Experiments

Research Methods for Human-Centered Computing

Experiments

Today's goal:

Introduce you to user experiments

Outline:

- What is a user experiment?
- From research questions to hypotheses
- An example

Literature outline

Mapping related work for your project proposal

Literature outline

Step 1: Decide on your research question!

I will give you feedback on the ones you submitted before the weekend

Step 2: Start the outline of your paper

Use the sigchi .docx or LaTeX template

Step 3: Put in the relevant sections

Introduction, related work, methods (proposed), results (expected), discussion, conclusion, references

Literature outline

Step 4: Introduce your research question in the introduction

- Give enough context (don't just put the question)

Step 5: Cite the literature that will motivate your research question in the introduction

- Why is it important to answer this research question?
- Example: "Knijnenburg et al. [5] argue that privacy concerns are an undying issue in social media."
- For now, you can put these citations in a bulleted list (no need to narrativize things)
- Use the correct reference format, though!

Literature outline

Step 6: Outline your related work section

Create sub-headings in which you organize the related work

Example (for a paper about children's privacy online):

- 2.1 Research on the online activities of children
- 2.2 Theories of online privacy
- 2.3 Children's privacy online

Step 7: Cite papers in each subsection

See step 5

Literature outline

Step 8: Cite the most common methods used in similar work in your methods section

> See step 5; e.g. “Knijnenburg et al. [7] create different IoT scenarios and ask participants whether they accept each scenario and to rate the perceived risk, usefulness, expectedness, and appropriateness of each scenario.”

Step 9: Cite papers with a similar argument (but a different setting/method/population than yours) in your discussion section

> e.g. “Knijnenburg et al. [9] also conclude that having fewer privacy options is better, but in a location-sharing context.”

User experiments

What is a user experiment?

User experiments

A scientific method to investigate factors that influence how people interact with systems*

Systems can be anything:

- Software
- Hardware
- Other people
- Organizations
- Policies

What to ask?

"Is my new travel system good?"

Problem...

What does **good** mean?

- Learnability? (e.g. number of errors?)
- Efficiency? (e.g. time to task completion?)
- Usage satisfaction? (e.g. usability scale?)
- Outcome quality? (e.g. survey?)

We need to define **measures**

Better...

"Does the user interface of my travel system score high on this usability scale?"

However...

What does **high** mean?

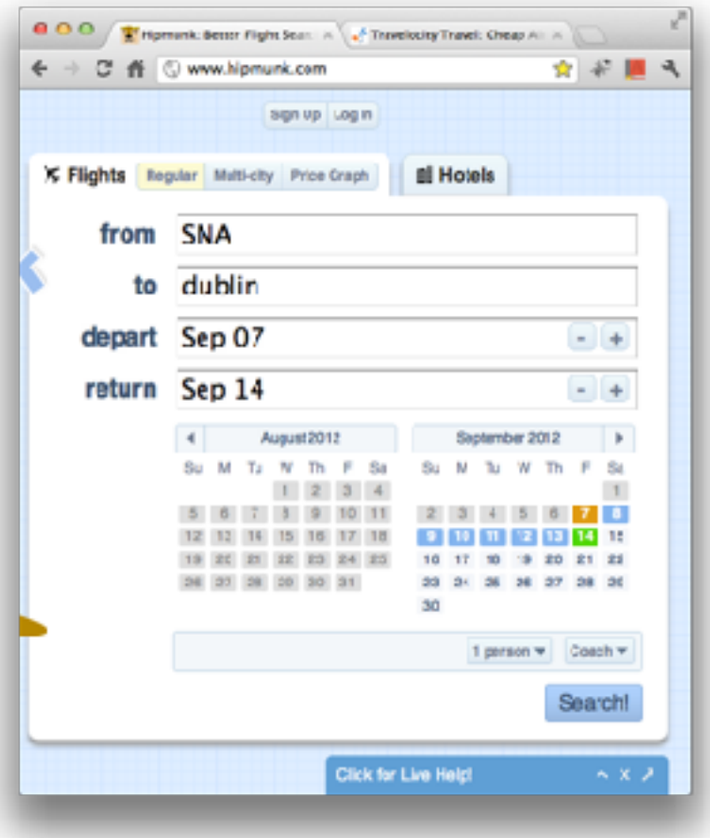
- Is 3.6 out of 5 on a 5-point scale “high”?
- What are 1 and 5?
- What is the difference between 3.6 and 3.7?

We need to **compare** the UI against something

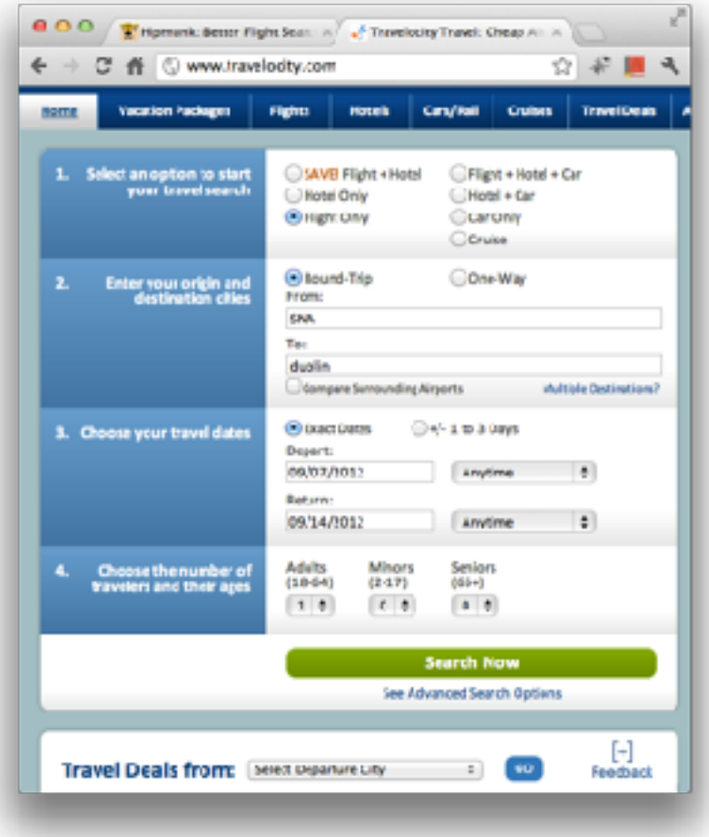
Even better...

“Does the UI of my system score high on this usability scale compared to this other system?”

Testing A vs. B



My new travel system



Travelocity

However...

Say we find that it scores higher on usability... **why** does it?

