

Types of Validity

By the next break, you will be able to identify and distinguish the four types of validity.

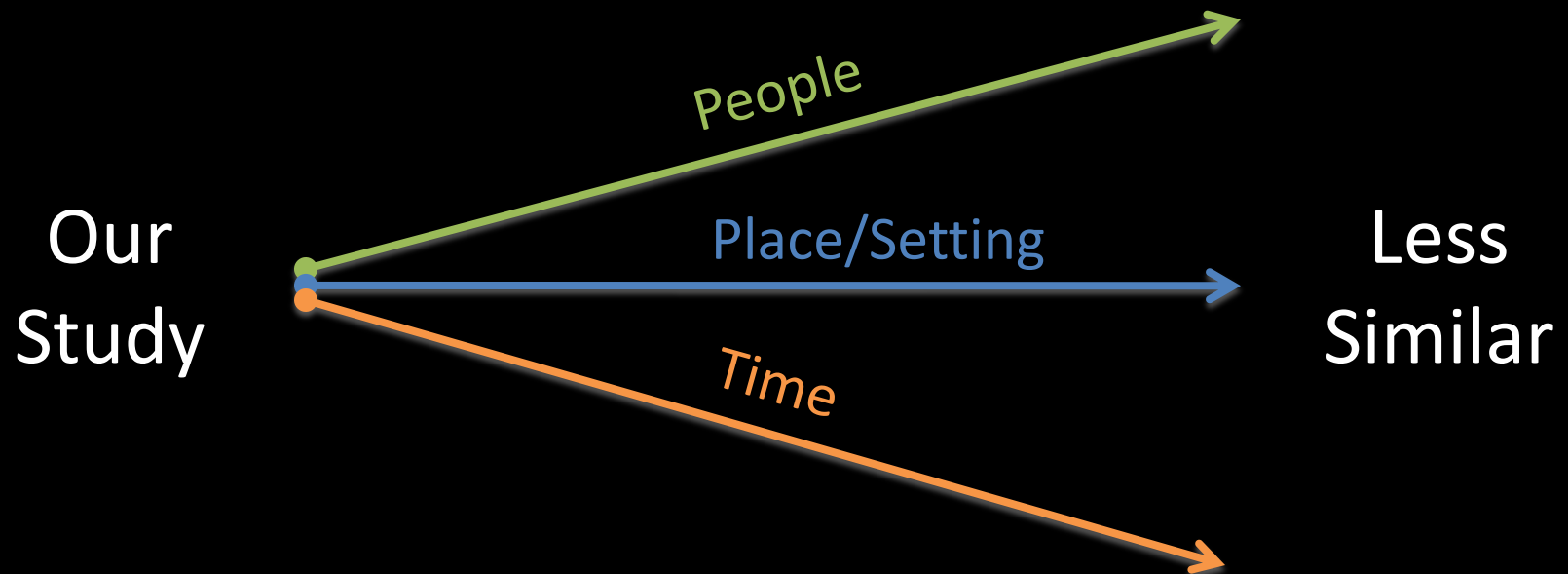
What makes a study valid?

