

Maya Cakmak, Matt Kay, Brad Jacobson, King Xia

EMPIRICAL USER-STUDIES

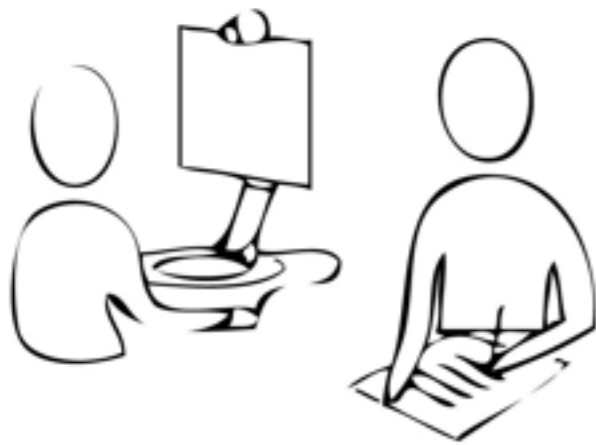


University of
Washington

human-computer interaction
CSE 440 WINTER 2015

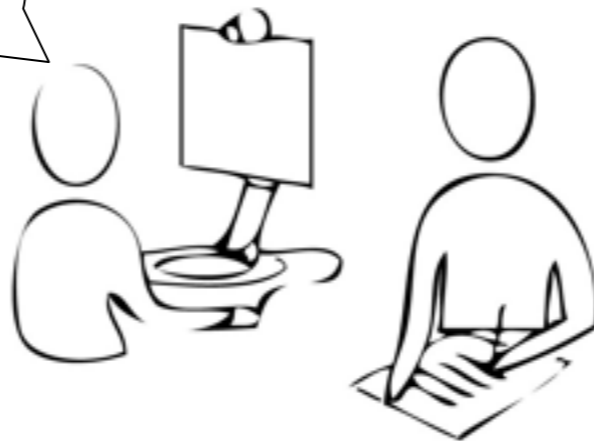
FEB 19 - WEEK 7 - THURSDAY

Methods for observing interaction

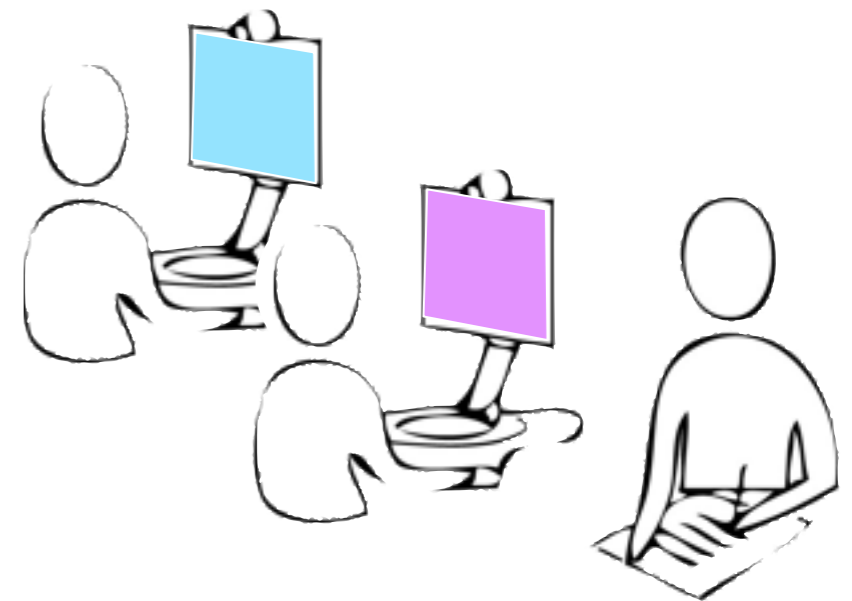


Passive observation

hmmmm
blah blah
blah bla



Think-aloud protocol

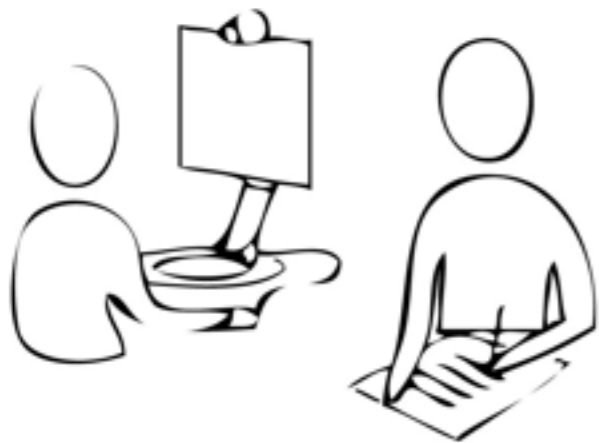


Comparative study



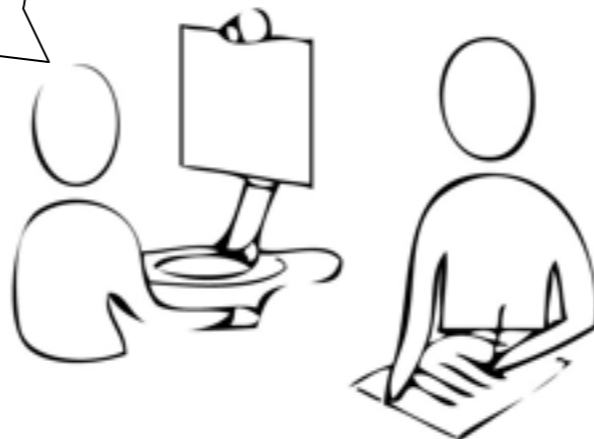
Last week

Methods for observing interaction

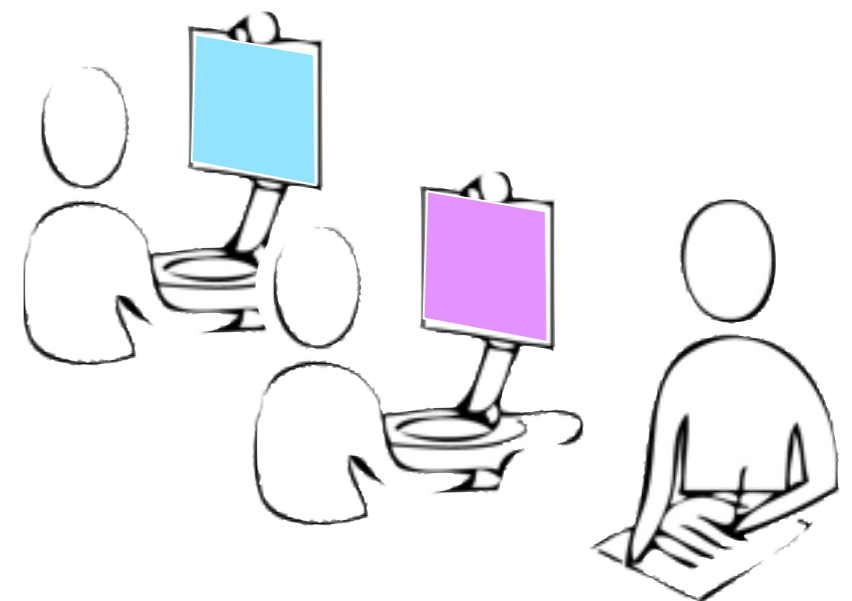


Passive observation

hmmmm
blah blah
blah bla



Think-aloud protocol



Comparative study

“Empirical user study”

“Controlled experiment”



Last week

Today

Evaluation Techniques (re-cap)

- Asking users
 - Questionnaires, interviews, focus groups
- Observing users
 - Passive observation, think-aloud protocol, ethnography, empirical user studies
- Make users observe themselves
 - Diaries, experience sampling
- Ask experts
 - Heuristic evaluation, cognitive walkthrough

Evaluation Techniques (re-cap)

- Asking users
 - Questionnaires, interviews, focus groups
- Observing users
 - Passive observation, think-aloud protocol, ethnography, empirical user studies
- Make users observe themselves
 - Diaries, experience sampling
- Ask experts
 - Heuristic evaluation, cognitive walkthrough

Designing an empirical study

Designing an empirical study

- What is being compared?
 - Independent variables

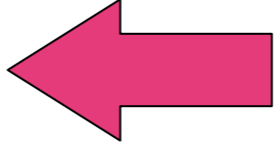
Designing an empirical study

- What is being compared?
 - Independent variables
- What are they being compared in?
 - Dependent variables (“metrics”)

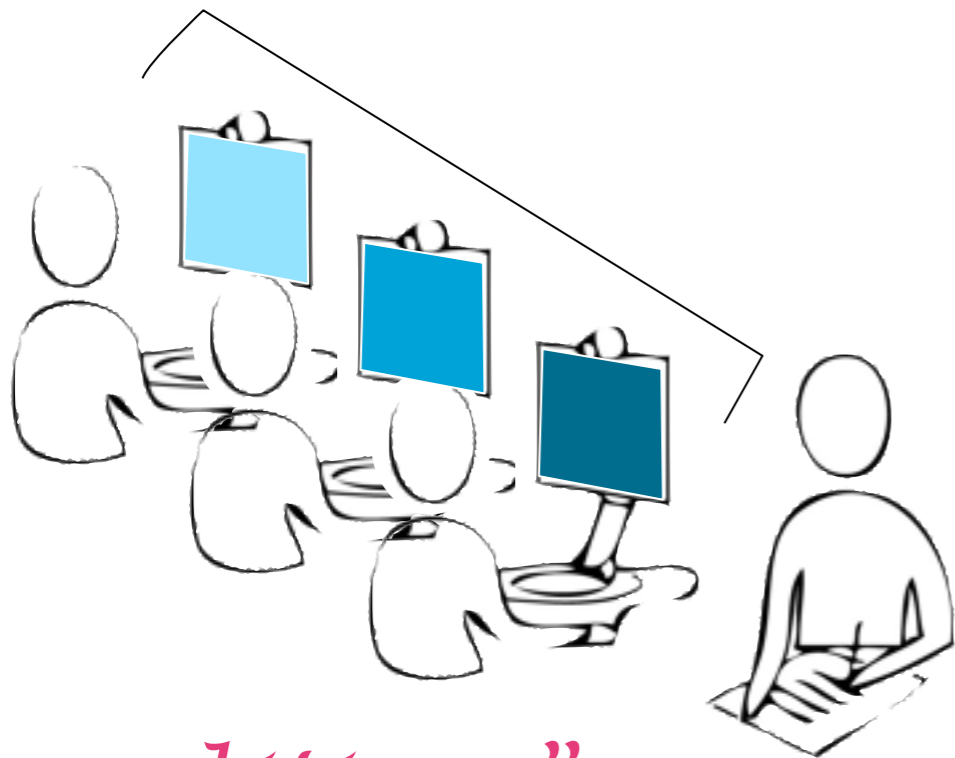
Designing an empirical study

- What is being compared?
 - Independent variables
- What are they being compared in?
 - Dependent variables (“metrics”)
- What (else) is being varied? What is kept constant?
 - Extraneous variables

Designing an empirical study

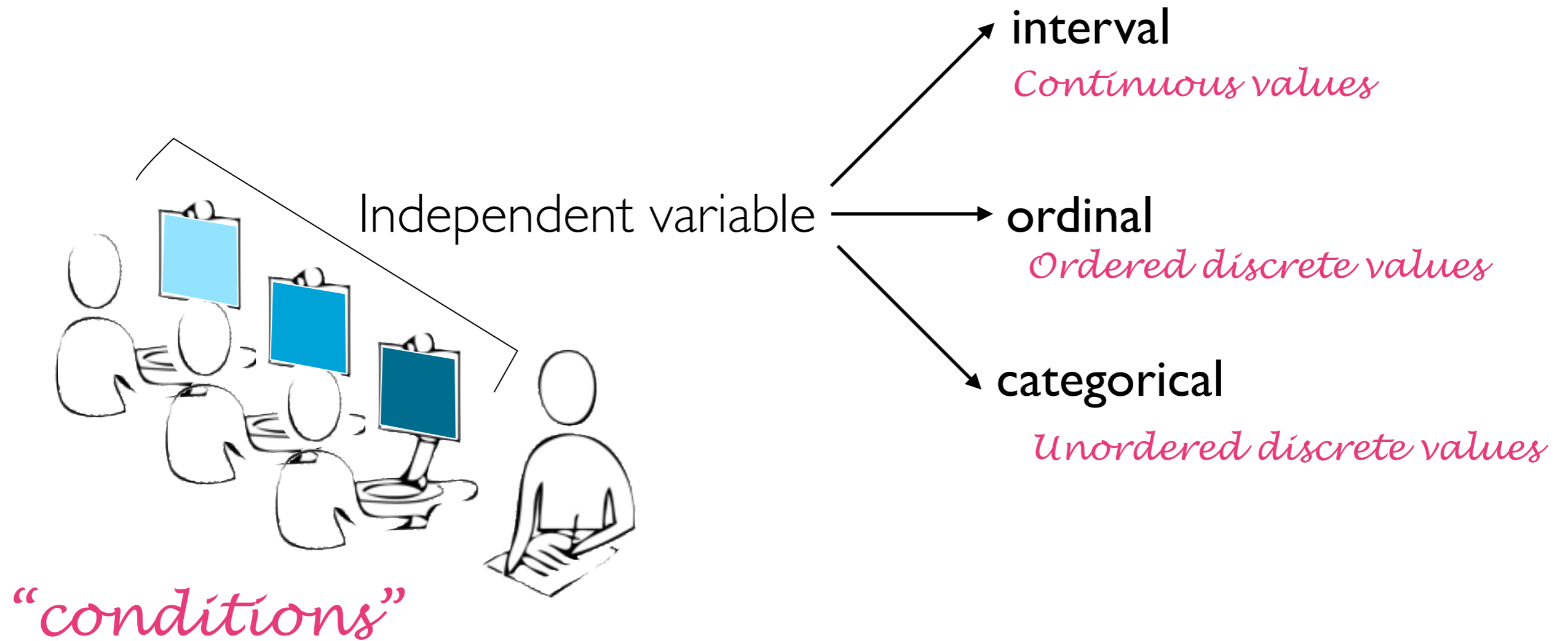
- What is being compared? 
 - Independent variables
- What are they being compared in?
 - Dependent variables (“metrics”)
- What (else) is being varied? What is kept constant?
 - Extraneous variables

What is being compared?