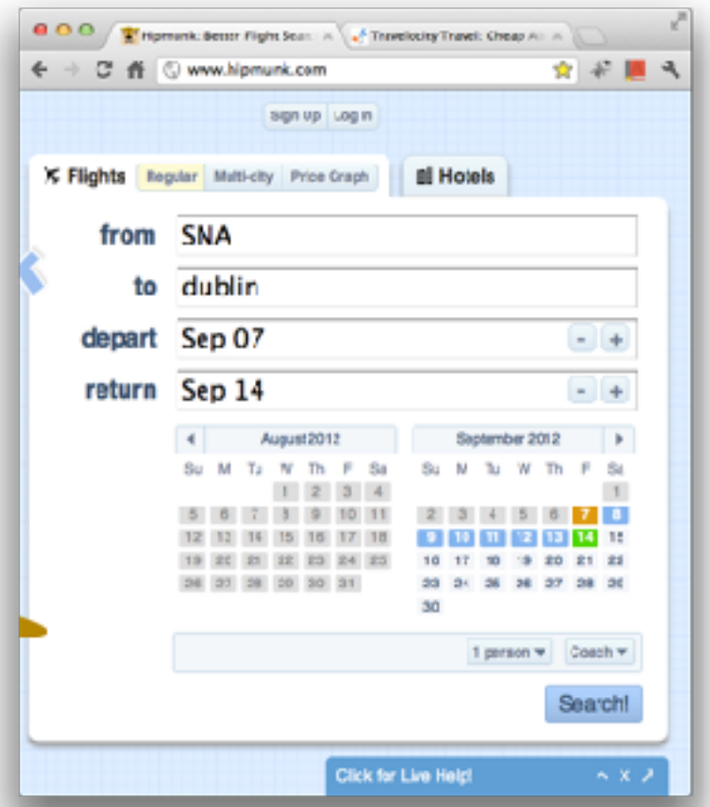
- different date-picker method
- different layout
- different number of options available

Apply the concept of **ceteris paribus** to get rid of confounding variables

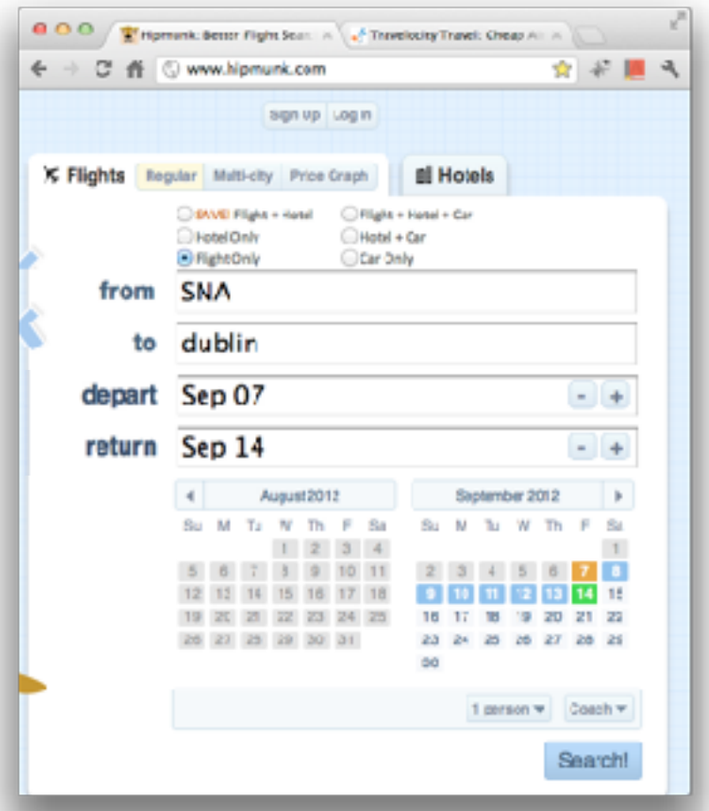
Keep everything the same, except for the thing you want to test (the manipulation)

Any difference can be attributed to the manipulation

Ceteris Paribus



My new travel system



Previous version
(too many options)

How it works

Create multiple versions of your system (intervention and control)

Recruit participants to take part in your study (a sample taken from a population)

You let people use one (or all) of these versions

You measure their behaviors and/or subjective evaluations (outcome)

You statistically evaluate the difference in outcome between intervention and control

Core components

"A user experiment systematically tests how different system aspects (manipulations) influence the users' experience and behavior (observations)."

Manipulations

Manipulations are the things that you believe will "make a difference"

- Also called "independent variables"

In HCC research, our main manipulations are **system aspects**

- We are not testing entire systems, but system aspects

Each manipulation consists of multiple **conditions** (different versions of the system aspect)

- Simplest variant, two conditions: intervention and control

Manipulations

Examples of manipulations:

- Recommendations vs. random items
- Number of recommendations: 5, 10, or 20
- Comic-based privacy policy vs. text-based privacy policy
- Comic length: short, medium, long

Manipulations

You can combine multiple manipulations in a single study!

Type of privacy policy (comic, text) X length of policy (short, medium, long)

In this case, experimental conditions multiply:

Short comic, medium comic, long comic, short text, medium text, long text

Don't go overboard!

Required sample size depends on the number of conditions!

Observations

Observations are the means by which you measure the differences between conditions

Also called “dependent variables” or “outcomes”

In HCC research, our observations are either **objective** or **subjective**

They are **always** quantitative (more on this next week)

Observations

Examples of observations:

Objective:

- Number of clicks
- Privacy knowledge (# correct answers on a privacy quiz)

Subjective

- Perceived privacy protection
- Perceived system effectiveness

Systematic

If you apply the concept of ceteris paribus...

...and you randomly assign participants to conditions...

...then any difference you find can be attributed to the manipulation!

i.e., the difference between the conditions!

The power of user experiments lies in the ability to make such **causal inferences**!

Survey/observation

What is the difference between men and women in Facebook usage satisfaction?

Downsides:

Purely correlational

- No manipulations!
- What causes what?
- Third variable problem

No ceteris paribus

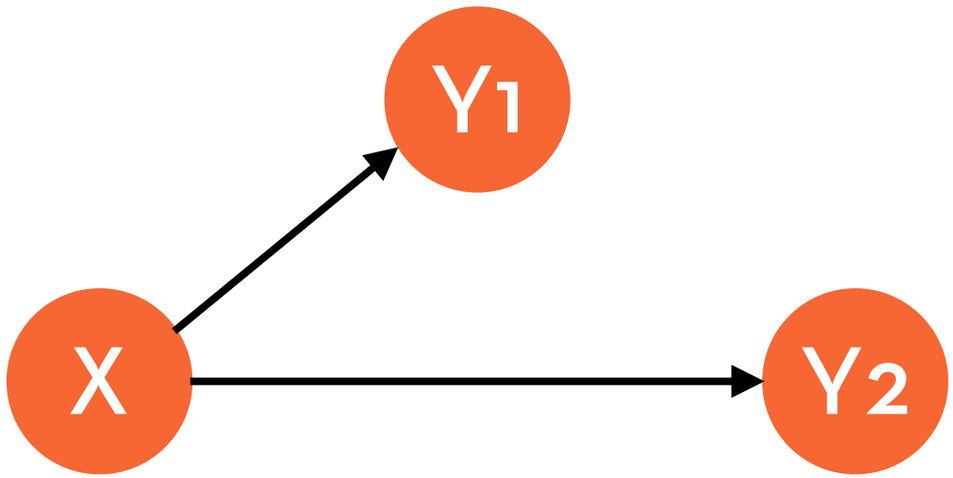
- Hard to get rid of confounding variables

Causal relations

X = comics (vs. text)

Y1 = privacy knowledge
(# correct quiz questions)