Types of Validity

- External
- Construct
- Internal
- Conclusion

External Validity

- Generalizability
- Sample → Population



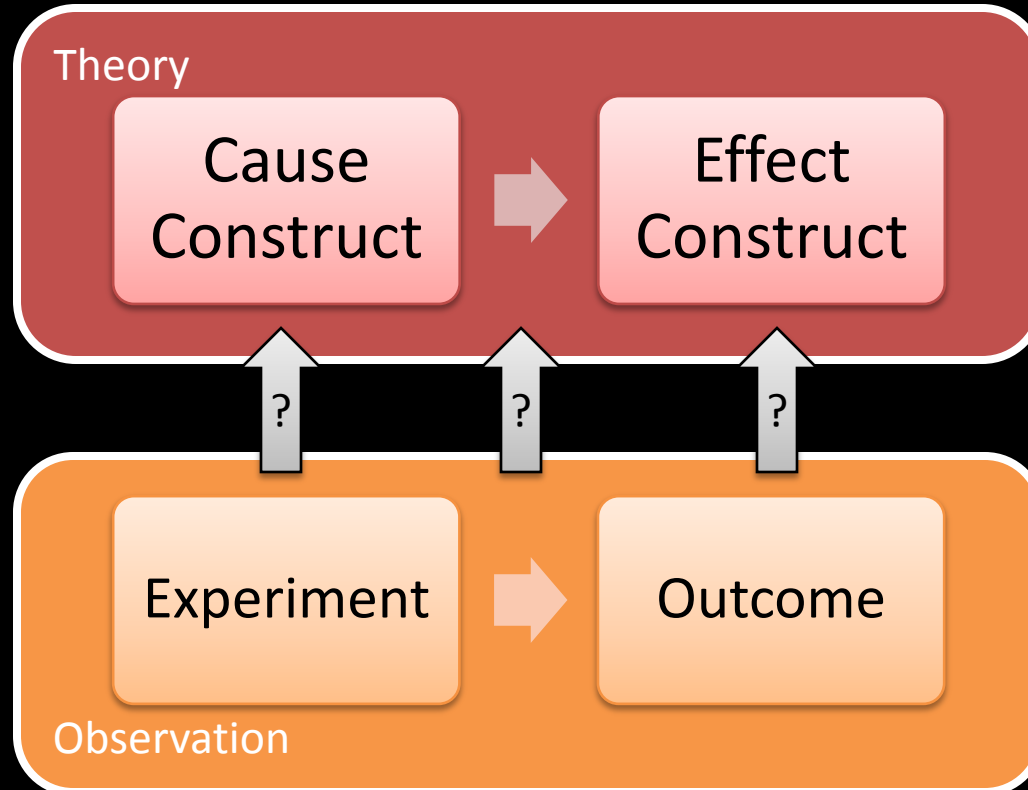
Example

- You perform a study on 12 university students. Each participant is asked to find three unique insights about the data.
- How similar would this population need to be to reproduce your results?

Construct Validity

- Do your *observations* correspond to the *theory* you are using to describe them?
- One interpretation: do you have the right labels?