“conditions”

What is being compared?



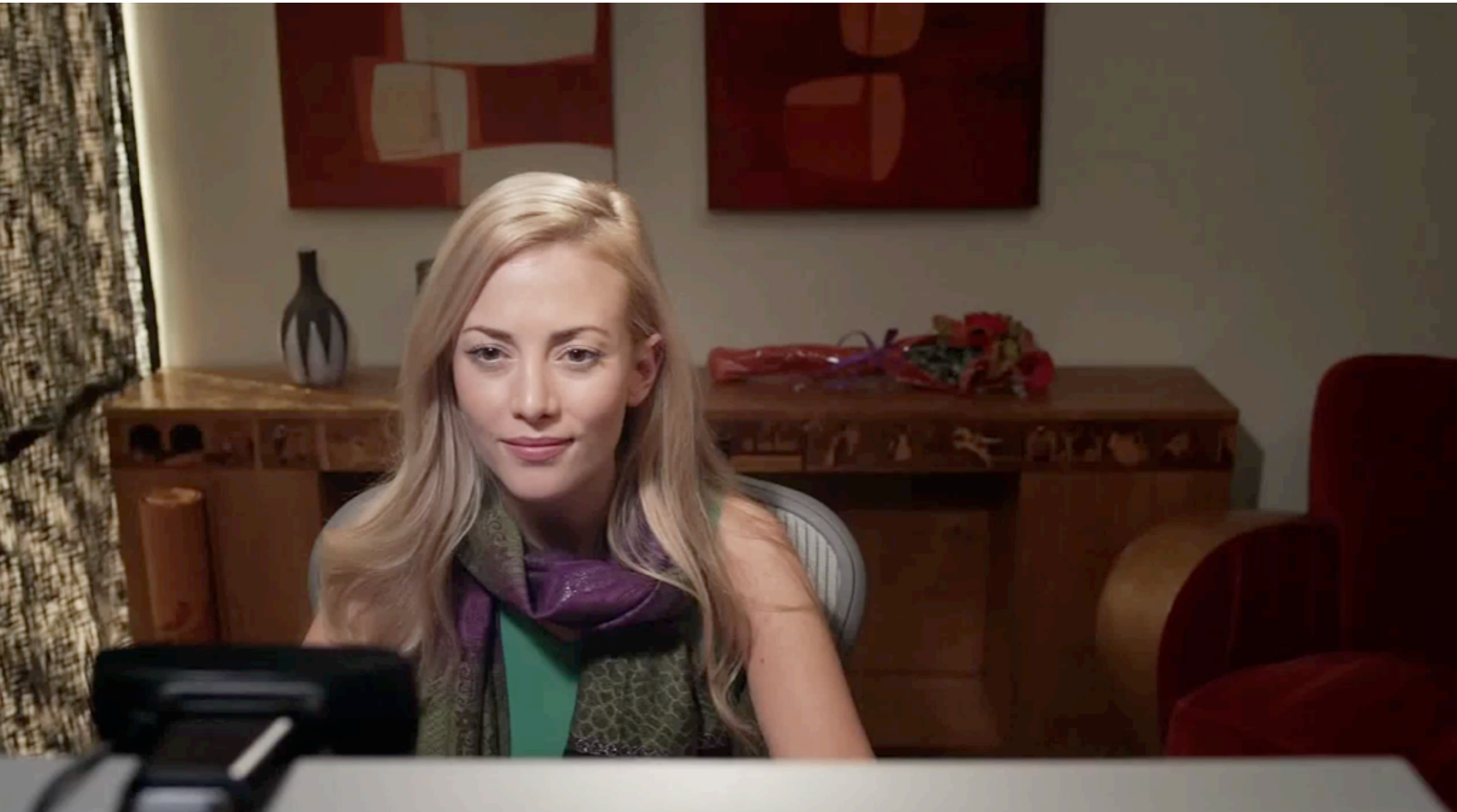
What is being compared?

- Example: **Interval** independent variable
 - What is the effect of **height** on telepresence systems?



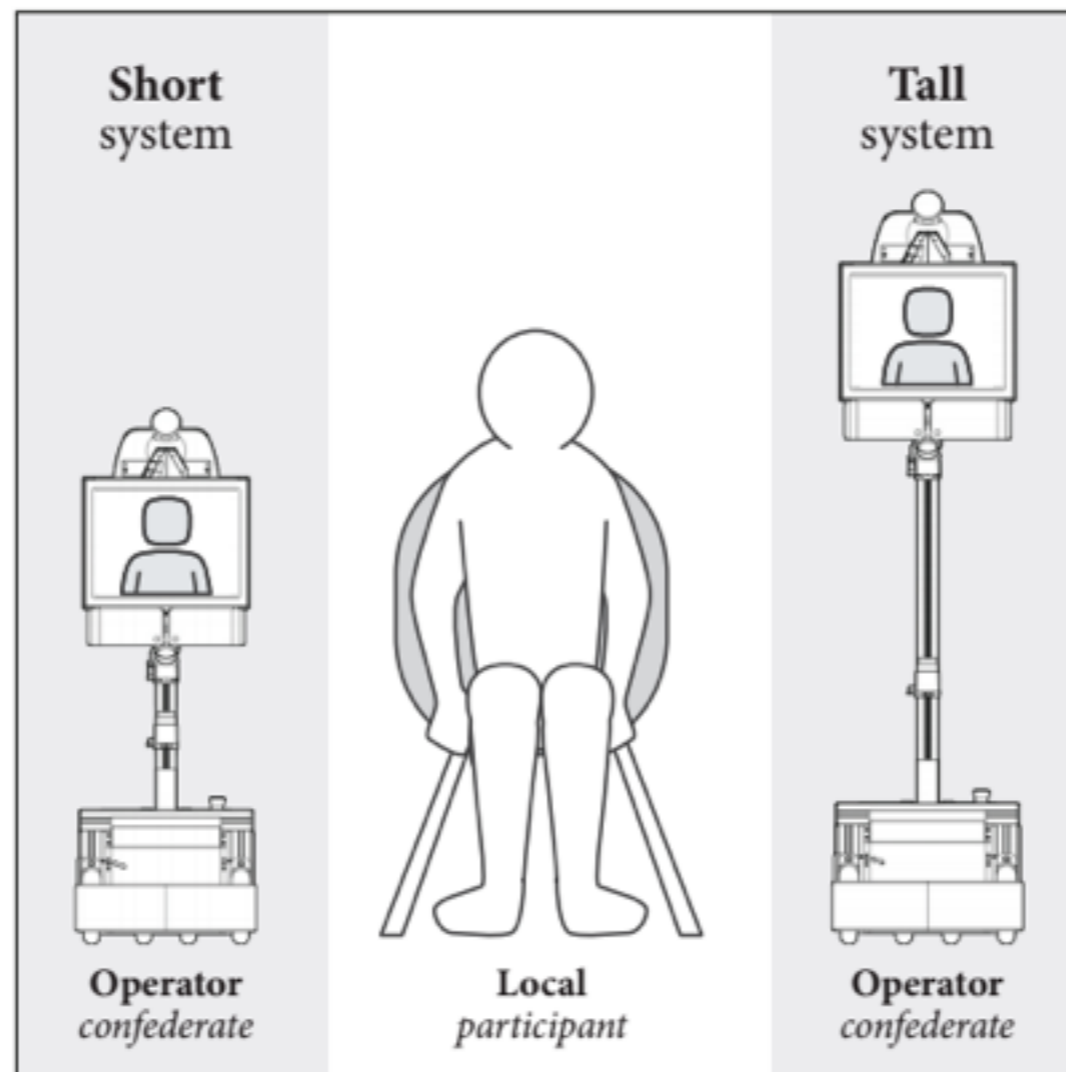
From Left: AnyBots QB, RoboDynamics TiLR, Gostai Jazz Connect, Mantaro's Mantaro Bot, and VGo

Robotic telepresence



What is being compared?

- Example: **Interval** independent variable
 - What is the effect of height on telepresence systems?



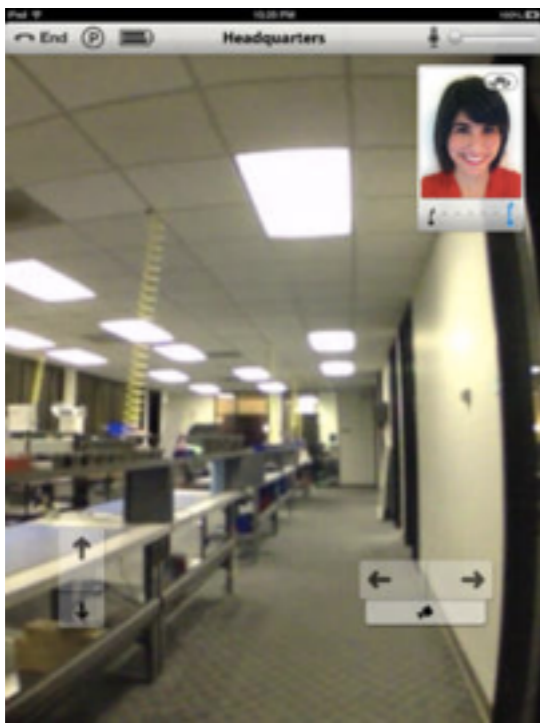
What is being compared?

- Example: **Ordinal** independent variable
 - What is the effect of **educational background** on acceptance of robots in the workplace?

high school < college < graduate degree

What is being compared?

- Example: **Categorical** independent variable
 - What is the effect of **input modality** on telepresence systems?