Y2 = perceived privacy
protection



Mediation

Manipulation -> perception
-> experience

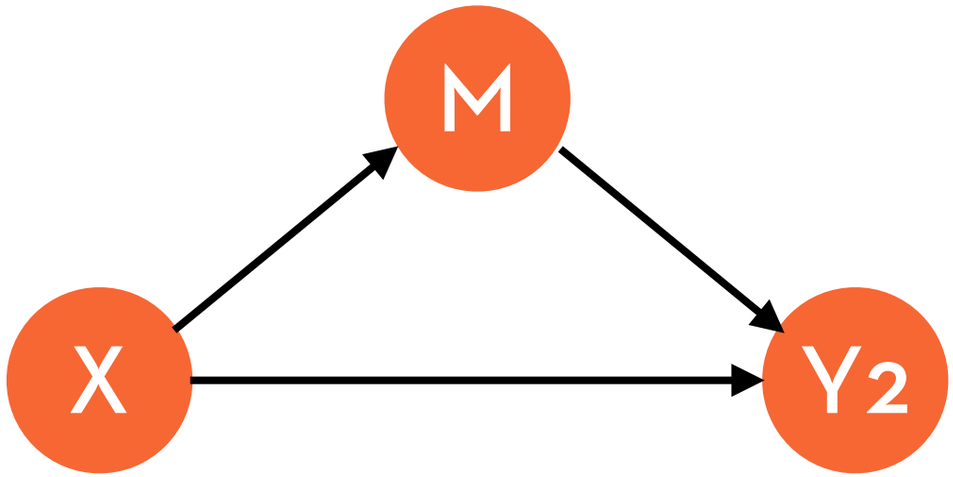
Comics result in higher protection because of higher knowledge

Types of mediation

Partial mediation

Full mediation

Negative mediation



Covariates

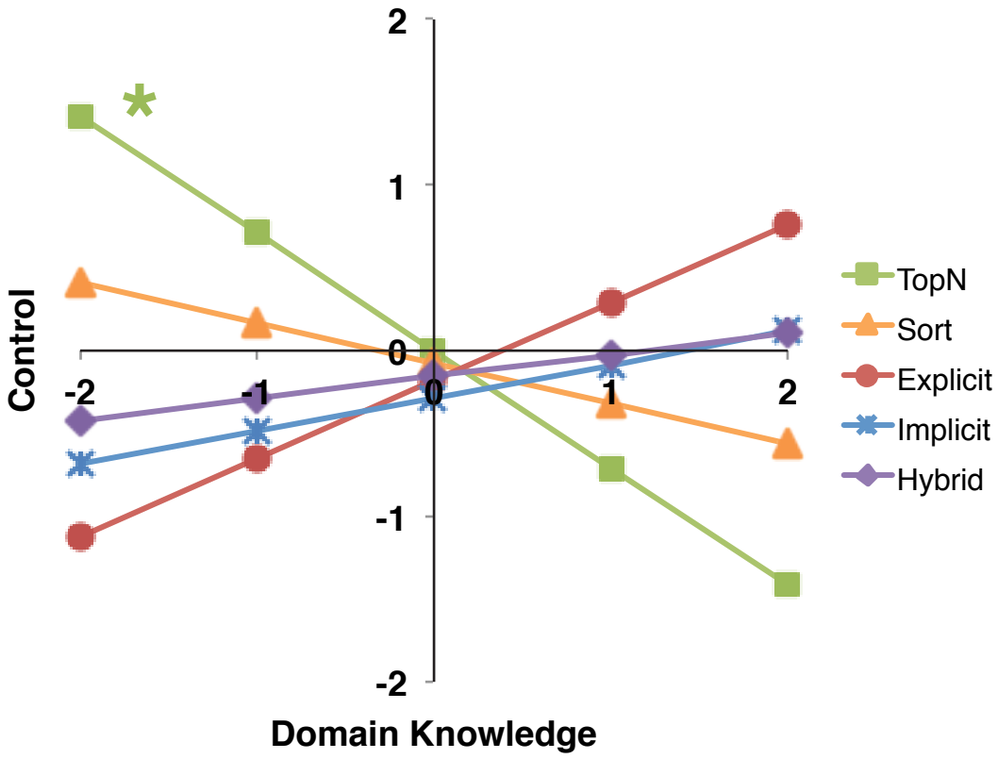
What about participant characteristics?

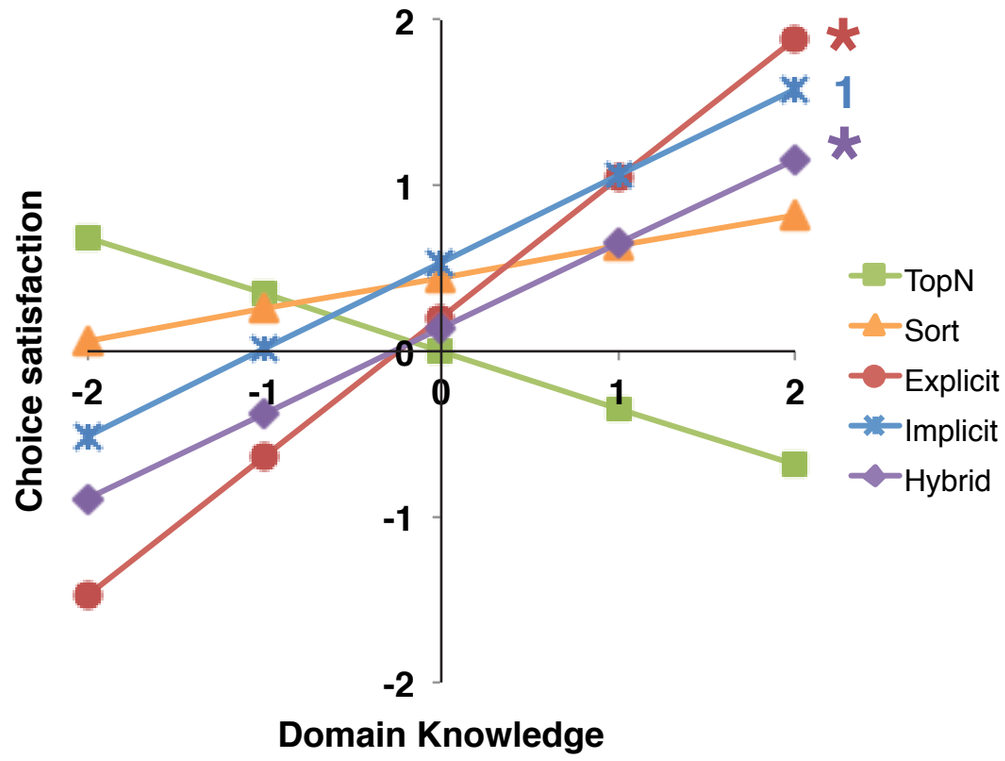
E.g., age, gender, etc.

We are usually interested in them as **covariates**

They change (**moderate**) the effect of the manipulation

Moderation


2
1
0
-1
-2
-2
-1
0
1
2
Control
Domain Knowledge
TopN
Sort
Explicit
Implicit
Hybrid


2
1
0
-1
-2
-2
-1
0
1
2
Choice satisfaction
Domain Knowledge
TopN
Sort
Explicit
Implicit
Hybrid

