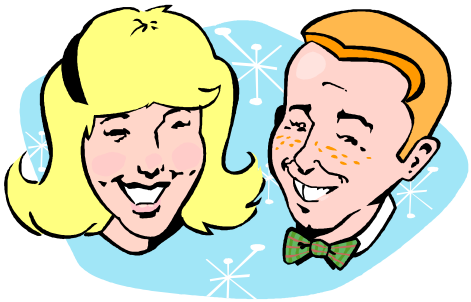
Totterdell et al, 1998; Barsade 2002; Bartel & Saavedra, 2000

Emotional contagion



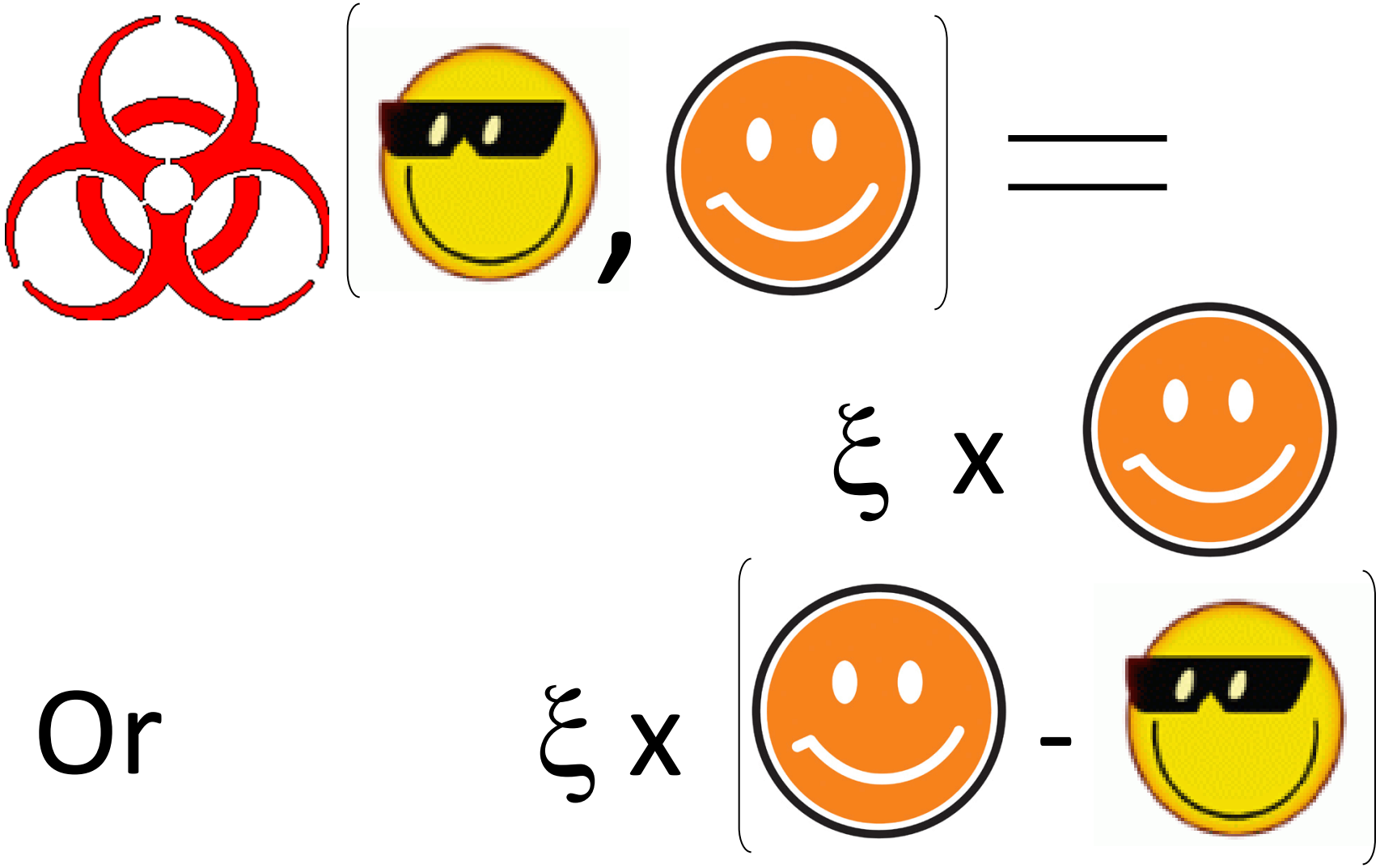
Totterdell et al, 1998; Barsade 2002; Bartel & Saavedra, 2000