– keyboard

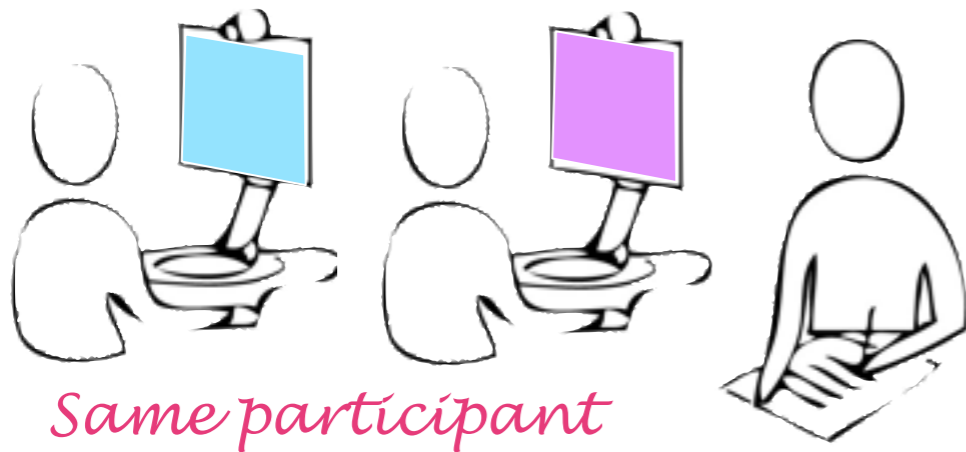


– mouse

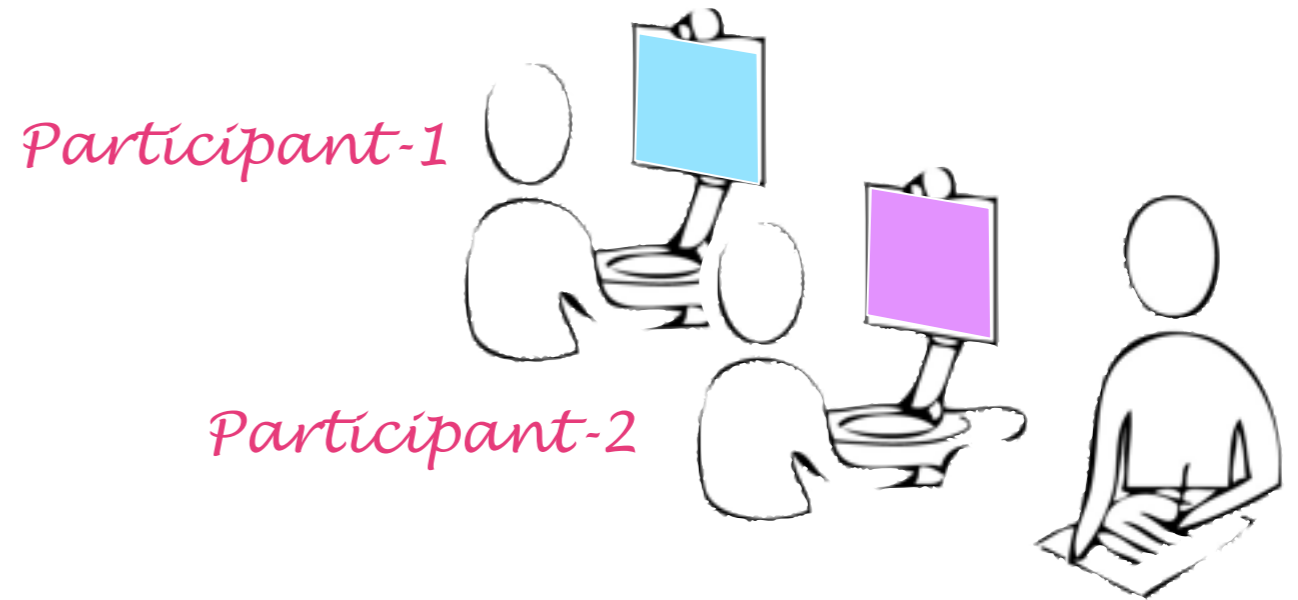


– joystick

Within-subject vs. between subject

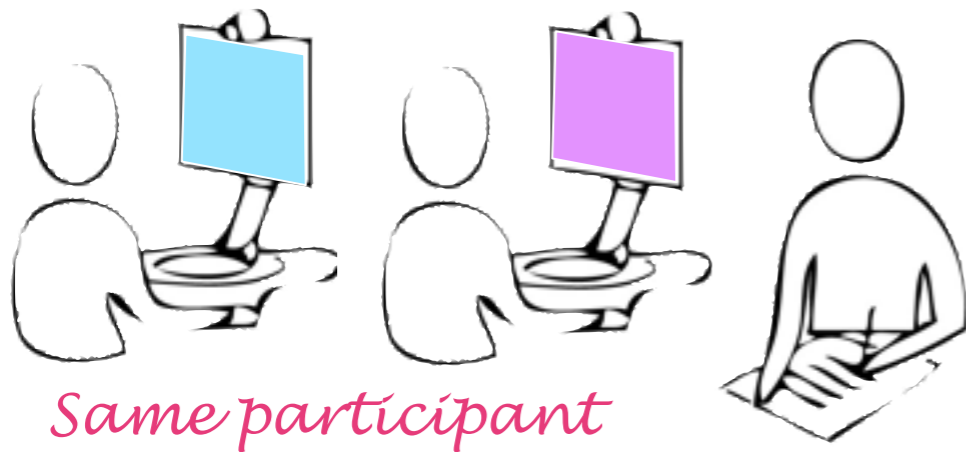


within



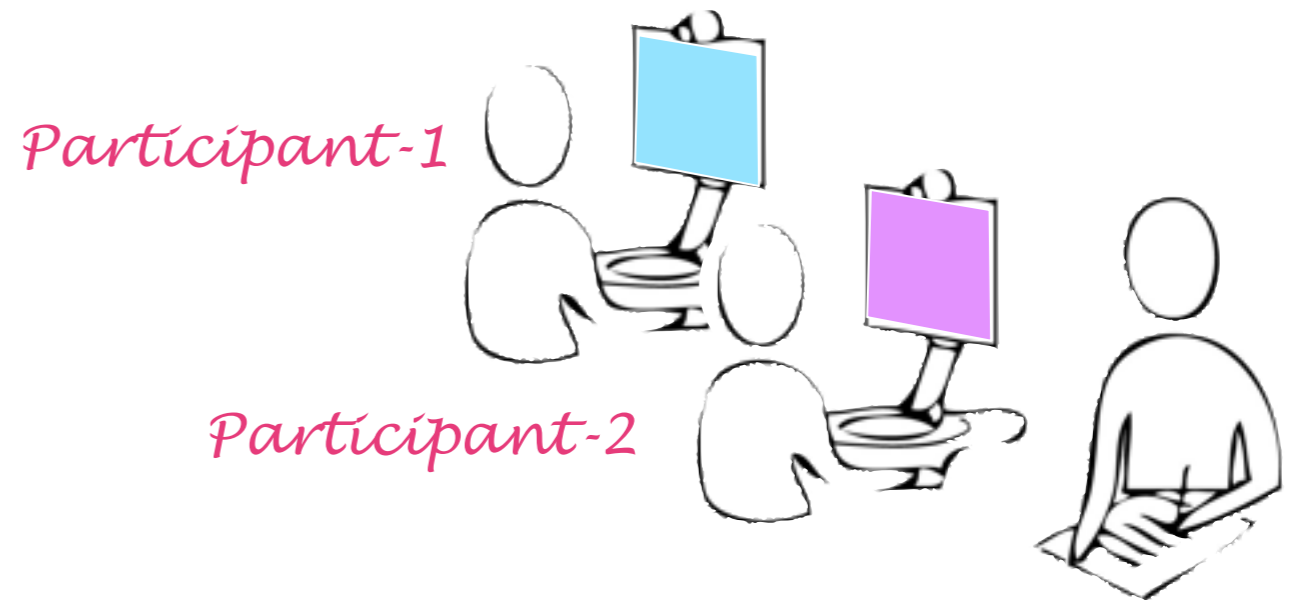
between

Within-subject vs. between subject



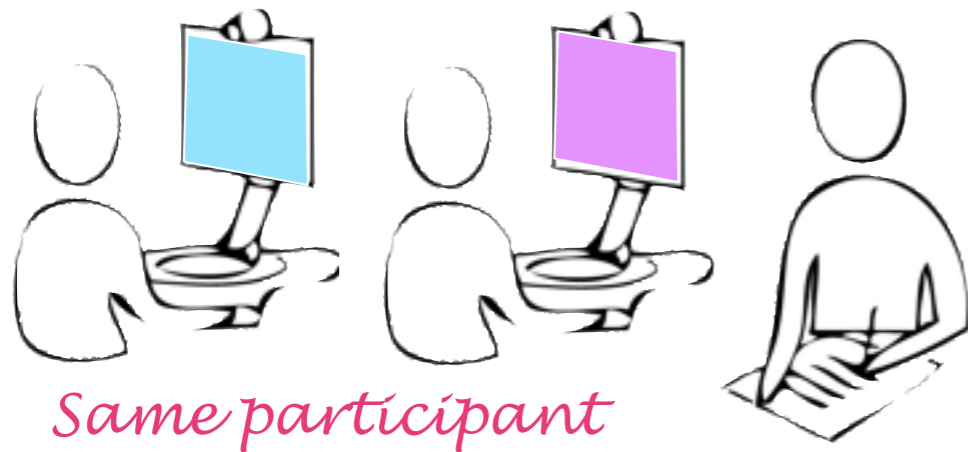
within

- + allows comparison
- + requires less participants
- subject to ordering effects



between

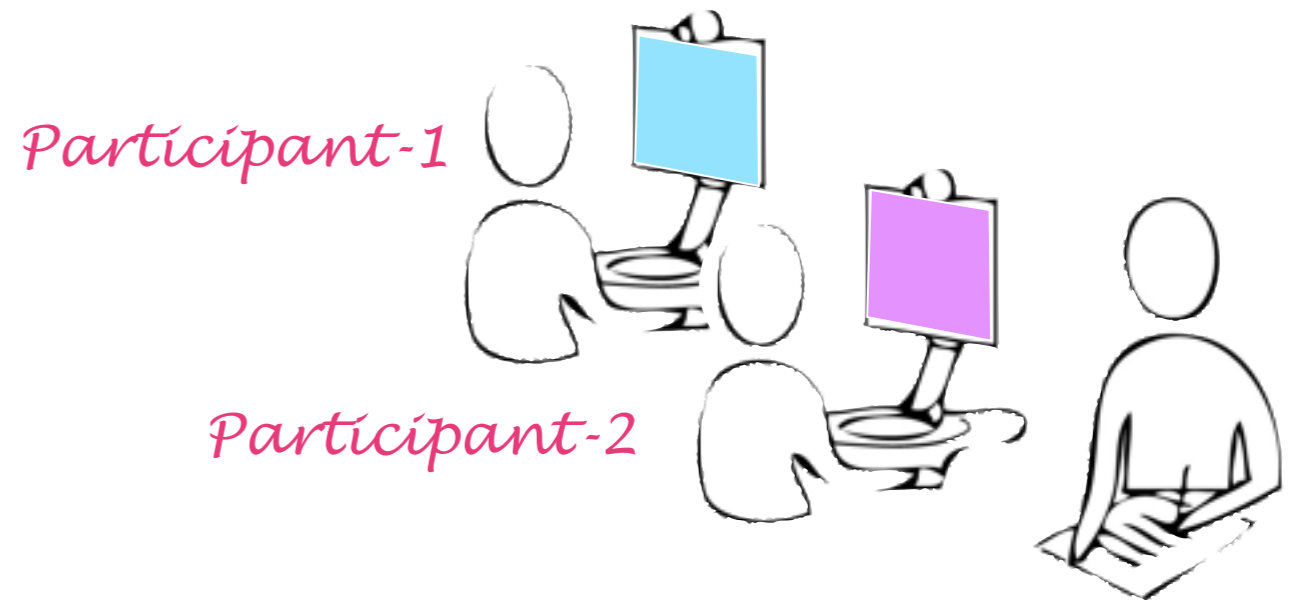
Within-subject vs. between subject



within

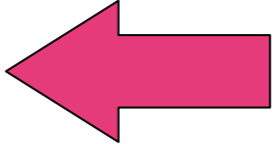
- + allows comparison
- + requires less participants
- subject to ordering effects

> Order counterbalancing



between

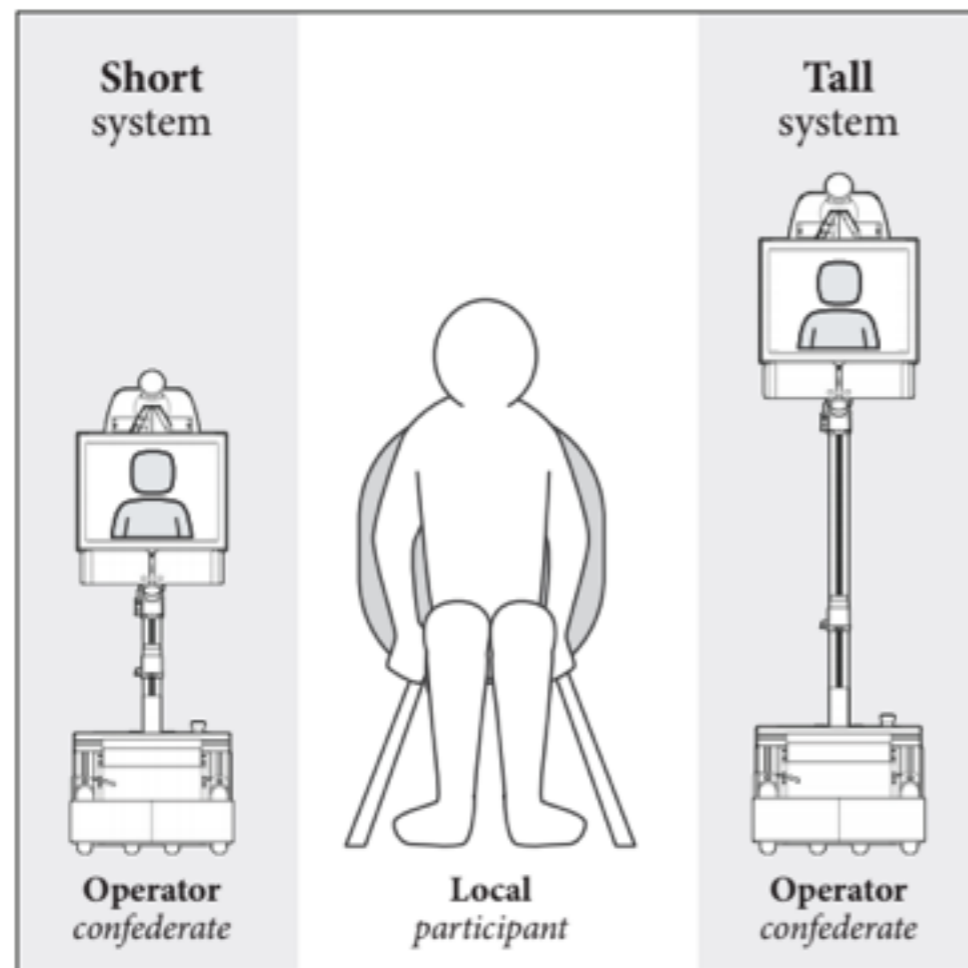
Designing an empirical study

- What is being compared?
 - Independent variables
- What are they being compared in? 
 - Dependent variables (“metrics”)
- What (else) is being varied? What is kept constant?
 - Extraneous variables