Moderation

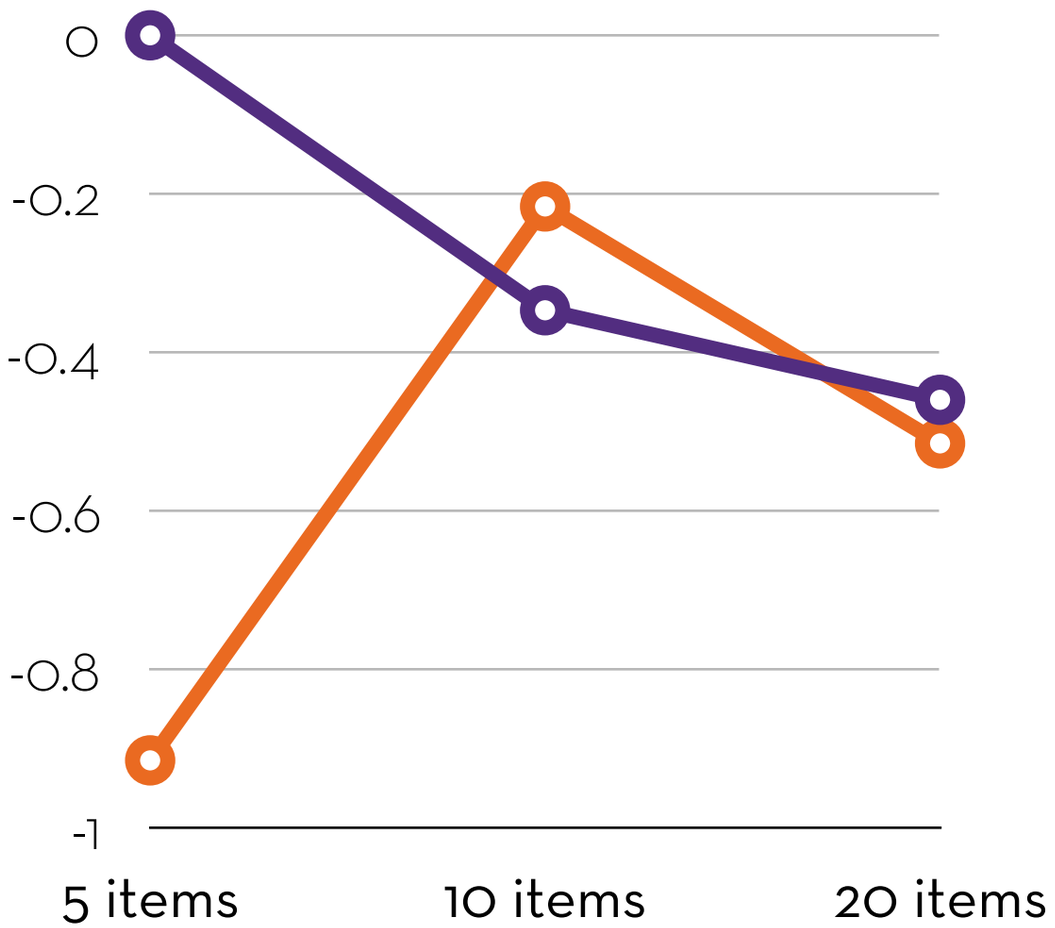
Also moderation: Two manipulations at the same time:

What is the combined effect of list diversity and list length on perceived recommendation quality?

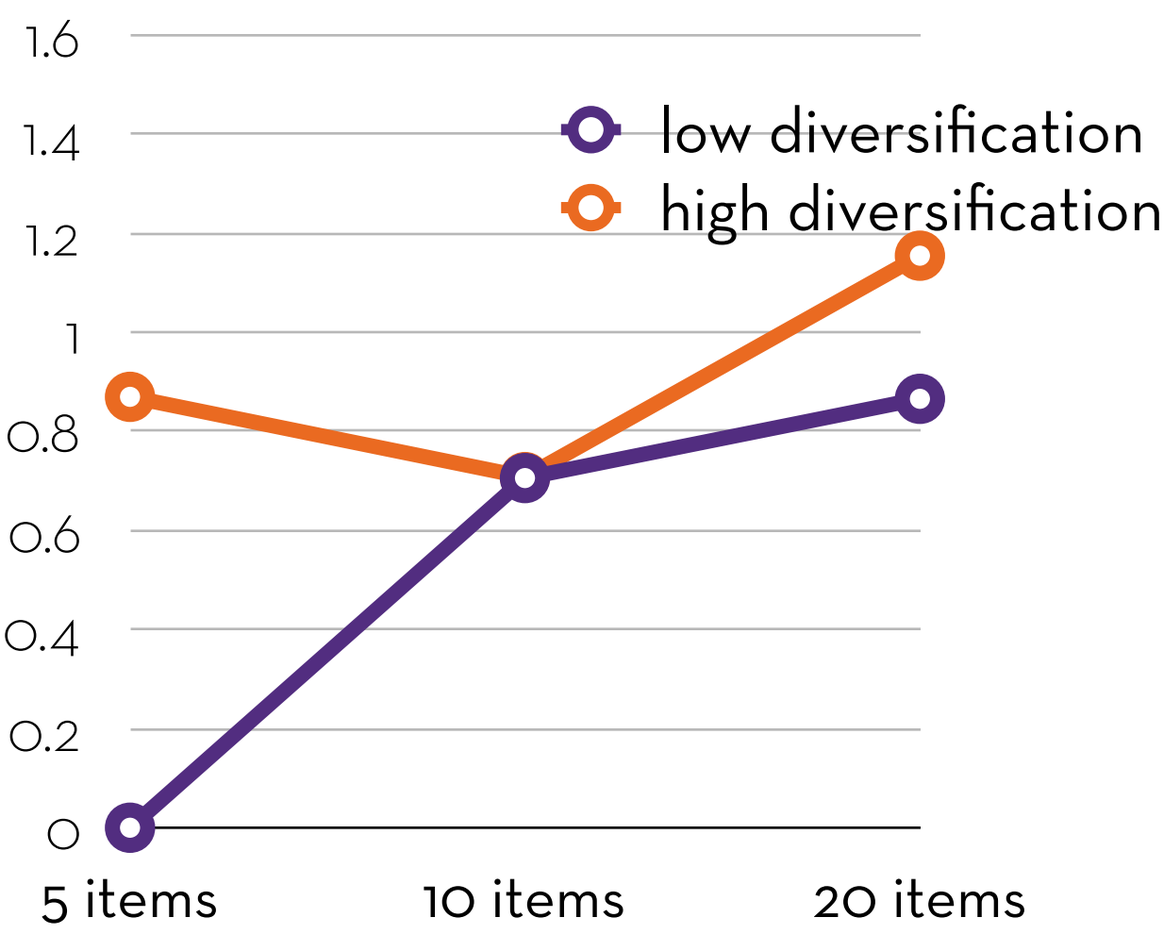
We call this an **interaction effect**

Interaction effect

Choice difficulty



Choice satisfaction



Willemsen et al.: “Understanding the Role of Latent Feature Diversification on Choice Difficulty and Satisfaction”, submitted to UMUAI

Hypotheses

How do we translate research questions into hypotheses?

Hypotheses

Experiments can answer **causal** research questions

i.e., how one variable influences the other

Example: Does a comic-based privacy policy increase privacy awareness compared to a text-based policy?

Hypotheses are predictions regarding the influence of your independent variables (manipulations) on your dependent variables (outcomes)

Example: compared to text, comic-based policies increase privacy knowledge

Hypotheses

Compared to text, comic-based policies increase privacy knowledge

Experimental hypothesis: H1: Mcomic > Mtext

Question: what if they could just as well be worse?

Calculate the means. Do they differ a lot?

Given no effect, we expect the means to be roughly equal

H0: Mcomic = Mtext

To test H1, we try to **reject H0**

Hypotheses

If the difference is larger than expected:

- We may still have found a difference by chance (no real effect), or...
- There is a real difference in means (H0 is incorrect).

The larger the difference, the more confident we are that H0 is incorrect. Then, H1 is **supported**

But never **proven**, because the first option may still apply!