Emotional contagion



Totterdell et al, 1998; Barsade 2002; Bartel & Saavedra, 2000

Emotional contagion



Susceptibility of emotional contagion

User Dependent

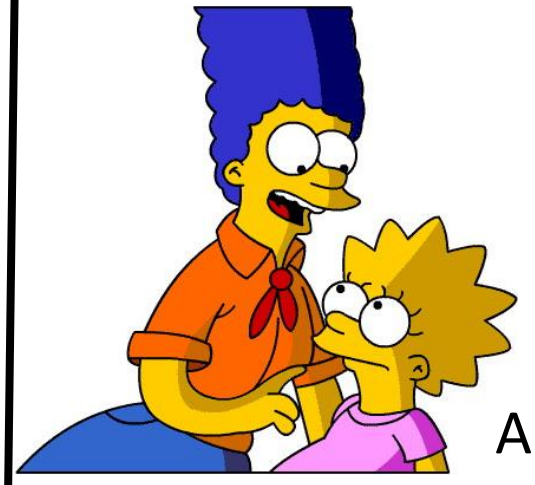
Laird et al, 1994

So, ξ should be user dependent

Types of relationship

"Somebody you share everything with, e.g. a best friend"

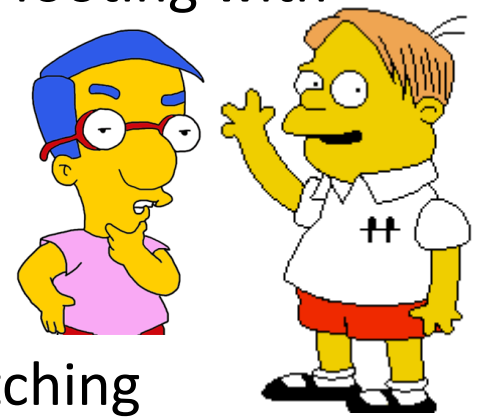
Communal Sharing



"Somebody you respect highly"

Authority Ranking

"Somebody you are on equal footing with"



Equality Matching

"Somebody you do deals with / compete with"

Market Pricing

Fiske, 1992; Haslam, 1994

Susceptibility and types of relationship

When calculating of by

Need to take account of 's susceptibility

And the relationship between and

ξ $= \sigma$ $\times \rho$

Exp5: Emotional contagion

- Susceptibility to emotional contagion measured using existing scale (Doherty, 1997)

- "Think of somebody [relationship type].
  Assume you and this person are watching TV together. You are enjoying the program a little. How would it make you feel to know that the other person is [enjoying it greatly / really hating it]? My enjoyment would…"

- We expect Authority Ranking and Communal Sharing to have more contagion.
- Will Market Pricing have negative ξ?

Exp5: Results

- Contagion happens
- More contagion for Authority Ranking and Communal Sharing relationships
- No difference between negative and positive contagion
- Susceptibility only seemed to make a difference for Communal Sharing relationships

Exp6: Opinion shifting

- Start made in Francesco Barile's work (2017 paper in UMAP workshop)
- Considered:
  - Tie strength (Weak, Intermediate, Strong)
  - Relationship type (Like, Indifferent, Dislike)
  - Closeness of initial ratings (Small, Large)
- Some evidence of:
  - positive opinion shifts when intial ratings far apart
  - negative shifts when initial ratings close but disliking relationship

Incorporating Group Attributes

What attributes matter?

- Remember the task I gave you at the start

- What attributes of the people in your group influenced the decision making (excluding their opinions on the music items)?