Construct Validity



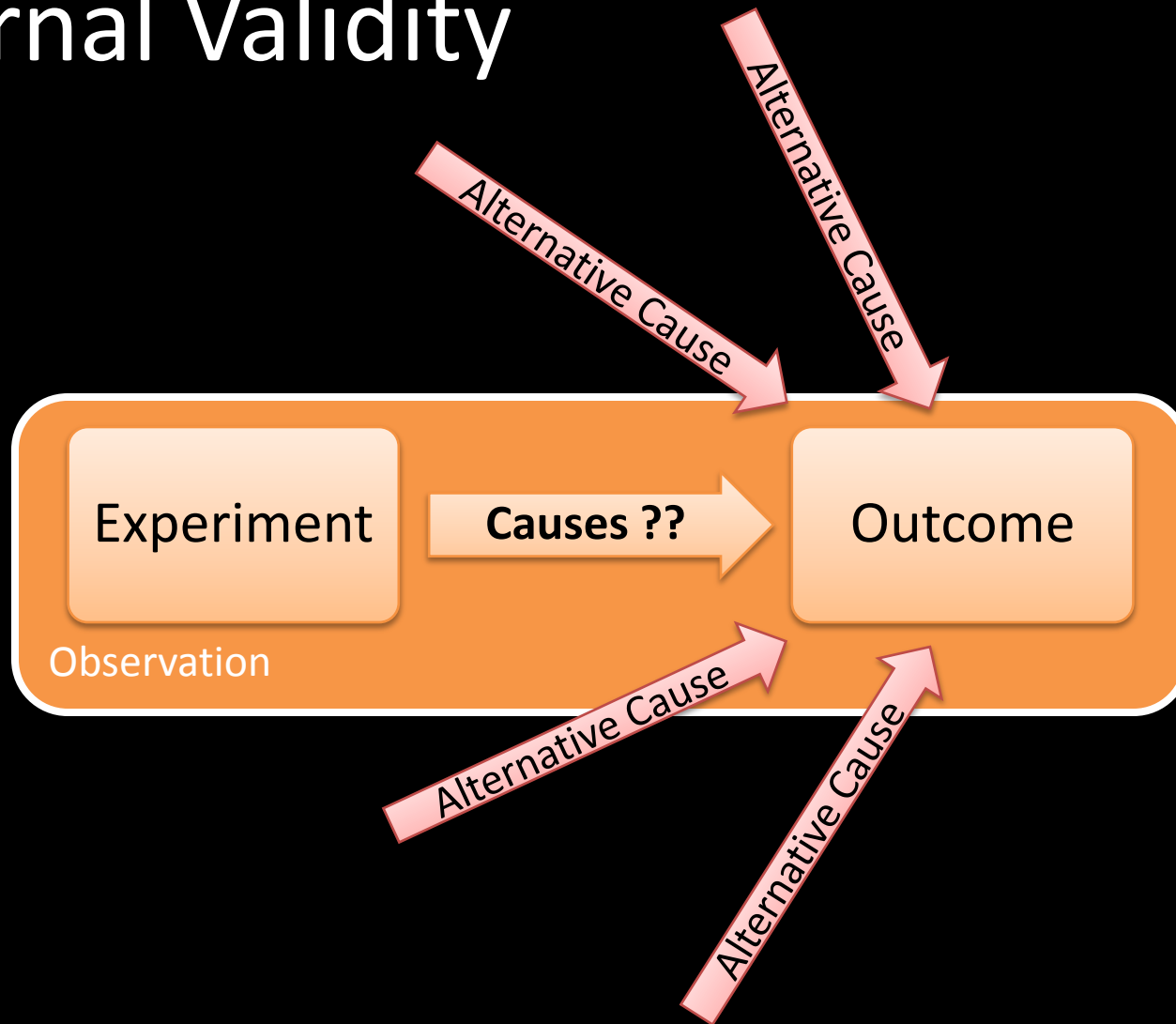
Example

- Our theory states that our new *type* of visualization will lead to faster discovery of *insight*.
 - Cause construct = type of visualization
 - Effect construct = speed of discovering insight
- Do your experiment and observations correspond to these *constructs*?

Internal Validity

- Can the *observed* changes be attributed to the factors *you* manipulated?
- Is there some alternative cause?
- Note: only concerned with what happened in *your study*!