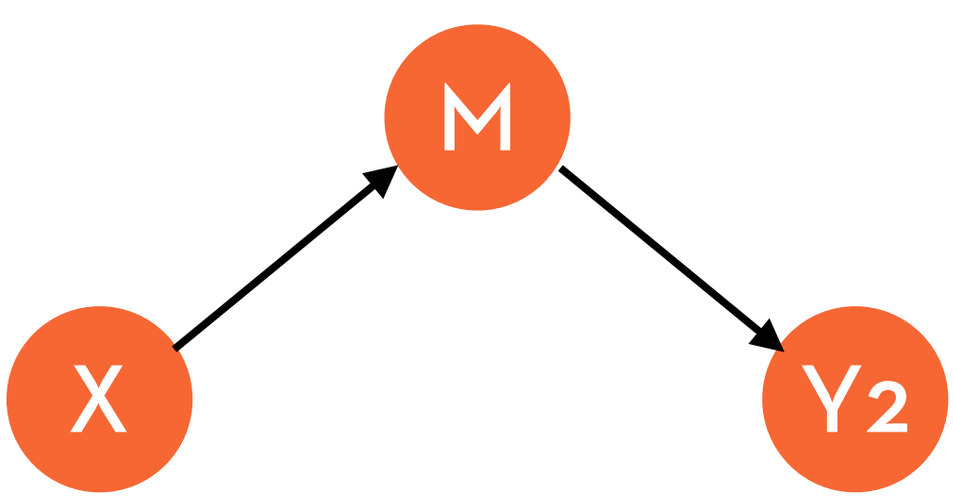
Mediation

What about mediation?

Multiple hypotheses!

Compared to text, comic-based policies (X) increase privacy knowledge (M)

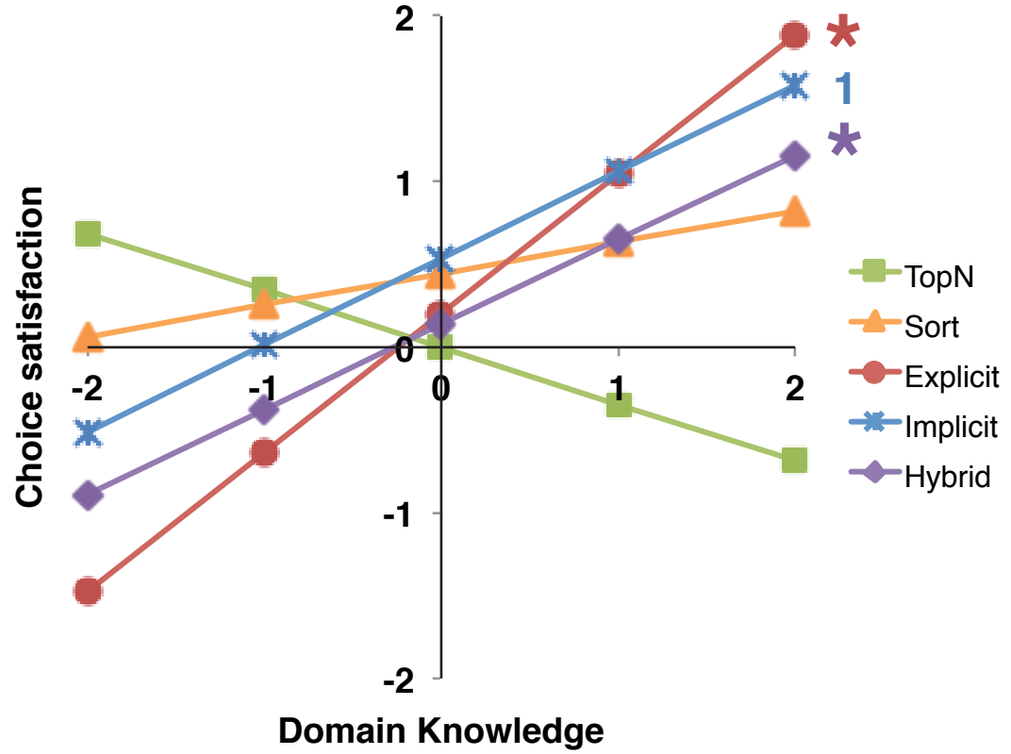
Privacy knowledge (M) is positively associated with perceived privacy protection (Y2)



Moderation

The effect of domain knowledge on choice satisfaction is moderated by PE method:

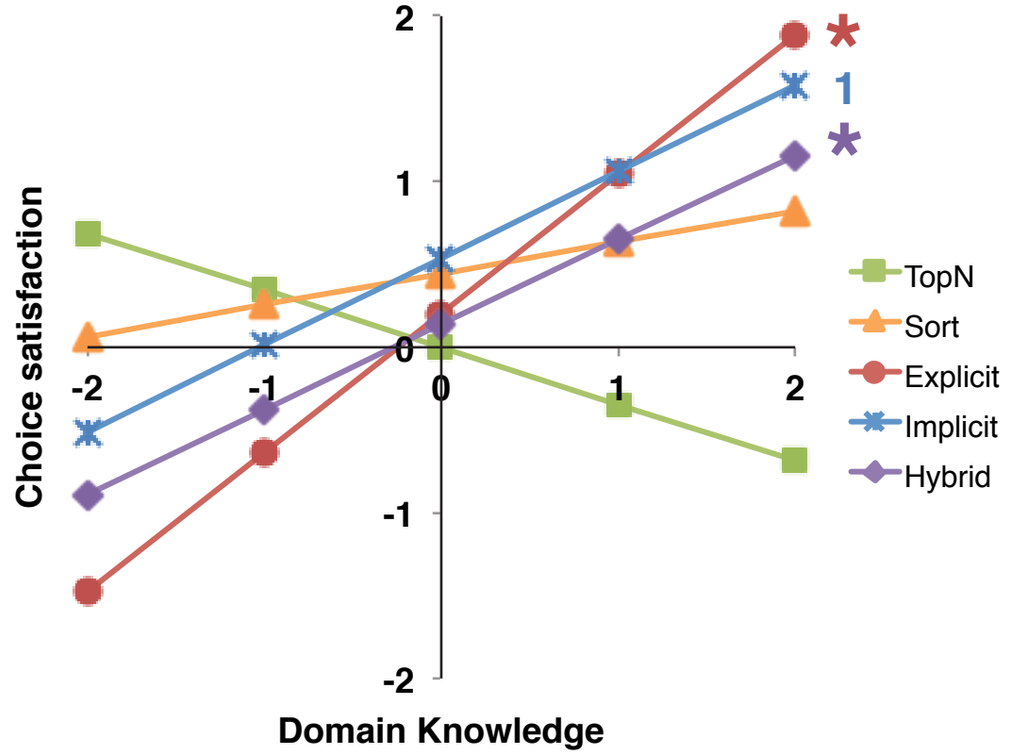
Domain knowledge is negatively associated with choice satisfaction in the TopN condition, but positively associated in the Explicit, Implicit, and Hybrid conditions



Moderation

The effect of PE method on choice satisfaction is moderated by domain knowledge:

For people with low domain knowledge, TopN performs significantly better than the other conditions; for people with high domain knowledge it performs worse