- Or could have influenced the decision making if they had been present in your group

Attributes of group members

- Demographics and roles [Ardissono et al, 2002; Senot et al, 2010]
- Personality
  - Propensity to emotional contagion
  - Agreeableness?
  - Assertiveness and cooperativeness [Quijano-Sanchez et al, 2013]

- Expertise [Berkovsky & Freyne; Gatrell et al, 2010, Herr et al, 2012]
- Personal impact/cognitive centrality [Liu et al, 2012; Herr et al, 2012]

Typically used to vary the weights of group members

Attributes of the group as a whole

- Relationship strength

  Gatrell et al (2010) propose:

  Most Pleasure for strong relations,
  Least Misery for weak, Average for intermediate

- Relationship type:

  Wang et al (2010) distinguish:

  – Positionally homogeneous vs heterogenous groups
  – Tightly coupled versus loosly coupled groups

Typically used to select a different strategy

Attributes of pairs in the group

- Relationship strength/social trust
  [Quijano-Sanchez et al, 2013]
- Personal impact
  [Liu et al, 2012; Ye et al, 2012, Ioannidis et al, 2013]

Typically used to adjust the ratings of an individual in light of the ratings of the other person in the pair.

Explaining Group Recommendations

Aim of explanations in any rec sys

Improve:

- Trust
- Effectiveness
- Persuasiveness
- Efficiency
- Transparency
- Scrutability
- Satisfaction

[Tintarev & Masthoff]

And these aims can conflict

Explanations may be even more important in group recommender systems

Which aims?

Sequence issue

- More work is needed on explaining sequences, particularly sequences that contain items the user will not like

Privacy issue

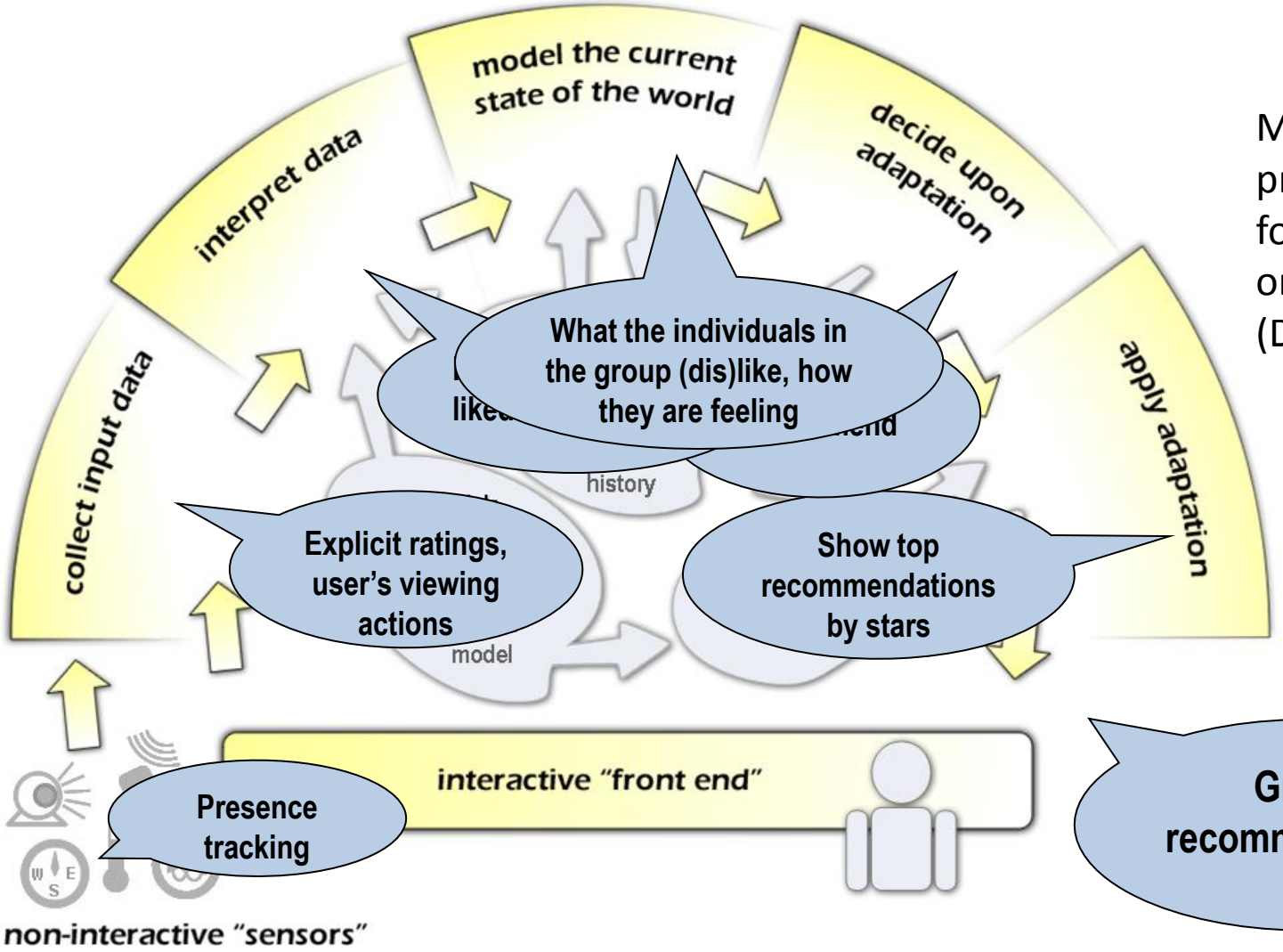
- Many aims may require explanations that reflect on other group members….
- How to do this without disclosing sensitive information?
- Even general statements such as “this item was not chosen as it was hated by somebody in your group” may cause problems