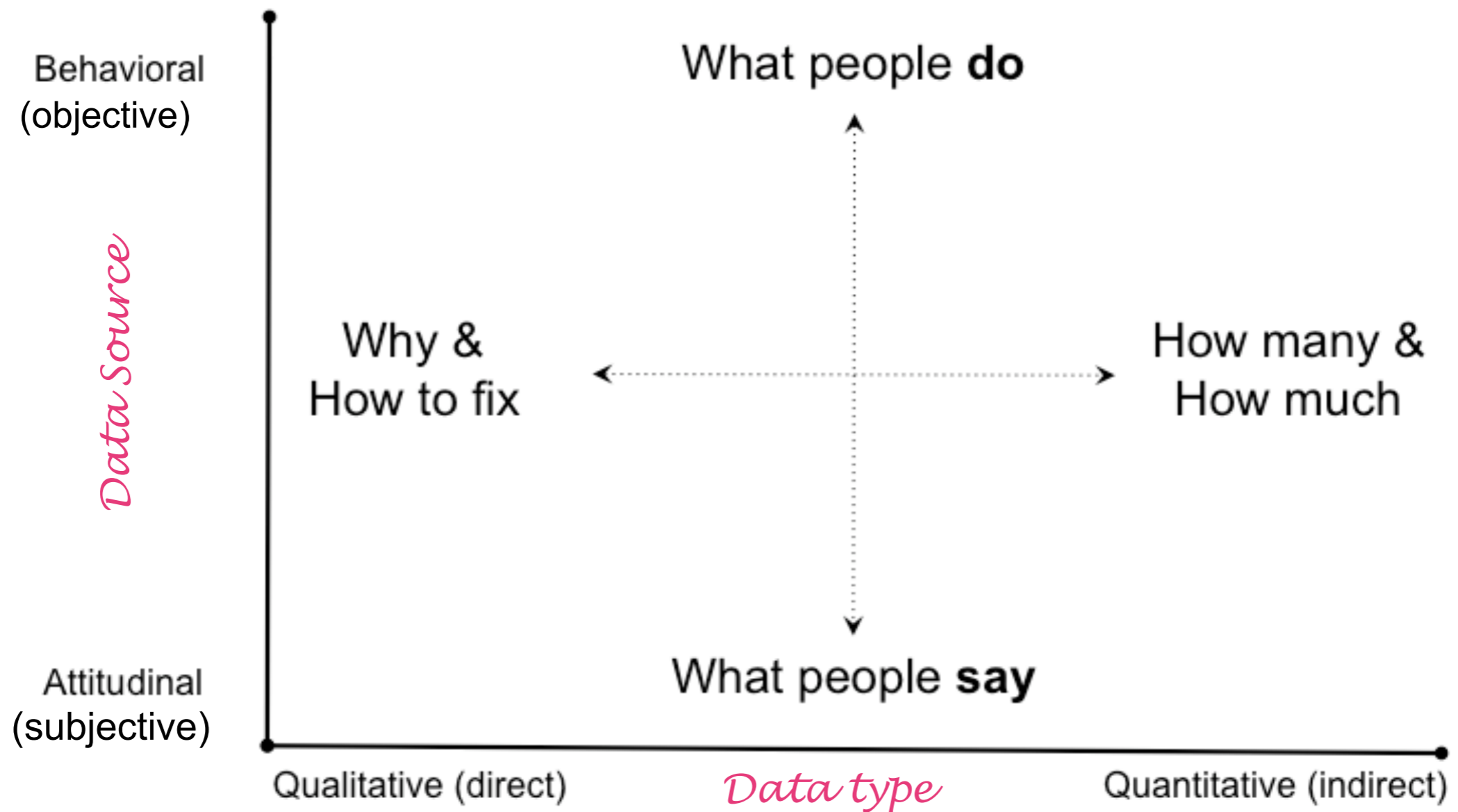
Independent vs. dependent variable

- Example:
 - What is the effect of height on telepresence systems?

in terms of what?

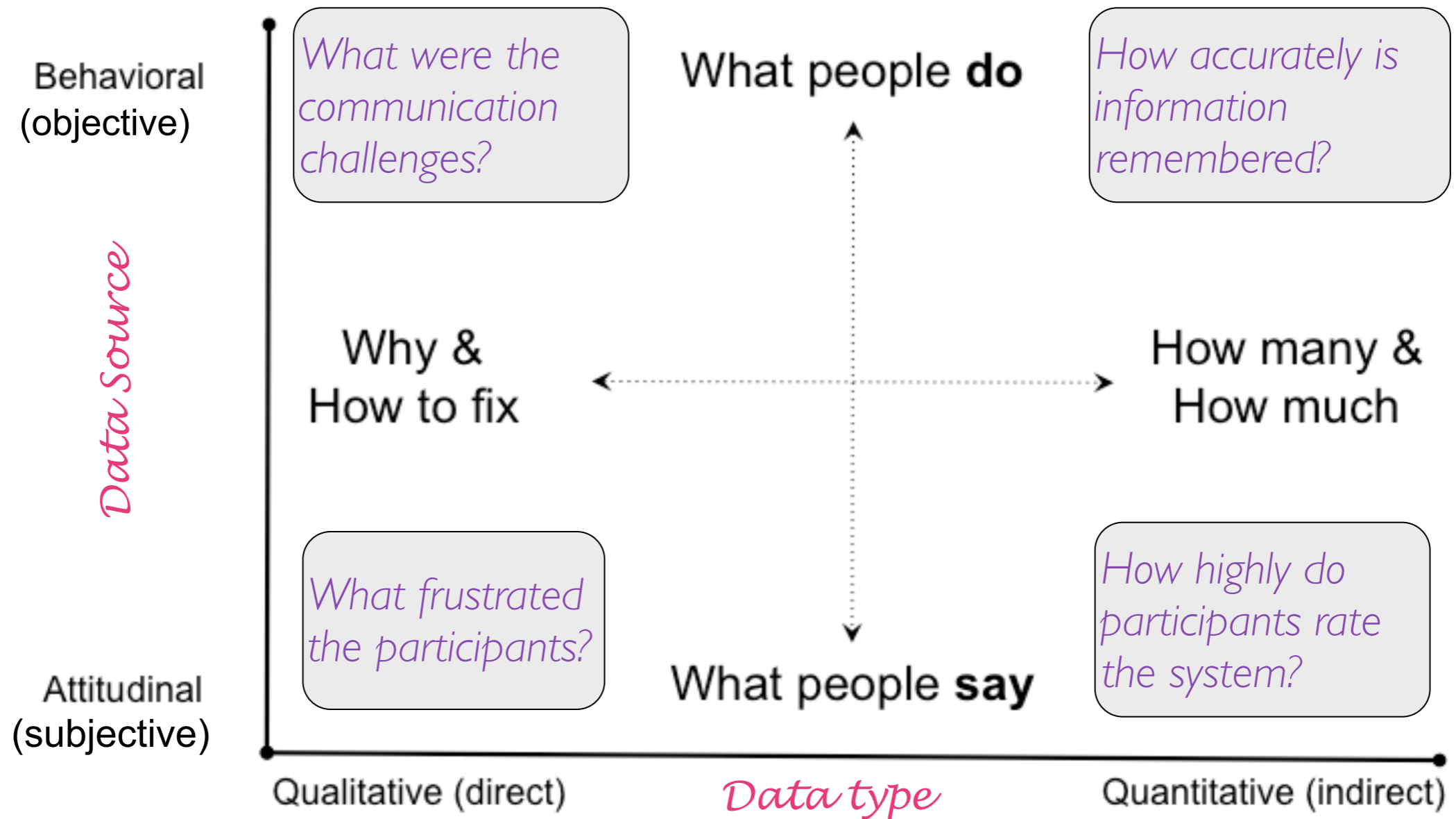


What to measure or observe?



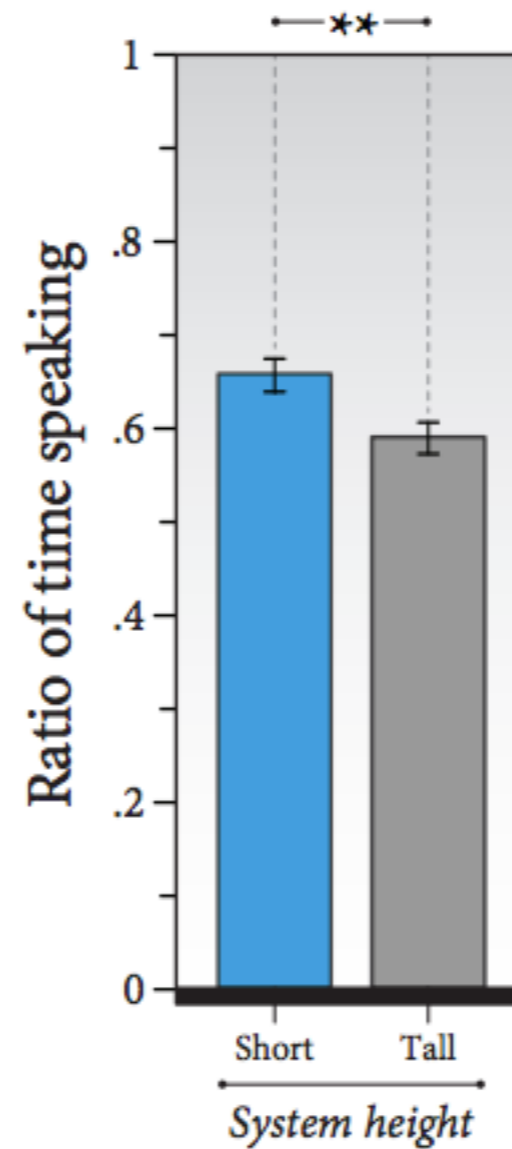
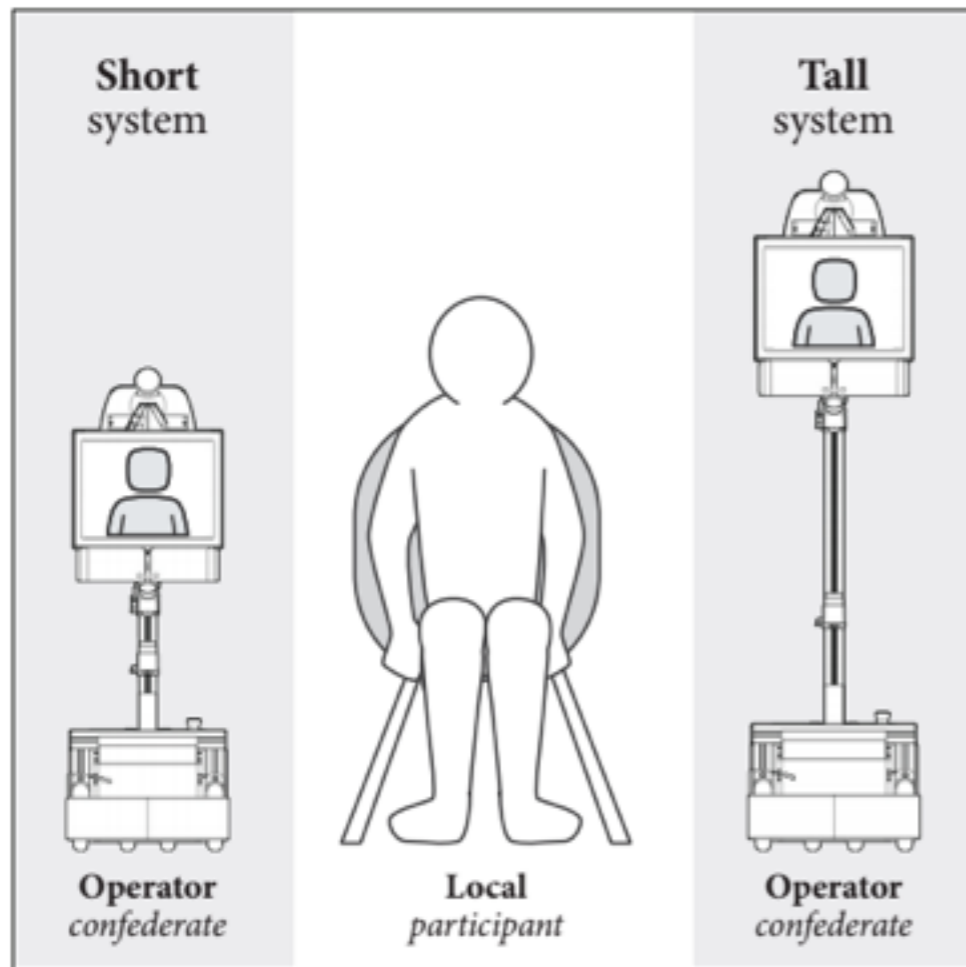
© 2008 by Christian Rohrer

What to measure or observe?