Interaction effect

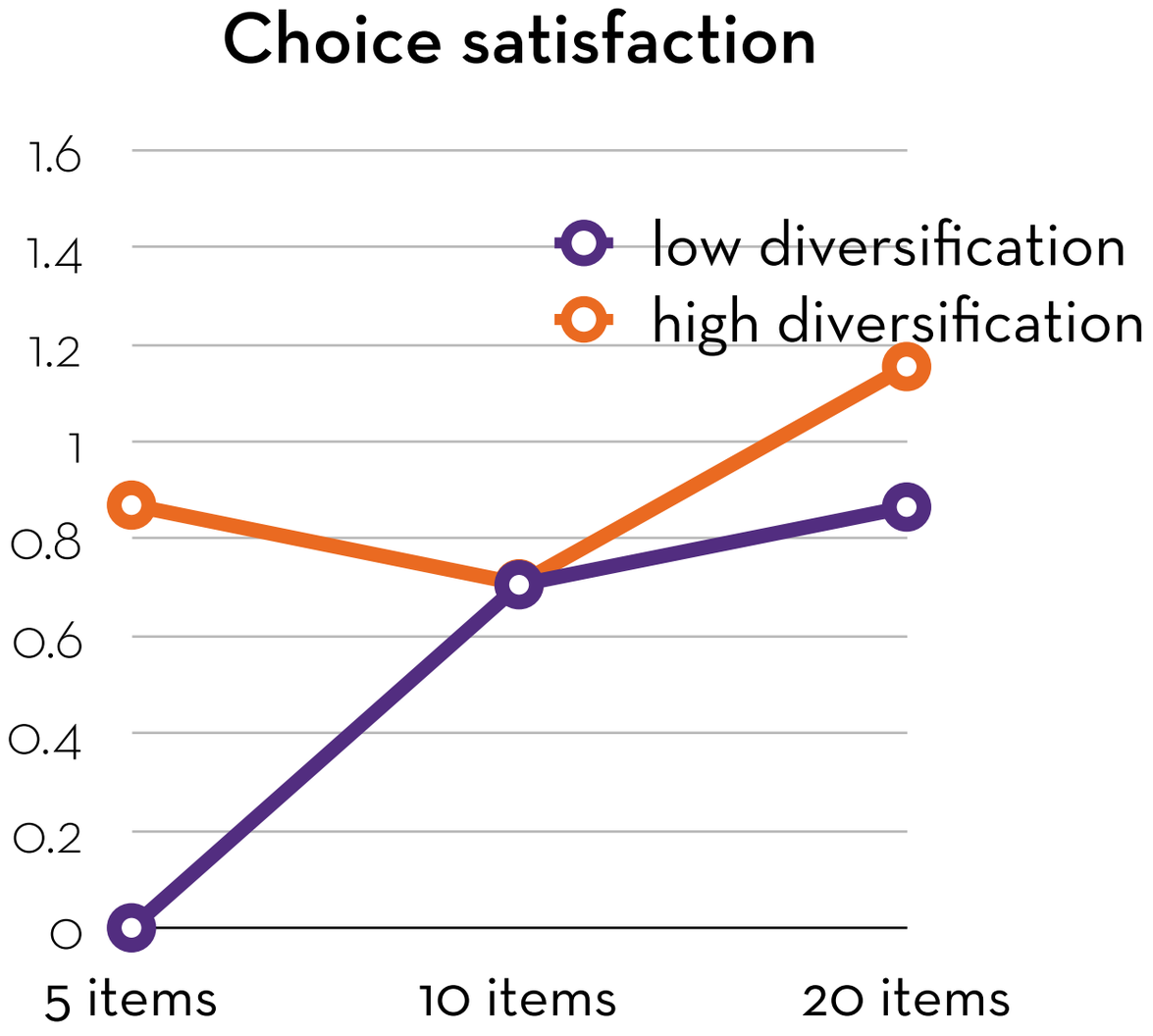
There is an interaction effect between diversification and domain knowledge on choice satisfaction

- High diversification leads to higher choice satisfaction, but only when 5 items are shown

Interaction effect

There is an interaction effect between diversification and domain knowledge on choice satisfaction

Or: Choice satisfaction is significantly lower for 5-item lists, but only when diversification is low



Example

Developing hypotheses

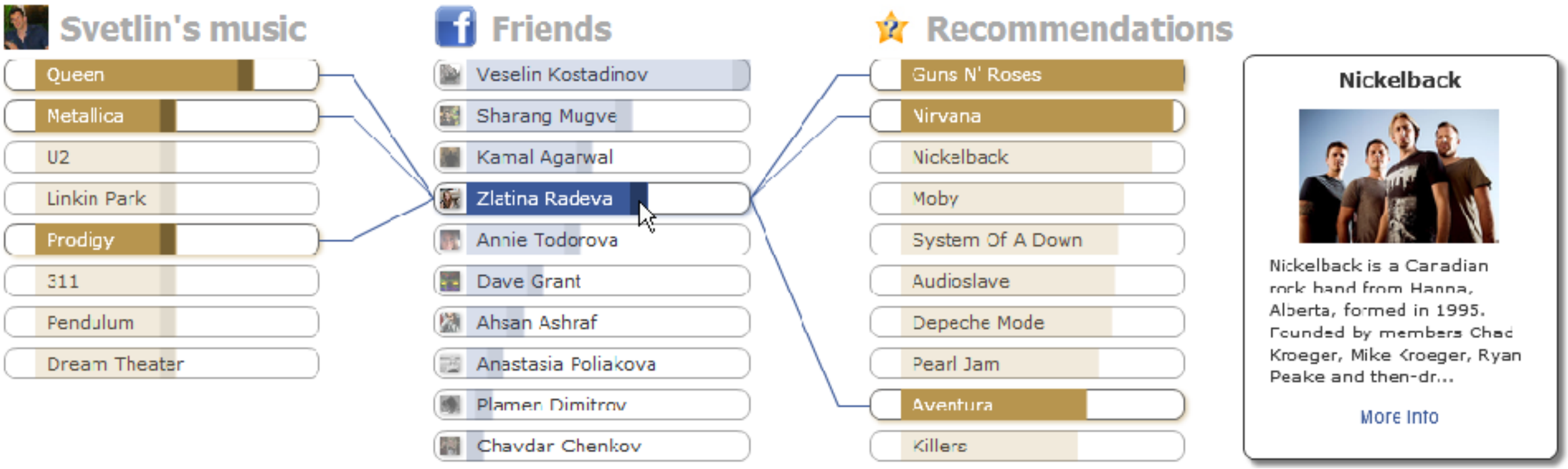
Example

Knijnenburg et al. (2012): "Inspectability and Control in Social Recommenders", *RecSys'12*

The TasteWeights system uses the overlap between you and your friends' Facebook "likes" to give you music recommendations.

- Friends "weights" based on the overlap in likes w/ user
- Friends' other music likes—the ones that are not among the user's likes—are tallied by weight
- Display to the user in a unique graph

Example


Svetlin's music
Queen
Metallica
U2
Linkin Park
Prodigy
311
Pendulum
Dream Theater
Friends
Veselin Kostadinov
Sharang Mugve
Kamal Agarwal
Zlatina Radeva
Annie Todorova
Dave Grant
Ahsan Ashraf
Anastasia Poliakova
Plamen Dimitrov
Chavdar Chenkov
Recommendations
Guns N' Roses
Nirvana
Nickelback
Moby
System Of A Down
Audioslave
Depeche Mode
Pearl Jam
Aventura
Killers
Nickelback
Nickelback is a Caradian rock band from Hanna, Alberta, formed in 1995. Founded by members Chad Kroeger, Mike Kroeger, Ryan Peake and then-dr...
More Info

Research question

How do inspectability (the cool graph) and control (the fact that I can set weights) influence the user experience?