Internal Validity



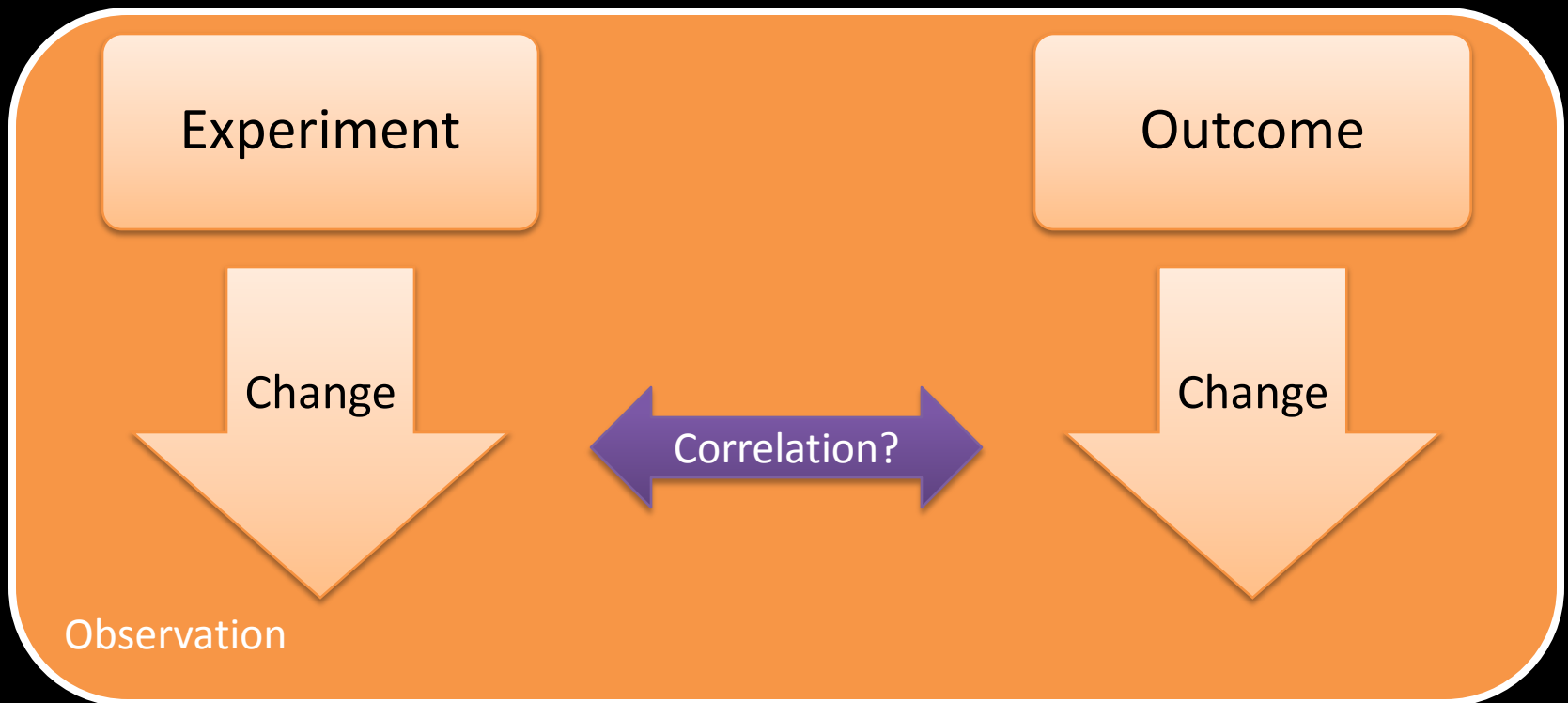
Example

- Study: compare BabelFish (a translator) to lattice uncertainty visualization (LUV).
- Observe: people who use LUV are more confident about their interpretation.
- Did the change in technique *cause* the observed change in confidence *in your study*?
- Is there another possible explanation?

(Statistical) Conclusion Validity

- Is the conclusion we make about the *relationship* between the independent and dependent variables valid?
- Not concerned with cause, only correlation
- Is our analysis correct?

(Statistical) Conclusion Validity



Example

- Our analysis revealed that there was a significant main effect of visualization technique ($F(\dots, \dots) = \dots, p < .05$).
- Is it reasonable to reach the conclusion that (in our study) changing the visualization technique is related to a change in the dependent variable?

Types of Validity

- Conclusion
 - Is there a *relationship*?
- Internal
 - Is the relationship *causal*?
- Construct
 - Can we *generalize* to the *constructs* (theory)?
- External
 - Can we *generalize* to other *people/places/times*?

Activity (four groups, 5 minutes)

- Conclusion
 - Is there a *relationship*?
- Internal
 - Is the relationship *causal*?
- Construct
 - Can we *generalize* to the *constructs* (theory)?
- External
 - Can we *generalize* to other *people/places/times*?

What form of validity?

- My theory states that people who spend all day typing have weaker wrists than those that don't.
- I measure how far two groups (typists and non-typists) can throw a Frisbee.
- Does what I observe in my study correspond to my theory about typists?

What form of validity?

- A longitudinal study on working habits was performed to measure the effect of working long hours on success.
- The study showed that people who worked long hours tended to be more successful.
- Is the conclusion that working long hours leads to success valid?

What form of validity?

- Two Mac users and two windows users were asked to rate their operating system on a scale of 1 (terrible) to 9 (fantastic). Results of a <analysis?> showed that people preferred Mac OS X to Windows.
- Were there enough people in this study to claim a significant result?

What form of validity?

- A study of 12 computer science students was performed to compare three 3D interaction techniques. Results showed that a 10-button mouse outperformed the arrow keys on a keyboard.
- Would this result be the same if the study was performed on 12 architects?

Summary

- Conclusion
 - Is there a *relationship*?
- Internal
 - Is the relationship *causal*?
- Construct
 - Can we *generalize* to the *constructs* (theory)?
- External
 - Can we *generalize* to other *people/places/times*?

Break: 15 Minutes

Is our list of forms of validity exhaustive?
(Note: I called them “*the* four types”)

After the next 10 minutes, you will be able to distinguish between *external validity* and *ecological validity*.

Has anyone been criticised about the validity of an experiment they ran (e.g., in a review)?

Ecological Validity

- How closely does the *experimental setting* correspond to the *real setting*?

Compare

Ecological Validity

- How closely does the *experimental setting* correspond to the *real setting*?

External Validity

- Does what we observed in *our study* generalize to what would happen with *different people*, in a *different place*, or in a *different time*?

How can one happen without the other?

Example

- We perform a study that compares how quickly people can select menu items in a *circular menu* and a *rectangular menu*.
- The menus were filled with different types of fruit in a random order and asked to select a target fruit. Time to select targets was measured.

Activity (same group, 2 minutes)

- Come up with an example of a study that has *high ecological validity* and *low external validity*?