Challenge 1: Evaluation and Metrics

Slicing and Dicing

- Want to know *why* a group recommender system works / does not work
- Slicing: Layered evaluation (Paramythis et al, 2010)
  - Break adaptation process down into its constituents ("layers")
  - Evaluate layers separately
- Dicing
  - Break system down into separate functionalities (e.g. provide recommendations, explain recommendations)
  - Evaluate functionalities separately

Layered evaluation



Most of this presentation focussed on one layer (DA or UM)

How to evaluate how good a strategy is?

- What does it mean for a group recommender strategy to be good?
- For the *group* to be satisfied?
- But how do you measure the satisfaction of a group?

Metrics (1)

- Utility for the group

This is what most researchers do, they take the average of the individuals' ratings (or average of a comparison of rankings of items).

What is the problem with this?

Metrics (2)

- Whether all individuals exceeded a minimum level of satisfaction

When? After a sequence of items? At each point in the sequence?

What is the problem with this?

Metrics (3)

Extent to which group members

- Think it is fair?
- Think it is best for the group?
- Accept the recommendation for the group?
- Do not exhibit negative emotions?

With or without having seen the options and individual preferences?

What is the problem with this?

Metrics (4)

Extent to which independent observers

- Think it is fair?
- Think it is good / best for the group?

Having seen the options and individual preferences

Having seen the reactions of the group members?

What is the problem with this?

Metrics (5)

Extent to which the recommendations correspond to

- What groups would decide themselves?
- What human facilitators would decide for the group?

What is the problem with this?

How to obtain groups for evaluation?

- Artificially construct groups
  - From existing data about individuals
  - Or: of invented individuals
- Use real groups:
  - But without group data
  - Or: to generate group data
    (e.g. What the group decides to watch when together)
  - Or: to provide recommendations and measure effect

Conclusion

Many challenges

Many opportunities

Some references

- Masthoff, J. (2015). Group recommender systems: aggregation, satisfaction and group attributes. Recommender Systems Handbook, pp743-776
  (See UMUAI articles in 2004 and 2006 for more detail)
- Tintarev, N., Masthoff, J. (2015). Explaining recommendations: Design and evaluation. Recommender Systems Handbook, pp353-382
  (See UMUAI article in 2012 for more detail)
- Paramythis, A., Weibelzahl, S., Masthoff, J. (2010). Layered evaluation of interactive adaptive systems: framework and formative methods. User Modeling and User-Adapted Interaction 20 (5), 383-453

Questions ?