Manipulations

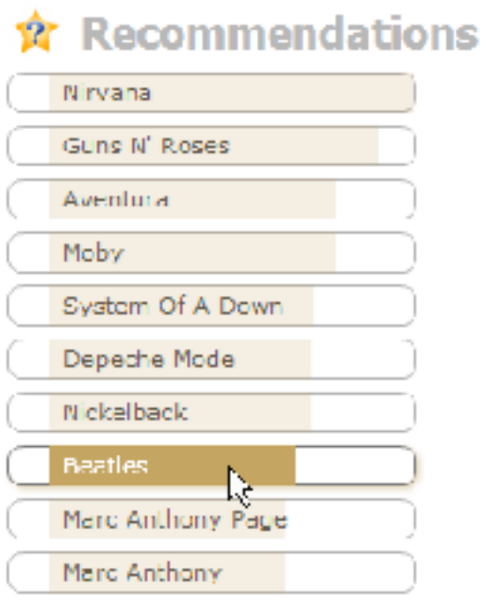
3 control conditions:

- No control (just use likes)
- Item control (weigh likes)
- Friend control (weigh friends)

Manipulations

2 inspectability conditions:

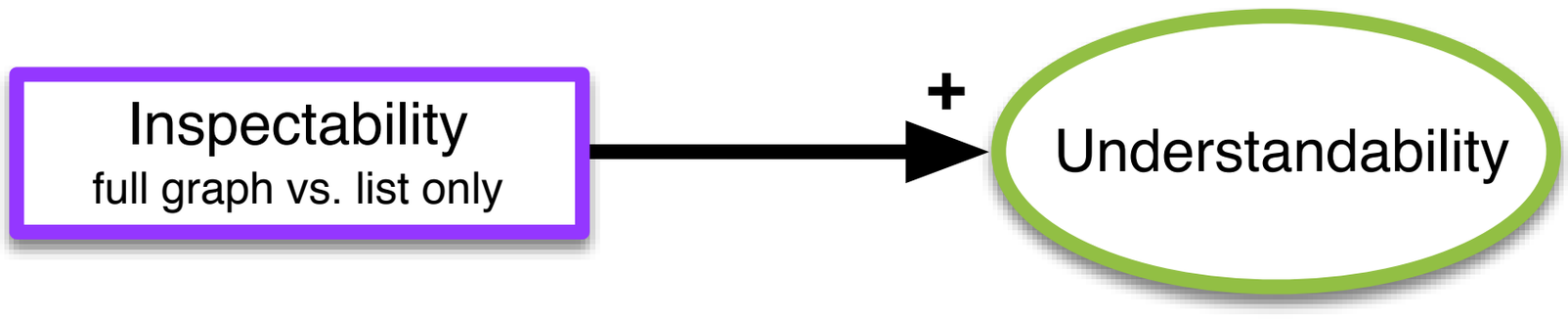
- List of recommendations vs. recommendation graph



Inspectability

Herlocker argues that explanation provides transparency, "exposing the reasoning behind a recommendation"

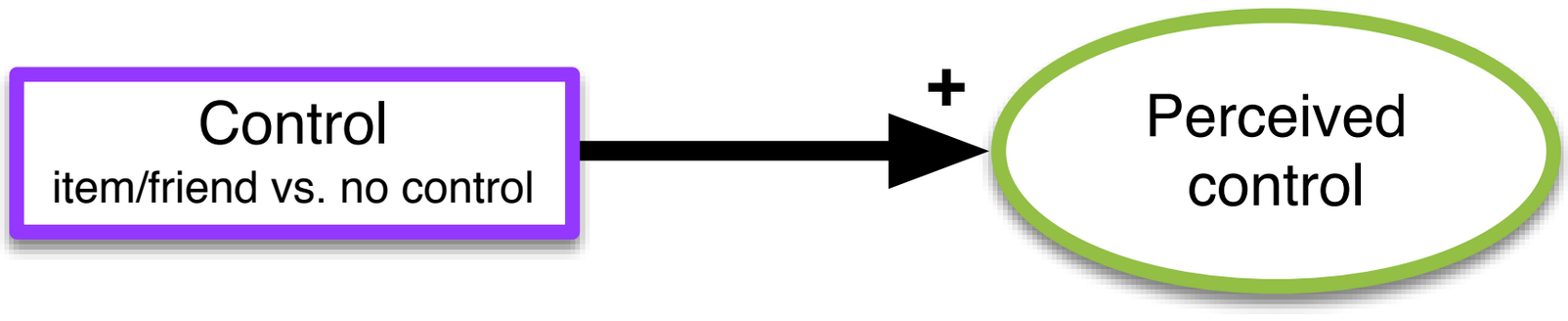
H1: The "full graph" condition results in higher understandability than the "list only" condition



Control

Multiple studies highlight the benefits of interactive interfaces that support control over the recommendation process.

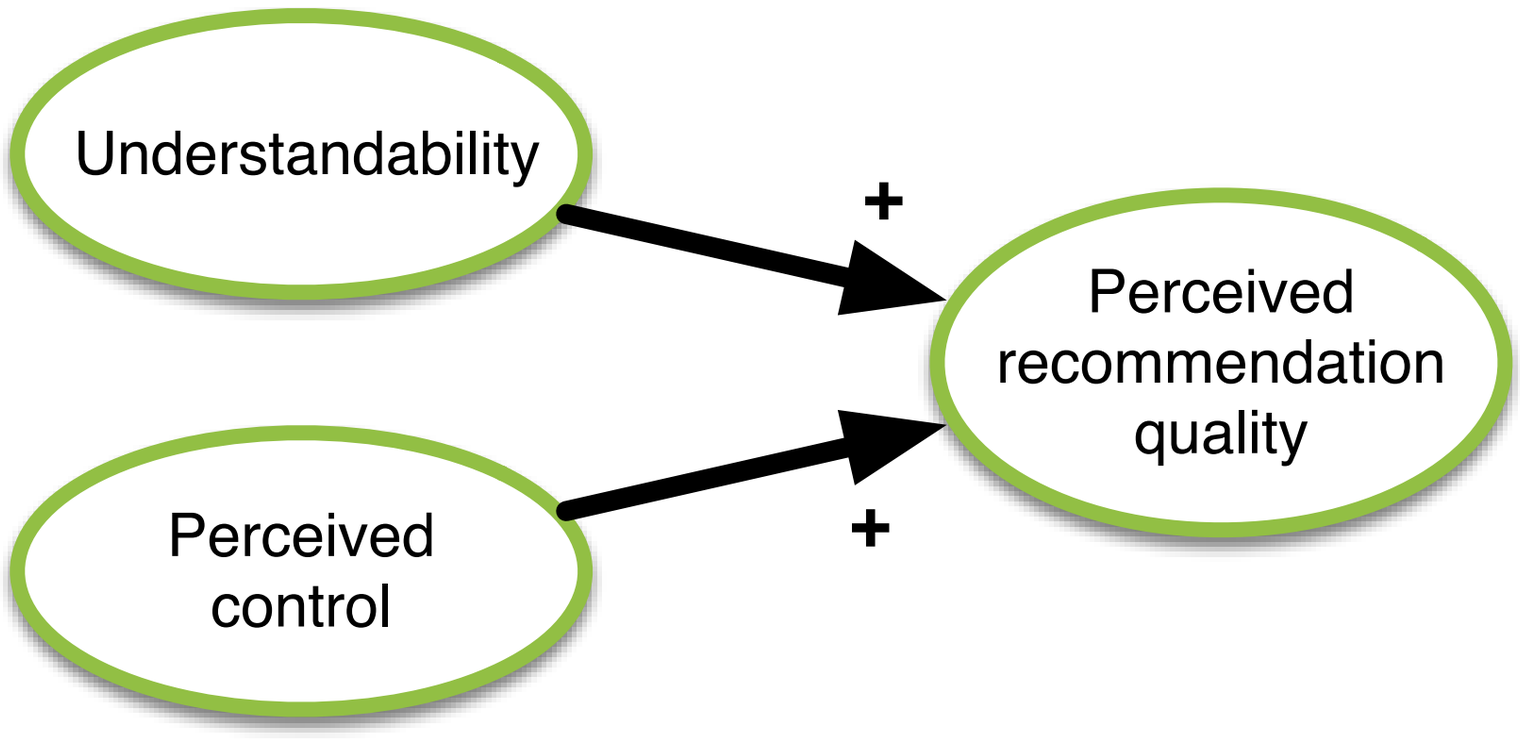
H2: The “item control” and “friend control” conditions lead to a higher level of perceived control than the “no control” condition



Perceived quality

Tintarev and Masthoff show that explanations make it easier to judge the quality of recommendations.

