Agenda

Today

- The basics of experimentation

Next Monday

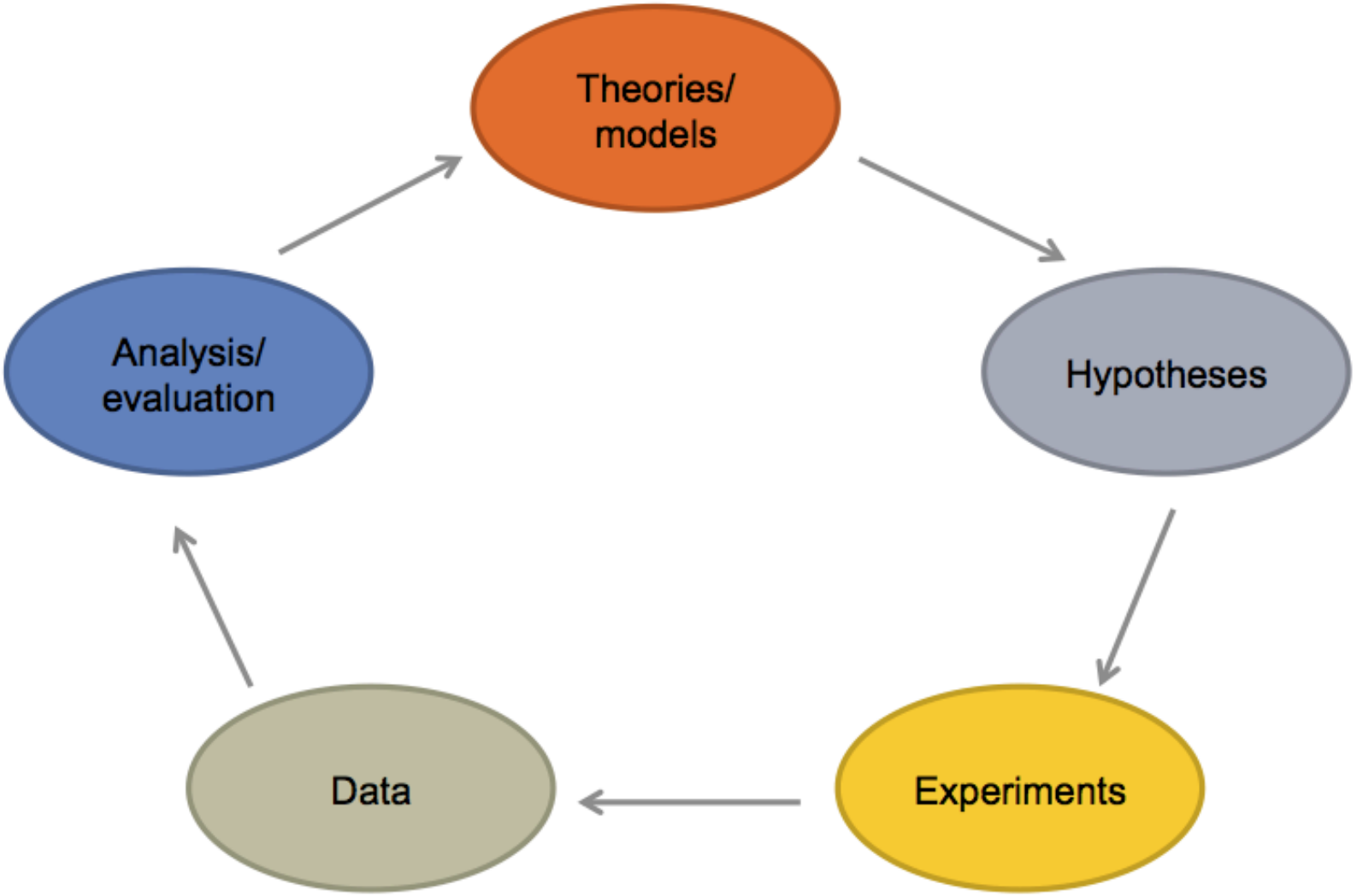
- Introduction to the course experiment
- **Have read**: Van Berkum et al. (2008)