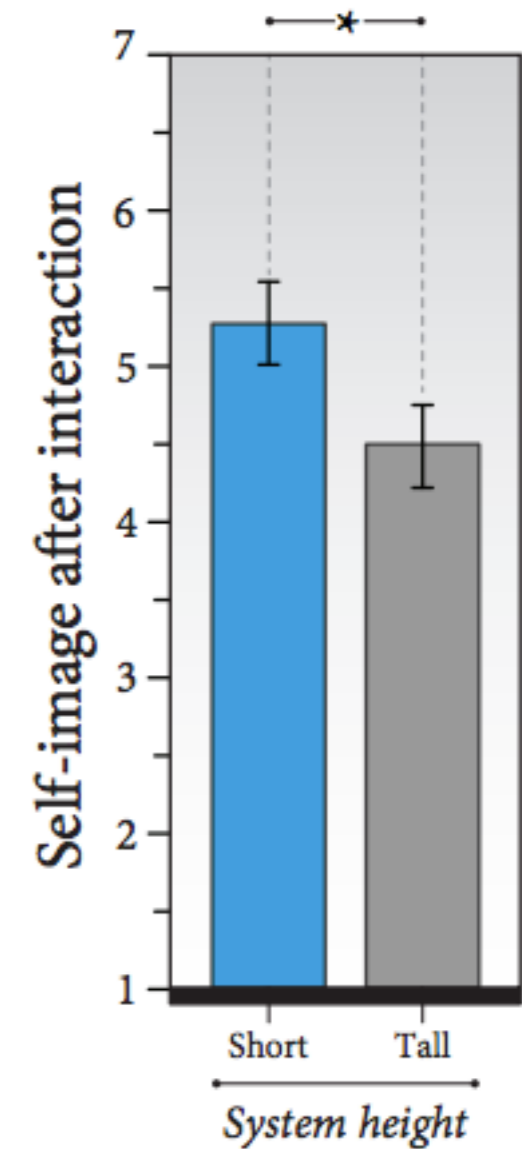


© 2008 by Christian Rohrer

Dependent variables



what people do..



what people say..

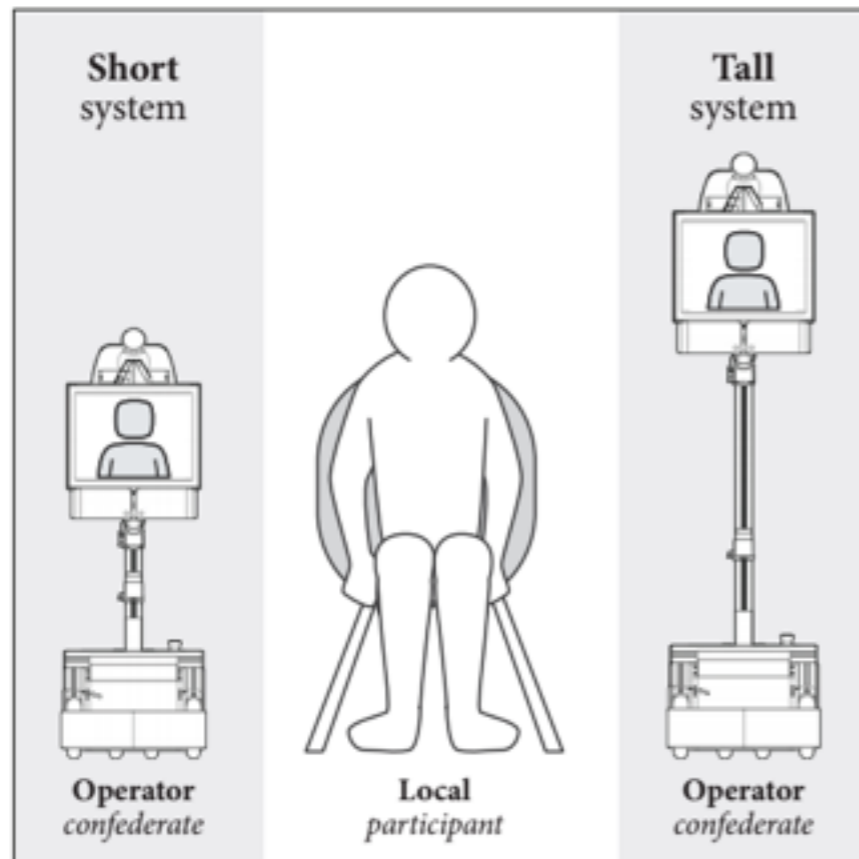
What is being measured?

- Example: **Interval** dependent variable

– What is the effect of height on conversation control?

- ratio of time speaking
- ratio of decisions influenced
- self assessment of control

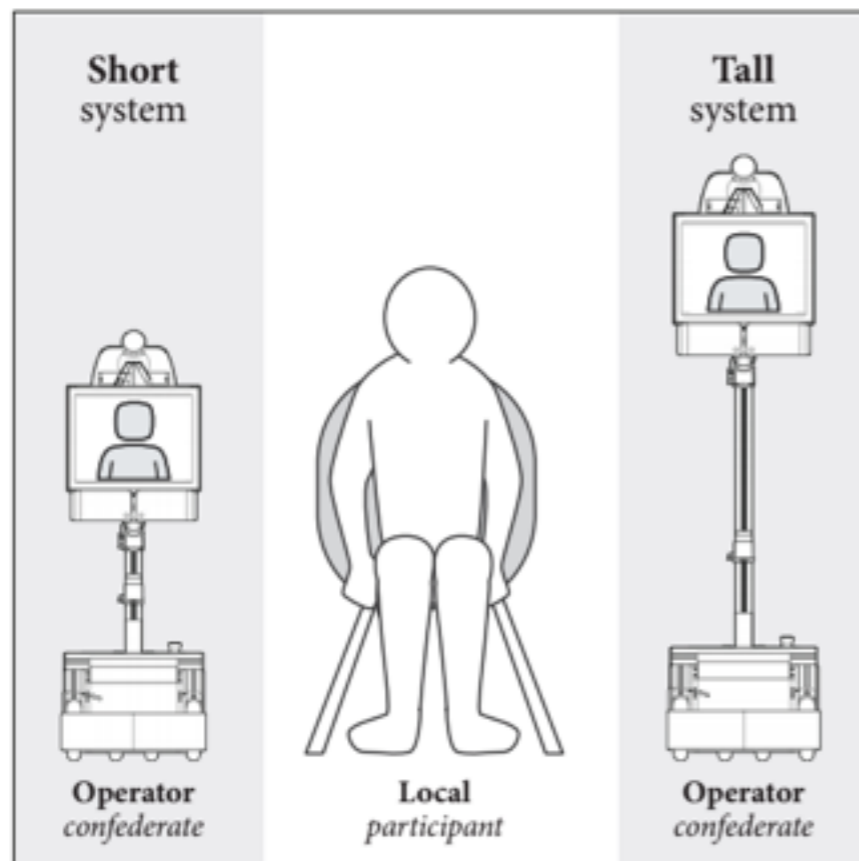
...



What is being measured?

- Example: **Ordinal** dependent variable
 - What is the effect of height on user preference?

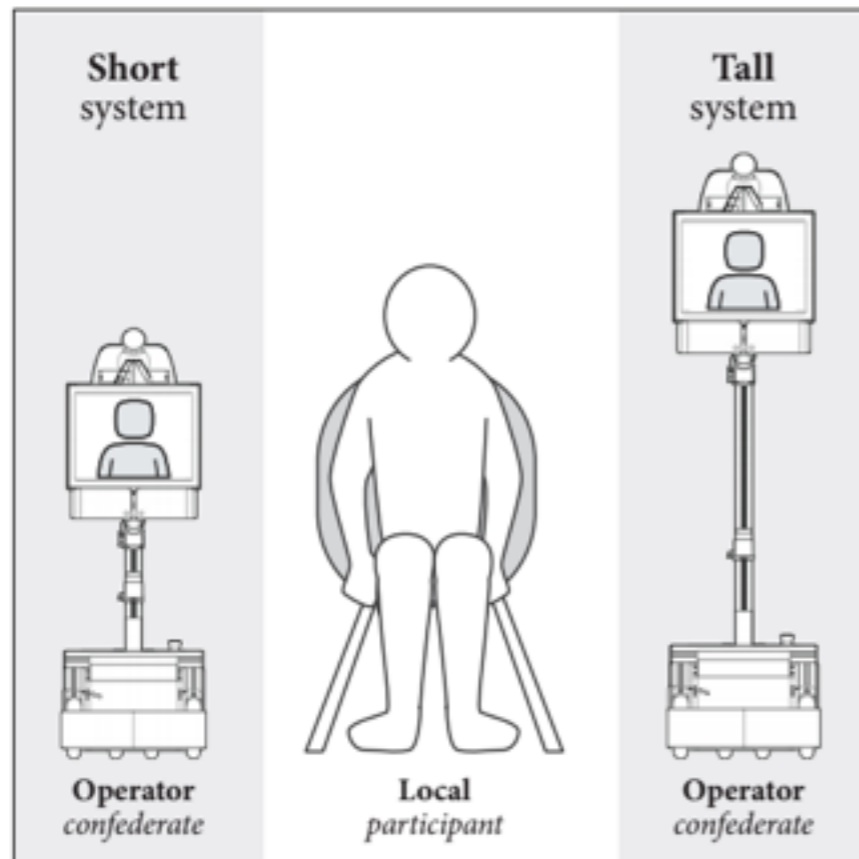
-user rating of the system



★☆☆☆☆	Hated it.
★★☆☆☆	Hated it.
★★★☆☆	Ashamed of liking it.
★★★★☆	Loved it.
★★★★★	Claimed to love it, but was actually a little bored.

What is being measured?

- Example: **Categorical** dependent variable
 - What is the effect of height on conversation control?

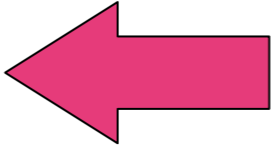


-choose one:

“I felt like the leader”

“I felt like the follower”

Designing an empirical study

- What is being compared?
 - Independent variables
- What are they being compared in?
 - Dependent variables (“metrics”)
- What (else) is being varied? 
- (What is kept constant?)
 - Extraneous variables

Extraneous variables

- Similar to independent variables but we are not looking for an effect
 - What is the effect of _____ on conversation control?
 - things that vary unless you control for them
gender, age, background of participants
 - things that you explicitly vary to demonstrate lack of effect
tasks performed using the system

Interpreting the results

- What is being compared?
 - Independent variables
- What are they being compared in?
 - Dependent variables (“metrics”)

Interpreting the results

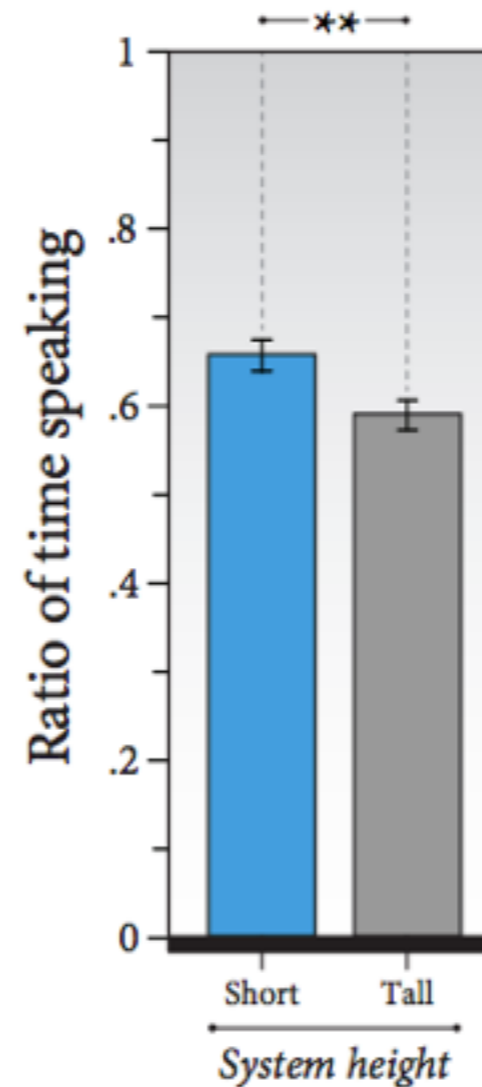
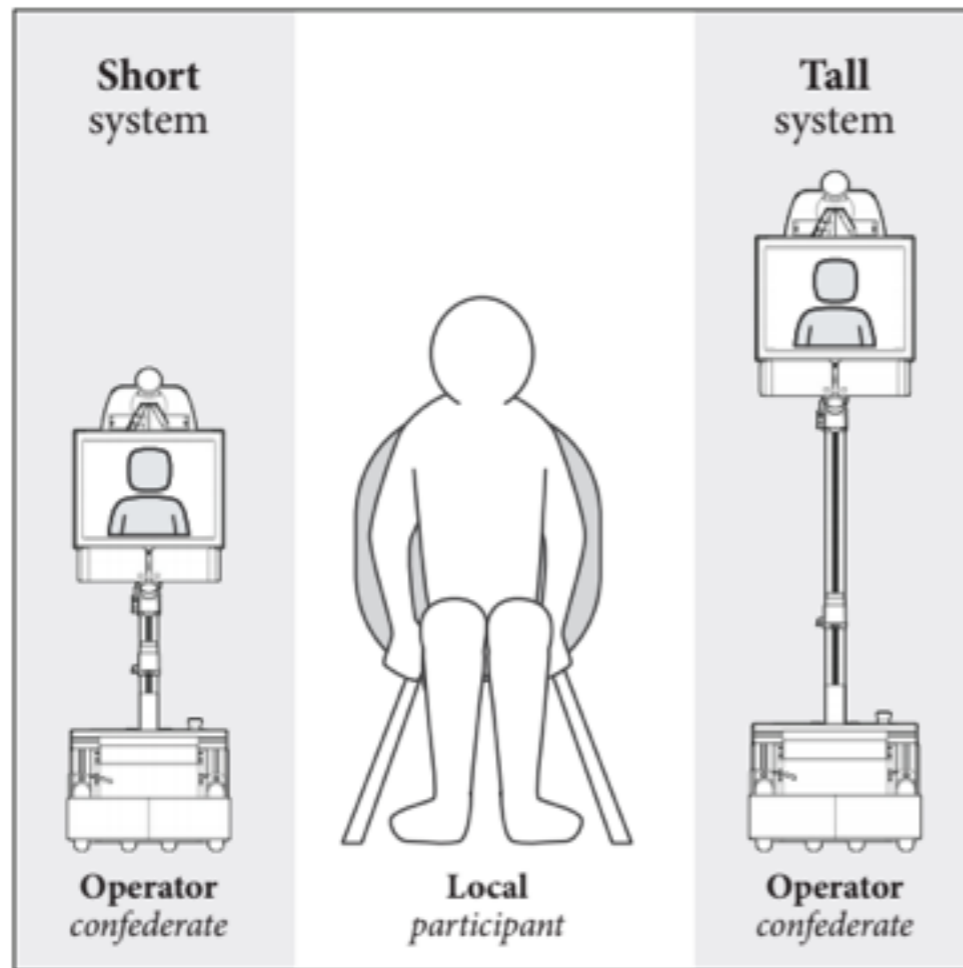
- What is being compared?
 - Independent variables
- What are they being compared in?
 - Dependent variables (“metrics”)