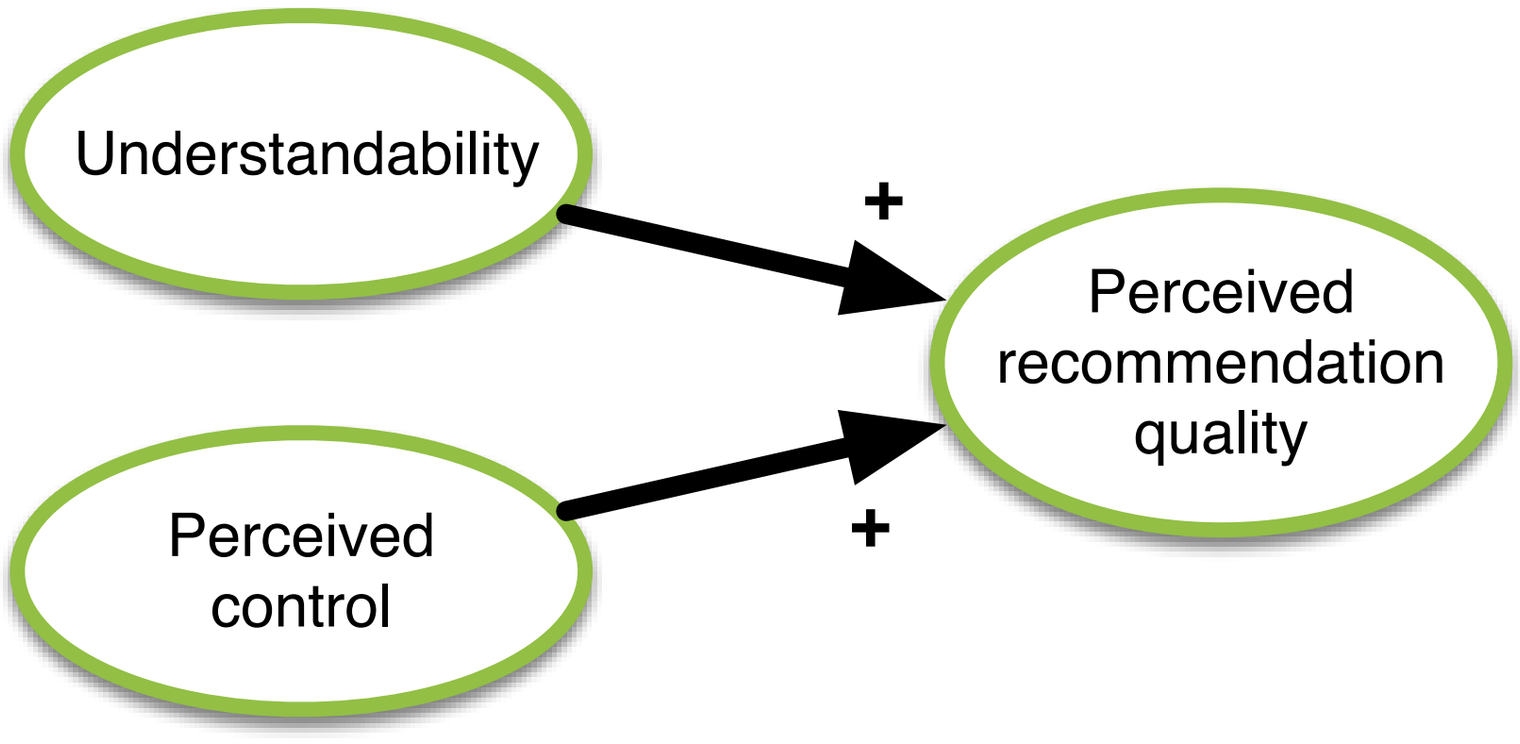
H3: Understandability is positively associated with perceived recommendation quality



Perceived quality

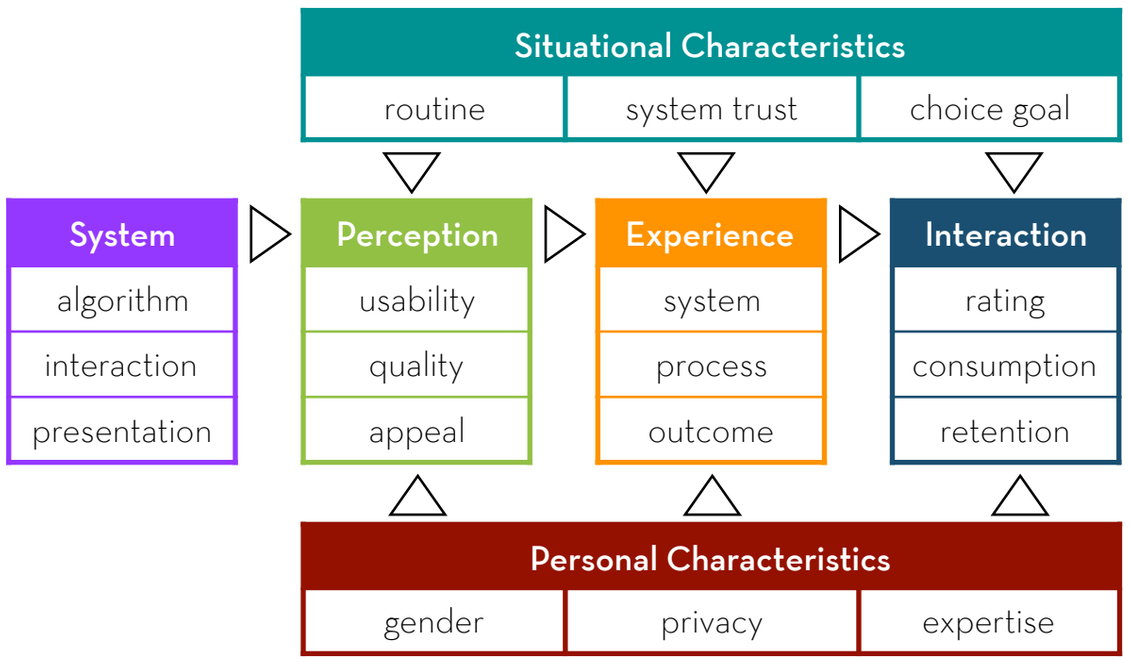
McNee et al. found that study participants preferred user-controlled interfaces because these systems “best understood their tastes”

H4: Perceived control is positively associated with perceived recommendation quality



Satisfaction

Knijnenburg et al. developed a framework that describes how certain manipulations influence subjective system aspects (i.e. understandability, perceived control and recommendation quality), which in turn influence user experience (i.e. system satisfaction).



Satisfaction

Understandability (H5), perceived control (H6) and perceived recommendation quality (H7) are positively associated with system satisfaction