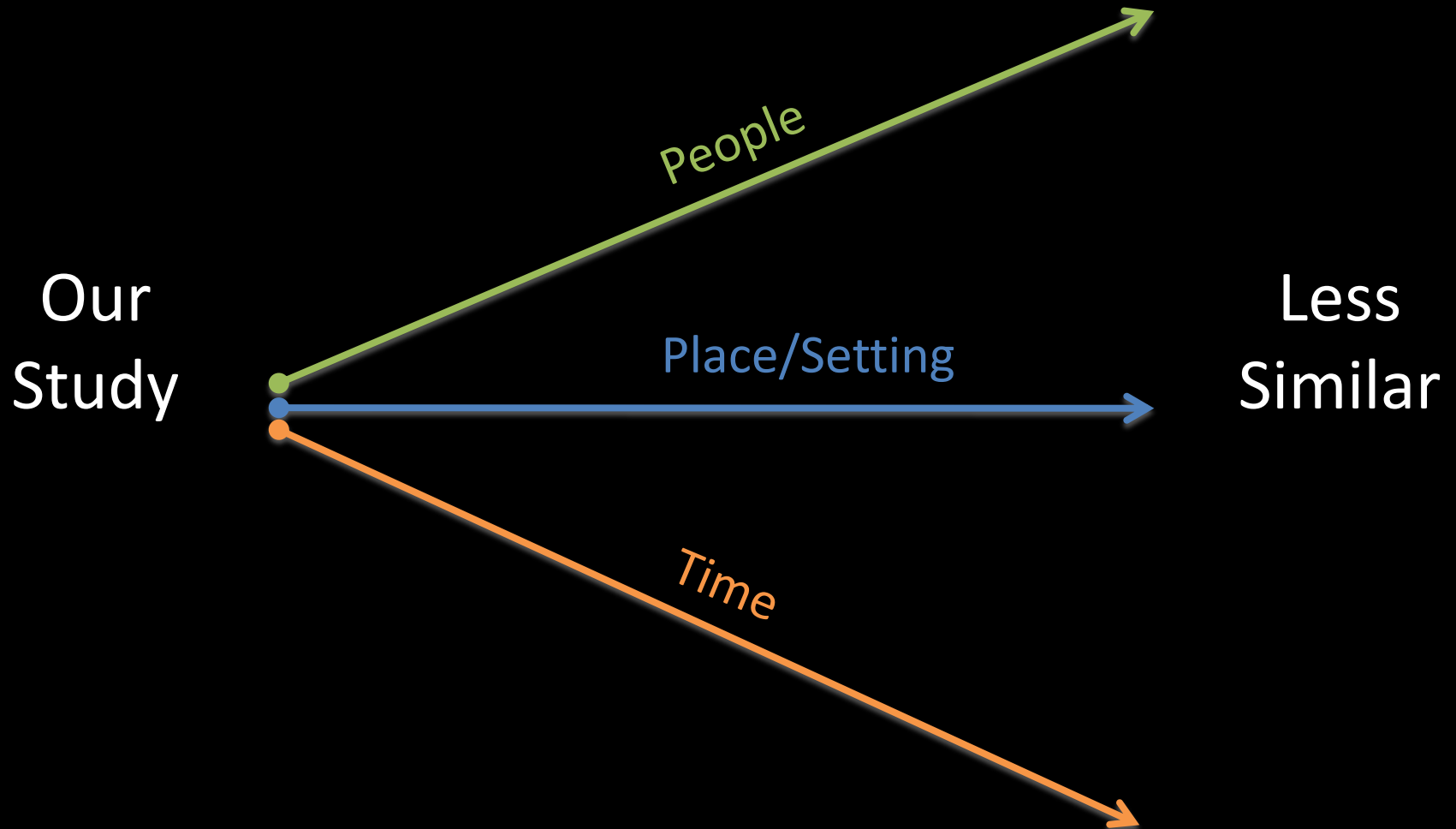
Summary

- Ecological validity
 - Does the experimental setting match its realistic counterpart?
- External validity
 - Can we generalize our results to other settings?

Threats to Validity

By the next break, you will be able to criticise a study according to the four types of validity.