Main question:

Does <independent variable> cause differences in <dependent variable>?

Interpreting the results

Does **height** effect **ratio of time speaking**?



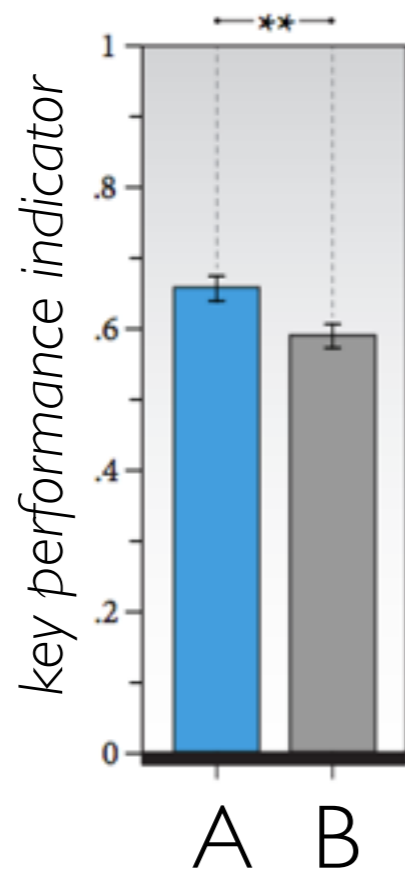
Yes/No?

Analyzing the data

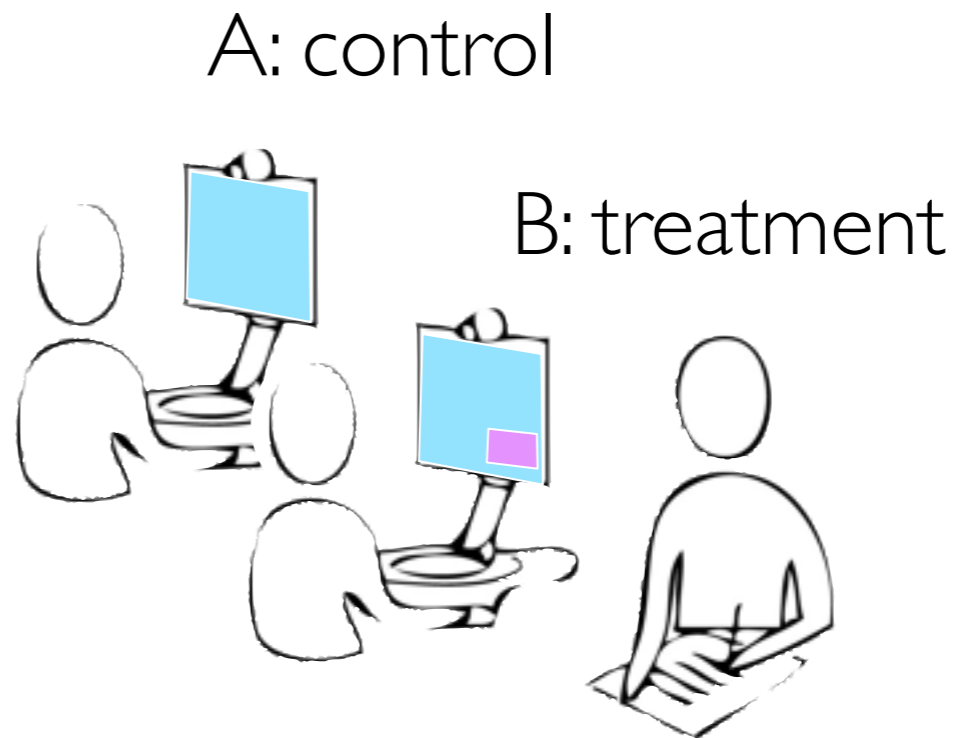
- Factors
 - Within vs. between groups
 - Number of variables
 - Type of dependent variables
 - Type of independent variables

A common case: A/B testing

- Two categorical independent variables (A vs. B)
- One interval dependent variable
 - key performance indicator



T-Test



(Student's) T-tests

- Check if two means (averages) are reliably different from each other
 - $t = (\text{variance between groups}) / (\text{variance within groups})$
 - Large t means **different groups**
 - Small t means **similar groups**

(Student's) T-tests Example

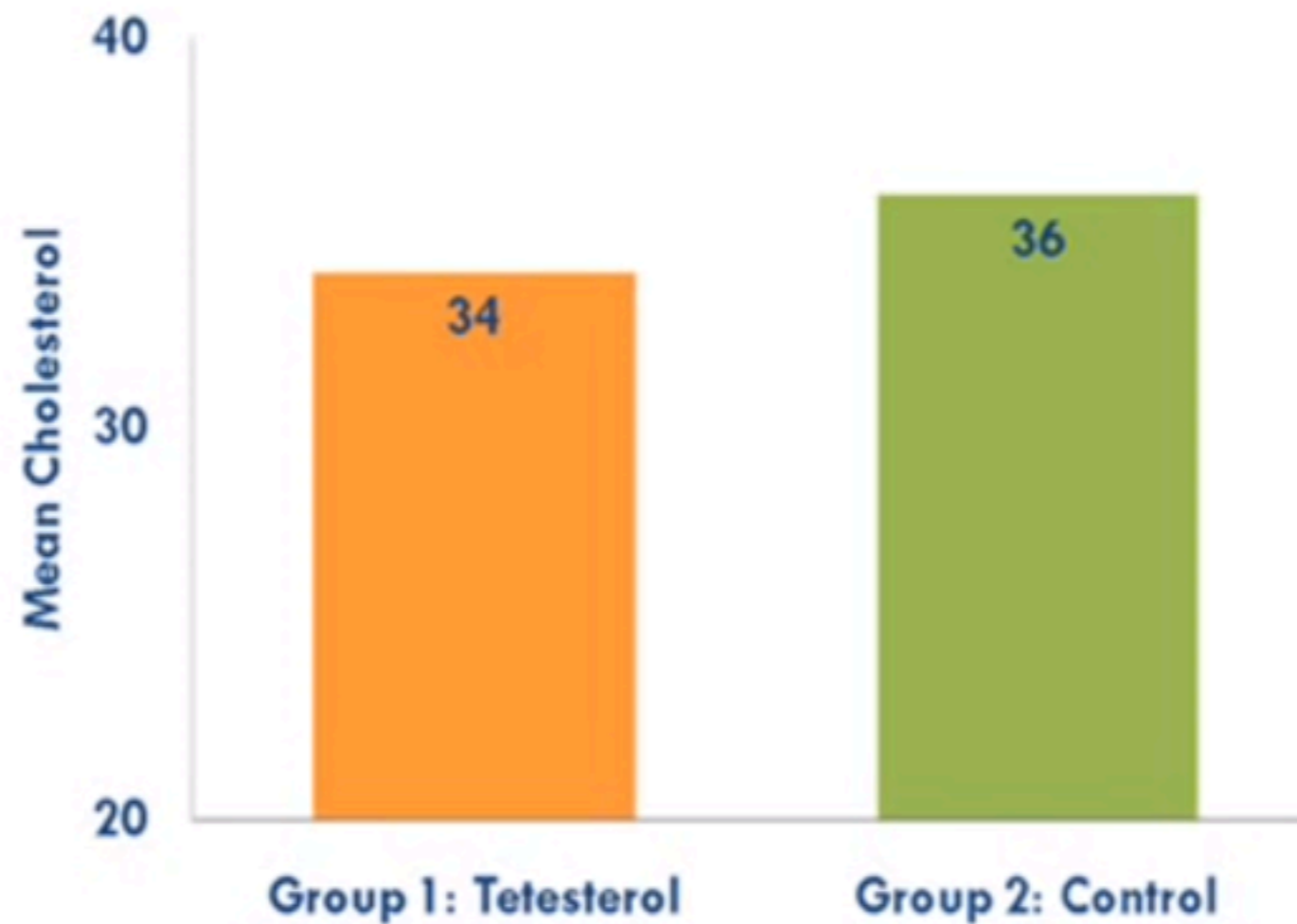
Tetesterol



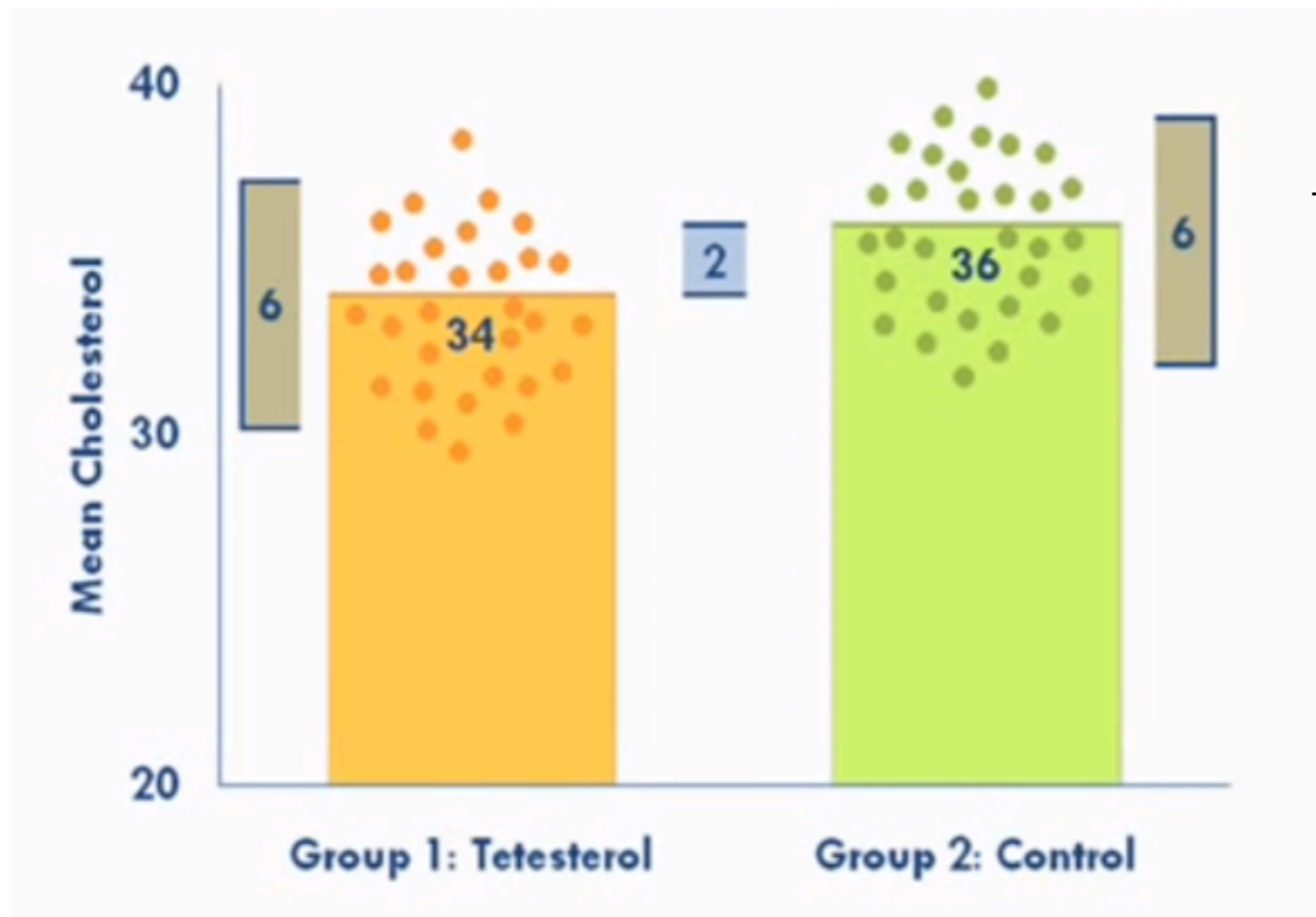
Control



(Student's) T-tests Example



(Student's) T-tests Example



$$t = 2/6$$

(Student's) T-tests

p-value: probability that our data could be produced randomly

(Student's) T-tests

p-value: probability that our data could be produced randomly

$$p < 0.05$$

(Student's) T-tests

p-value: probability that our data could be produced randomly

$$p < 0.05$$

This means that there is only a 5% chance that there is no real difference between the two groups.