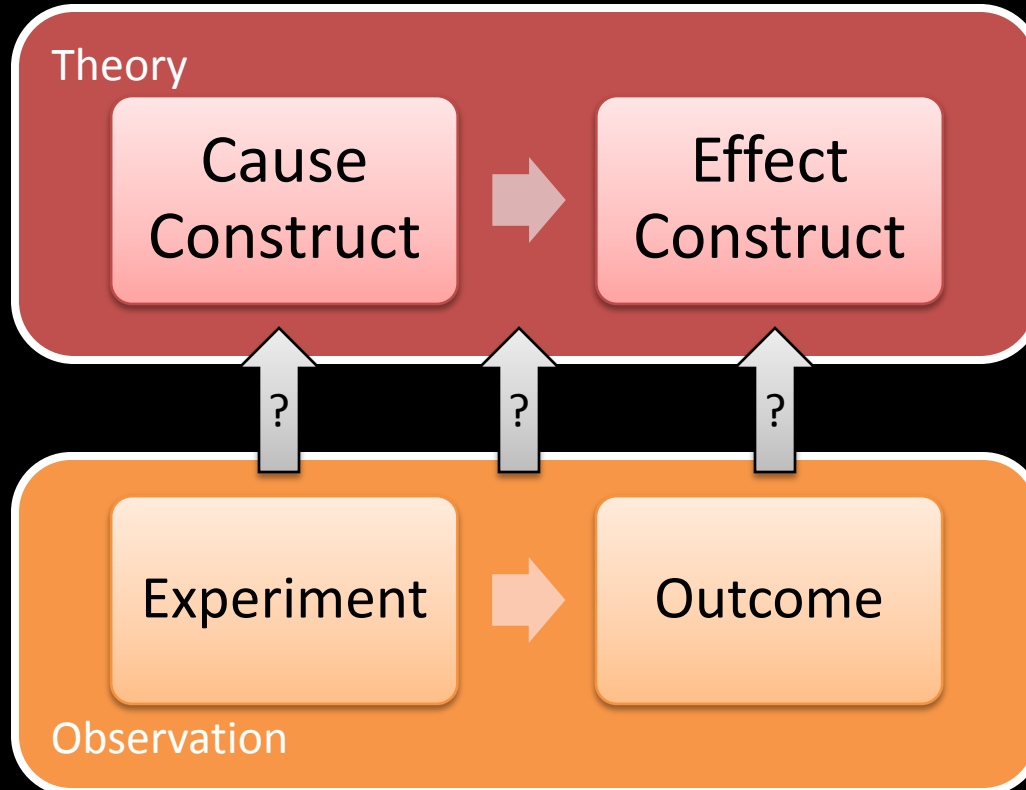
Threats to External Validity



Examples

- Criticism #1: you used only computer science students (people)
- Criticism #2: you performed the study in a lab setting (place)
- Criticism #3: you performed the study right after the Wii was released (time)

Threats to Construct Validity



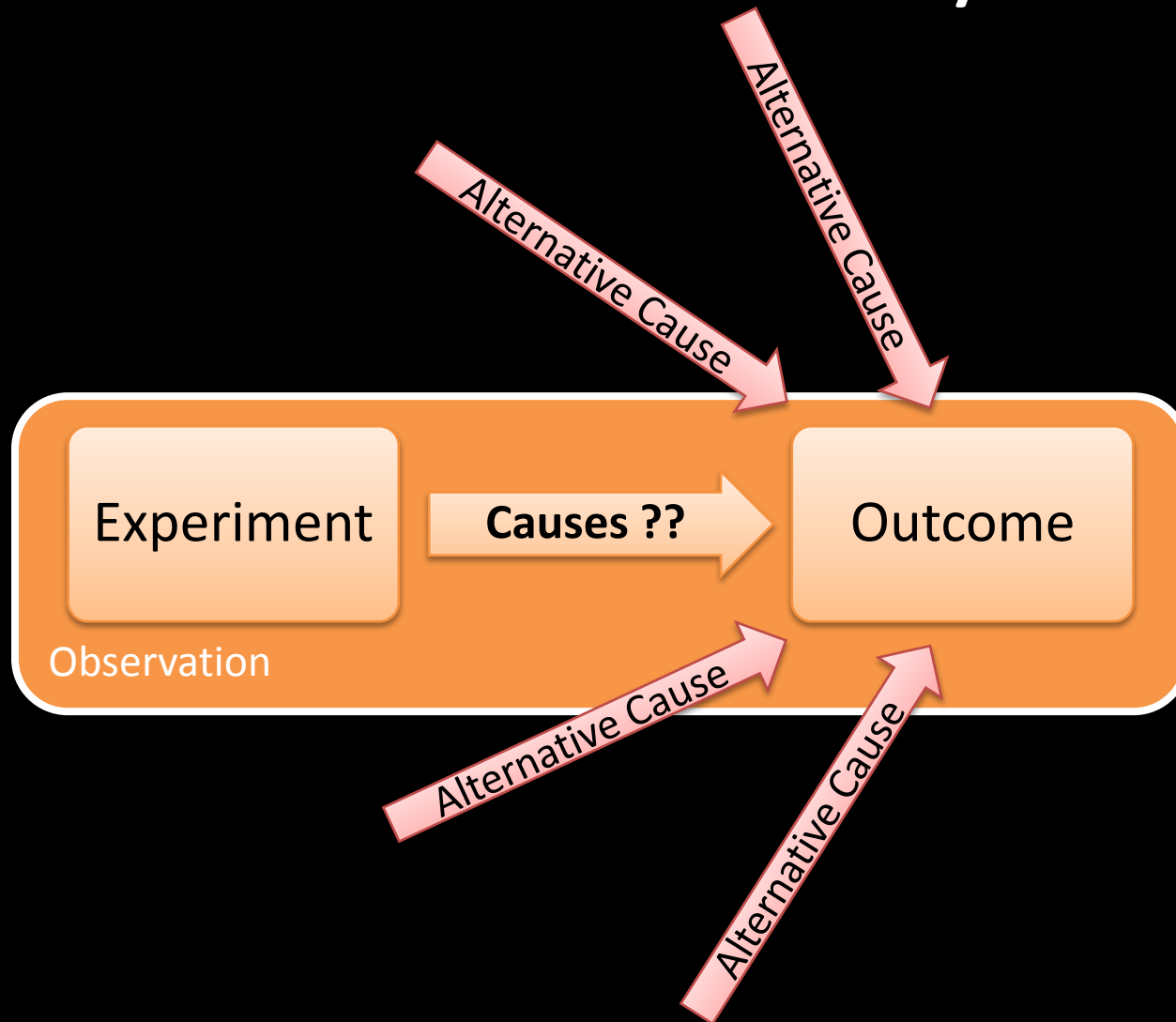
Threats to Construct Validity

- Poorly defined construct
- Only one representative:
 - cause construct (e.g., one multi-D vis.)
 - effect construct (e.g., one measure of “insight”)
- Interaction:
 - cause construct (e.g., combination of causes)
 - effect construct (e.g., experiment + cause)

Threats to Construct Validity

- Unintended consequences
 - e.g., label interaction technique as “effective” when it is faster, but has side effect of being less accurate
- Confound in Levels of Construct
 - e.g., conclude that use of “lenses” helps find targets, but only test with one lens.

Threats to Internal Validity



Threats to Internal Validity

- History Threat (e.g., Wii released)
- Maturation Threat (e.g., learning effect)
- Testing Threat
 - (e.g., pre-test: ask about table use)
- Instrumentation Threat (e.g., wear on device)
- Mortality Threat (e.g., people drop out)
- Regression Threat (e.g., novices get better)

Threats to Conclusion Validity

		Reality	
		H_0 true	H_0 false
Decision	H_0 true	✓	Type II
	H_0 false	Type I	✓

Threats to Conclusion Validity

- Type I Error:
 - Repeated tests (fishing)
- Type II Error:
 - Small sample size, small effect size
 - Noisy data: measurement error, experimenter error, setting changes (e.g., lighting), natural differences in people

Activity (same groups)

- What threat to validity lead to the invalidity in your previous examples?

Summary

- Threats to External Validity
 - People, place, or time
- Threats to Construct Validity
 - Incorrect labelling
- Threats to Internal Validity
 - Alternative explanations/causes
- Threats to Conclusion Validity
 - Type I and Type II errors

Break: 15 Minutes

Assignment 3

Experimental Design

By the end of this course (!), you will be able to design and analyse your own experiment.

Has anyone performed their own experiment
and analysis?

Method

- What is the problem?
- What is your hypothesis?
- How can you test your hypothesis?
 - What factors might be interesting?
 - What/how can you measure?
 - How can you avoid the threats to the four types of validity?

Activity/Discussion

- Design a study.

Summary of (Final) Day

- Four forms of validity
 - Conclusion, internal, construct, external
- Ecological validity
- Threats to Validity
- Experimental Design

Thanks for being a great class :)