(Student's) T-tests

p-value: probability that our data could be produced randomly
–depends on number of participants



With two groups of 5,
when $t = 2.0$, $p = .04$.



With two groups of 10,
when $t = 2.0$, $p = .03$.

(Student's) T-tests

p-value: probability that our data could be produced randomly
–depends on number of participants



With two groups of 5,
when $t = 2.0$, $p = .04$.



With two groups of 10,
when $t = 2.0$, $p = .03$.

Types of t-tests



“independent”
“unpaired”
“between samples”



“dependent”
“paired”
“within subjects”
“repeated measures”

Limitations of t-tests

- Generalizes to similar population
- Assumes that your data has Normal (Gaussian) distribution
- Sample size should be roughly the same
- All data should be independent/ not influenced by each other
- Interval type variables (will not work for rankings)

Lots of statistical tools available

1. Choose data entry format

- Enter up to 50 rows.
- Enter or paste up to 2000 rows.
- Enter mean, SEM and N.
- Enter mean, SD and N.

Caution: Changing format will erase your data.

2. Enter data

[Help me arrange the data.](#)

Label:

	Group 1	Group 2
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

3. Choose a test

- Unpaired t test.
 - Welch's unpaired t test (used rarely).
 - Paired t test.
- [Help me decide.](#)

4. View the results

Calculate now

Clear the form

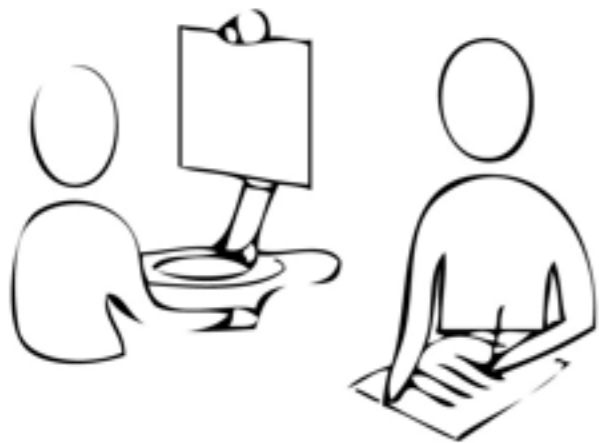
<http://www.graphpad.com/quickcalcs/ttest1.cfm>

Which statistical test to use?

1 IV with 2 levels (independent groups)	interval & normal	2 independent sample t-test
	ordinal or interval	Wilcoxon-Mann Whitney test
	categorical	Chi-square test Fisher's exact test
1 IV with 2 or more levels (independent groups)	interval & normal	one-way ANOVA
	ordinal or interval	Kruskal Wallis
	categorical	Chi-square test
1 IV with 2 levels (dependent/matched groups)	interval & normal	paired t-test
	ordinal or interval	Wilcoxon signed ranks test
	categorical	McNemar

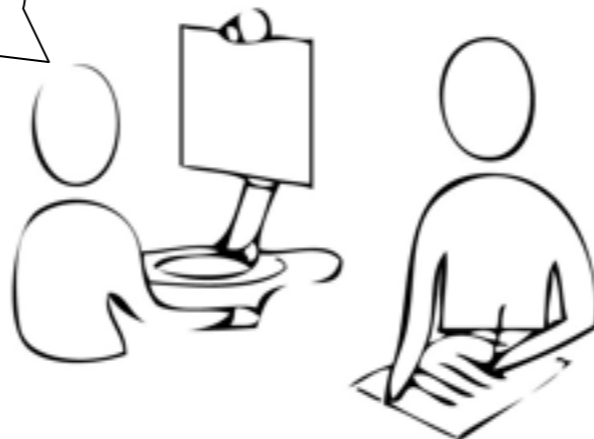
1 IV with 2 or more levels (dependent/matched groups)	interval & normal	one-way repeated measures ANOVA
	ordinal or interval	Friedman test
	categorical	repeated measures logistic regression
2 or more IVs (independent groups)	interval & normal	factorial ANOVA
	ordinal or interval	ordered logistic regression
	categorical	factorial logistic regression
1 interval IV	interval & normal	correlation
	interval & normal	simple linear regression
	ordinal or interval	non-parametric correlation
	categorical	simple logistic regression

Comparisons in observational studies

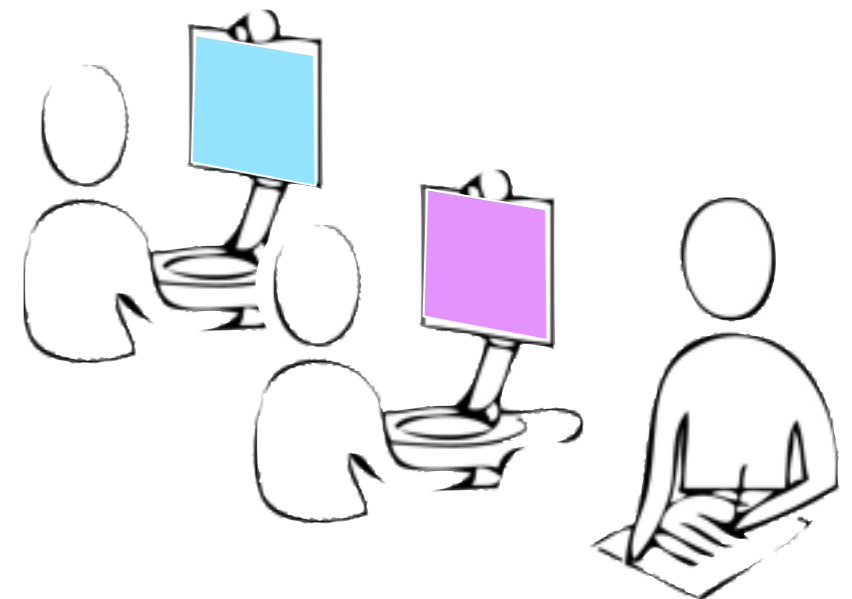


Observational study

hmmmm
blah blah
blah bla



Think-aloud protocol

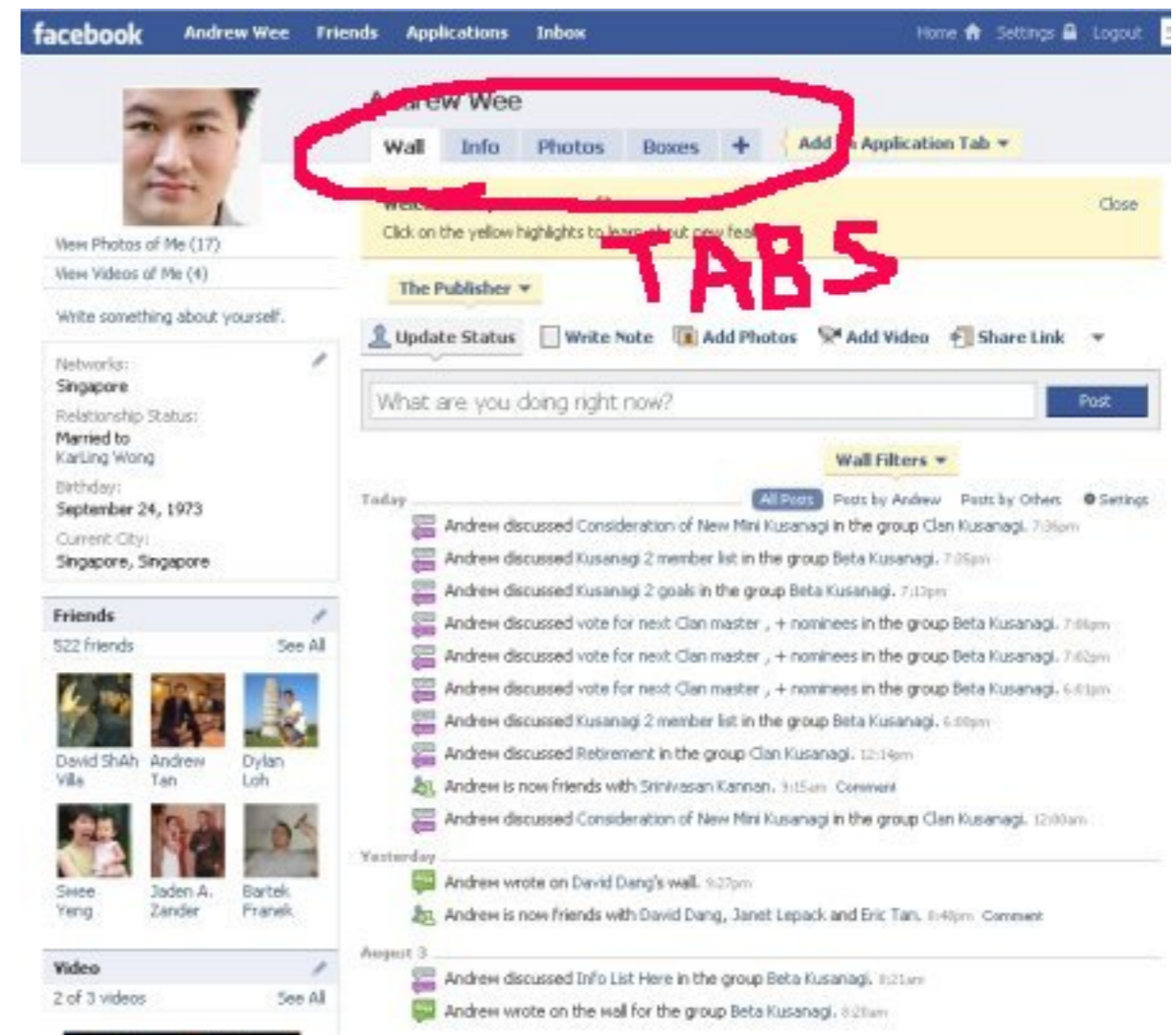


Comparative study



Post-hoc analysis

A/B testing



A/B testing example

The screenshot shows the Amazon.com shopping cart interface. At the top, it says "amazon.com" and "Hello, James Landay. We have recommendations for you. (Not James?)". Below this are links for "James's Amazon.com", "Today's Deals", "Gifts & Wish Lists", and "Gift Cards". A search bar is visible with "All Departments" selected. The cart is for "James Landay" and has a subtotal of \$219.99. The main item in the cart is "Microsoft Windows 7 Ultimate Upgrade" with a price of \$199.99. Below the cart items, there is a section titled "About the Shopping Cart" with three bullet points: "Items in your Shopping Cart always reflect the most recent price displayed on their product", "Items remain in your Shopping Cart for 90 days.", and "Learn more about the Shopping Cart and how to buy items at Amazon.com". At the bottom, there is a section titled "Customers Who Bought Items in Your Shopping Cart Also Bought" which displays three recommended products: "Microsoft Windows 7 Professional Upgrade" for \$199.99, "Microsoft Windows 7 Home Premium Upgrade" for \$119.99, and "Mac OS X versio Snow Leopard" for \$25.00 with a 4.5-star rating from 58 reviews.

A: No recommendations at checkout

B: Recommendations based on cart content

Pro: cross-sell more items

Con: distract people at check out

B wildly successful!

A/B testing example

A
Solitaire

Windows Marketplace Find. Try. Buy.

Downloads All Software Hardware
Top Sellers Experience Vista Game Downloads Security Downloads IE Add-Ons

Search: Windows Marketplace Go

Downloads > Game Downloads > Card & Casino Game Downloads

Narrow Your Selection

By Cost

- Free - \$9.00
- \$10.00 - \$19.00
- \$20.00 - \$29.00
- \$30.00 - \$49.00
- \$50.00 - \$69.00
- \$70.00 - \$129.00

By Manufacturer

- DreamQuest
- DeepNet Technologies
- Havareark Software

Card & Casino Game Downloads

Solitaire Master 2
by Egames
Solitaire Master 2 contains 300 great games including Klondike, Free Cell, Deuces Queens, Cats Cradle, Four Leaf Clover, Spider Web and many more! Plu...
Price: \$12.99
ADD TO CART

Spotlight on

Championship Gin Pro Card Game for Windows XP
Read reviews | Download now
Price: \$20.00

Tight Shark Hold'em Advisor
Read reviews | Download now
Price: \$25.00

B
Poker

Windows Marketplace Find. Try. Buy.

Downloads All Software Hardware
Top Sellers Experience Vista Game Downloads Security Downloads IE Add-Ons

Search: Windows Marketplace Go

Downloads > Game Downloads > Card & Casino Game Downloads

Narrow Your Selection

By Cost

- Free - \$9.00
- \$10.00 - \$19.00
- \$20.00 - \$29.00
- \$30.00 - \$49.00
- \$50.00 - \$69.00
- \$70.00 - \$129.00

By Manufacturer

- DreamQuest
- DeepNet Technologies
- Havareark Software

Card & Casino Game Downloads

5 Card Draw Poker
by Andrei Jurko
Five card draw is one of the most popular poker variation around the world. This software features smart AI players, nice graphics and sound. Have fun...
Price: \$9.95
ADD TO CART

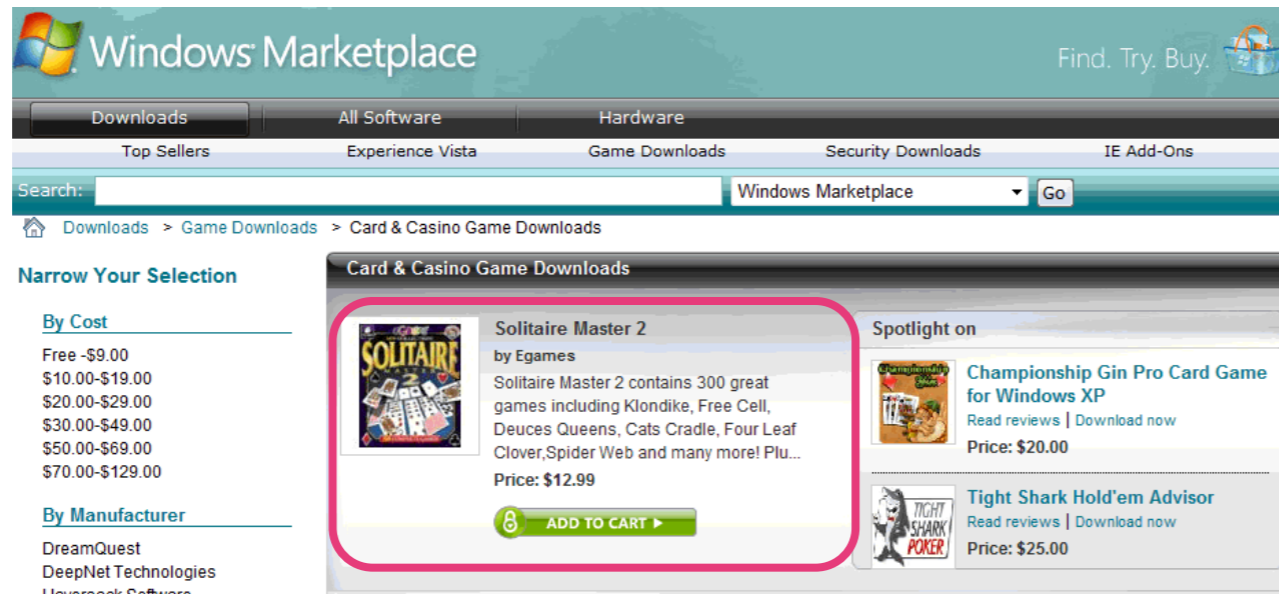
Spotlight on

Championship Gin Pro Card Game for Windows XP
Read reviews | Download now
Price: \$20.00

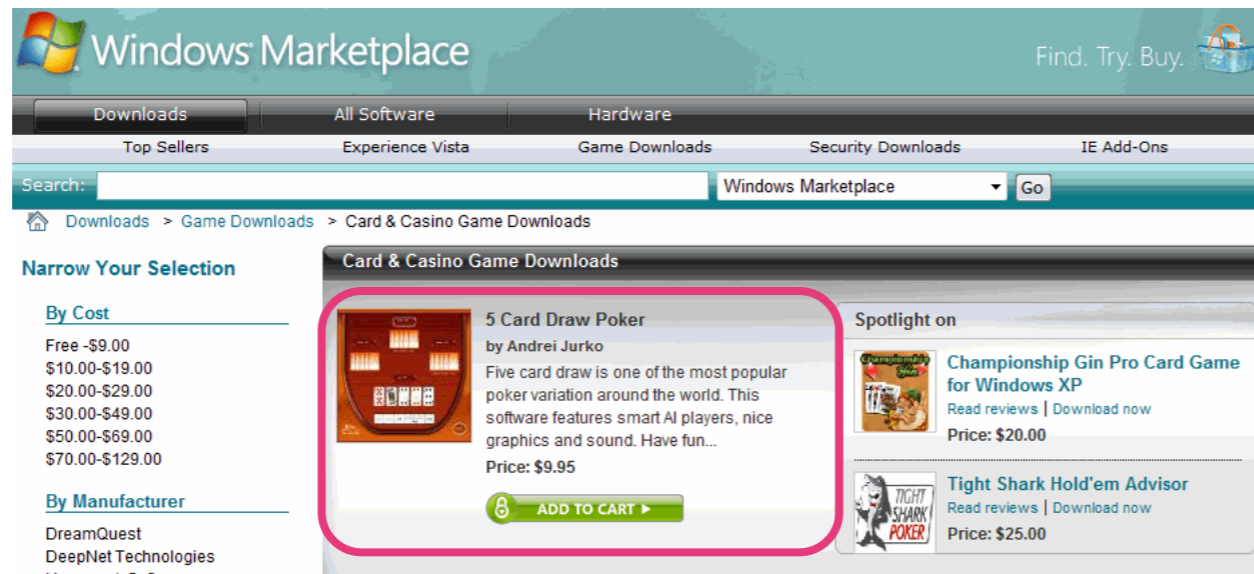
Tight Shark Hold'em Advisor
Read reviews | Download now
Price: \$25.00

A/B testing example

A
Solitaire



B
Poker



A is 61% better!

A/B testing example

A

Ask why
by default

Please let us know if this content was helpful.

Rate this content:



Tell us why you rated the content this way (optional):

Remaining characters: 650

Submit

B

Ask why
if user
gives rating

How helpful was this information?

Click a star.

Not helpful  Very helpful

Click to rate: 3 out of 5 stars



How helpful was this information?

Click a star.

Not helpful  Very helpful

Why did you rate the information this way?

Remaining characters: 650

Submit

A/B testing example

A

Ask why
by default

Please let us know if this content was helpful.

Rate this content:



Tell us why you rated the content this way (optional):

Remaining characters: 650

Submit

B

Ask why
if user
gives rating

How helpful was this information?

Click a star.

Not helpful  Very helpful

Click to rate: 3 out of 5 stars



How helpful was this information?

Click a star.

Not helpful  Very helpful

Why did you rate the information this way?

Remaining characters: 650

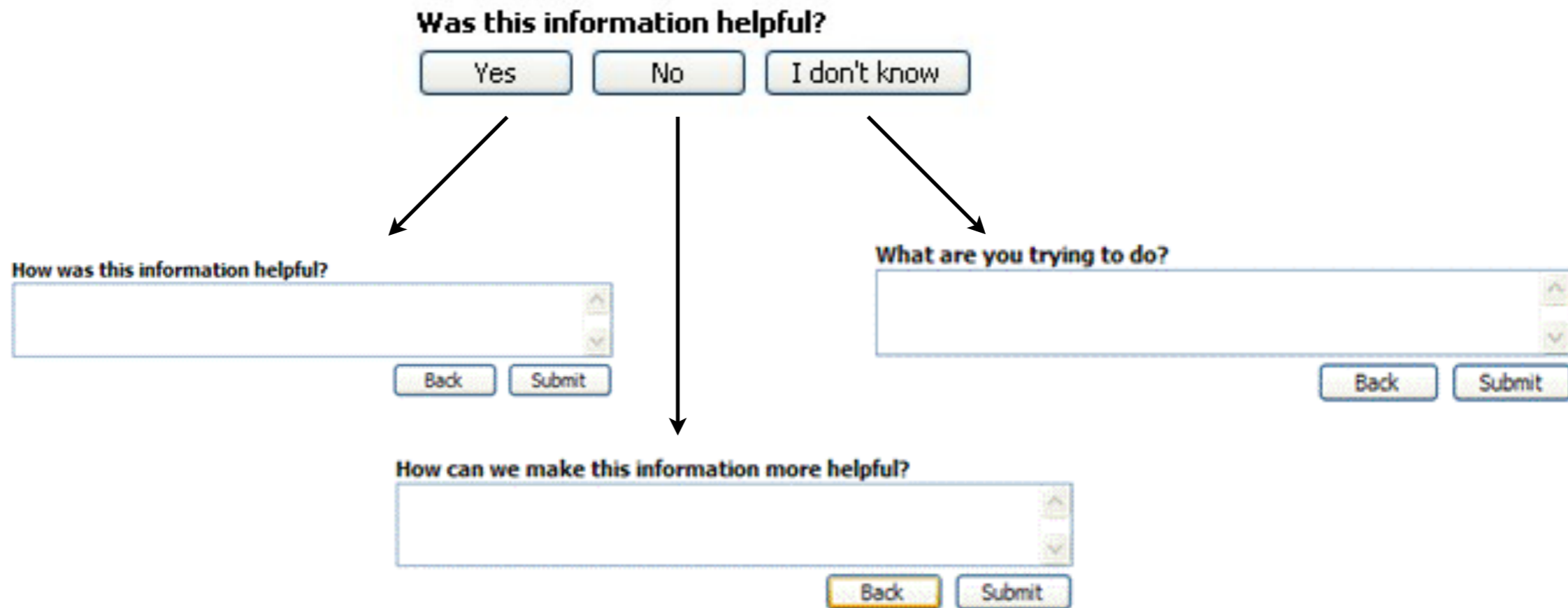
Submit

*More than double
response rate!*

A/B testing example

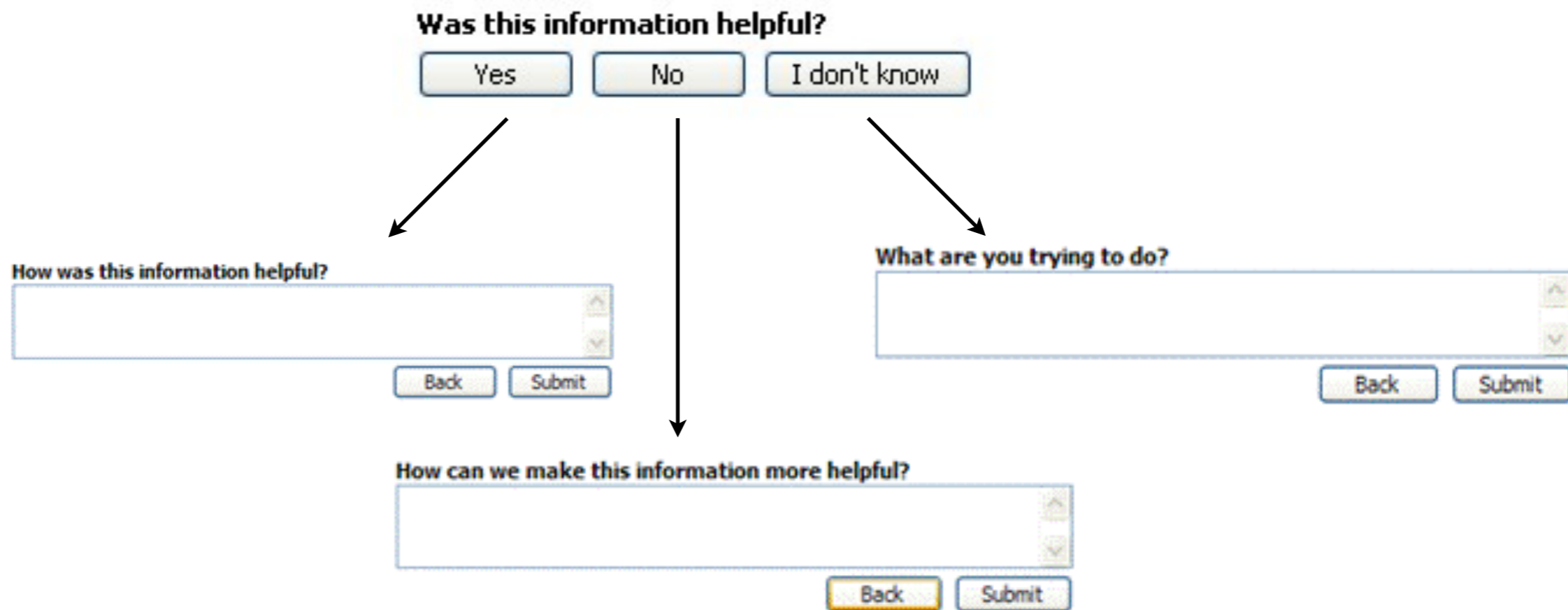
C

Ask a different question based on step 1



A/B testing example

C
Ask a
different
question
based on
step 1



*C outperforms B
by a factor of 3.5!*

Limitations of A/B testing

- Hill climbing, will not re-invent anything