Next Wednesday

- Methods and measures

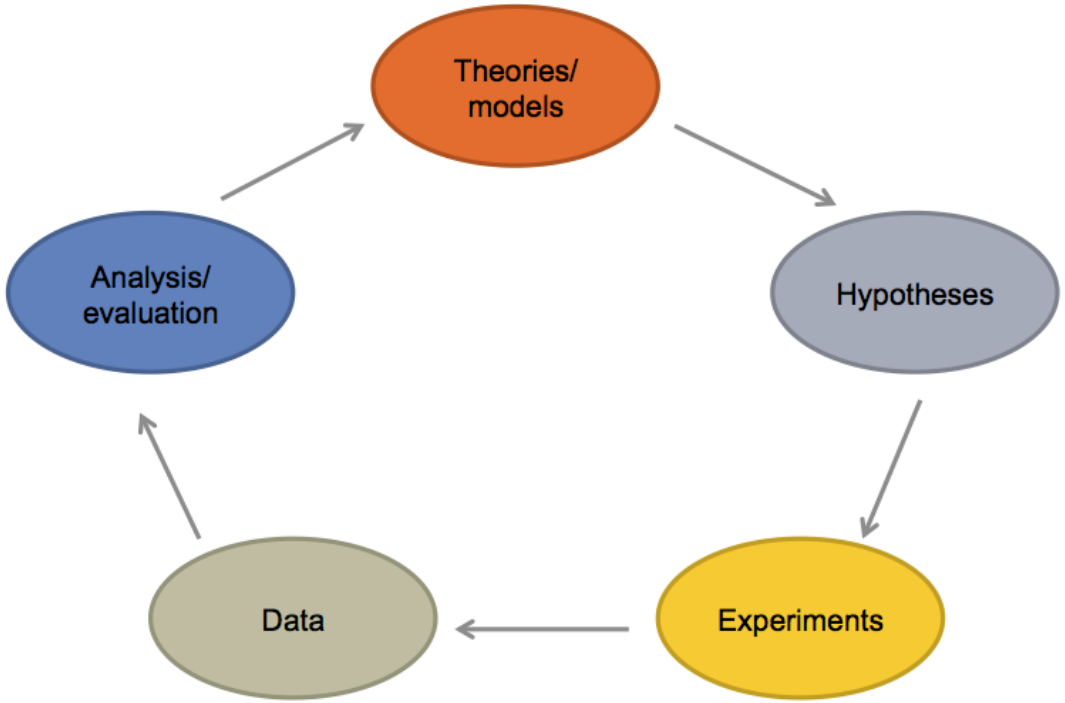
EXPERIMENTAL RESEARCH

Scientific Method



Scientific Method

1. Observation
2. Develop a theory or hypothesis
3. Generate a testable prediction
4. Make systematic observations
5. Evaluate the evidence
6. Repeat



Scientific Method

Step 1: Observation (Empiricism)

- Pay attention to the world around you and look for generalizations in a population's behavior
- Two classes of generalizations

Experimentation?

  - **Descriptive** – Just describe how it was, what was seen, how frequently it occurred, without making predictions ✖
  - **Cause and effect** – Make predictions about the observed relationship between two (or more) things based on an assumption of ***determinism***, that phenomena have identifiable causes ✔

Scientific Method

Step 2: Develop a theory or hypothesis

- Identify the ***variables*** associated with your observations
  (i.e., the characteristics of the behavior and the surrounding context)

- Come up with a plausible explanation for the observed behavior(s)
  - How are the variables related to one another?
  - Source of explanations may be based on past research, common sense, intuition, logic, etc.

Scientific Method

Step 3: Generate a testable prediction

- **Testable**: Must be specified in a way that can be assessed via observation
  - The relevant variables must be clearly defined and observable
- **Falsifiable:** Scientists don't try to prove theories, they try to ***refute*** theories
  - Refutable hypotheses must be stated in a way that allow the potential for it to be wrong
- **Focused**: Is the idea specified enough (***operationalized***) to be manageable
- **Replicable**: Can the results be reproduced? Or is it a "one time only" effect?

Scientific Method

Step 4: Make systematic observations

- Which variables will we examine?
- How do we measure these variables?
- Which variables can we systematically manipulate?
- What variables need to be controlled?
- Where (from whom) will we collect the observations?

Scientific Method

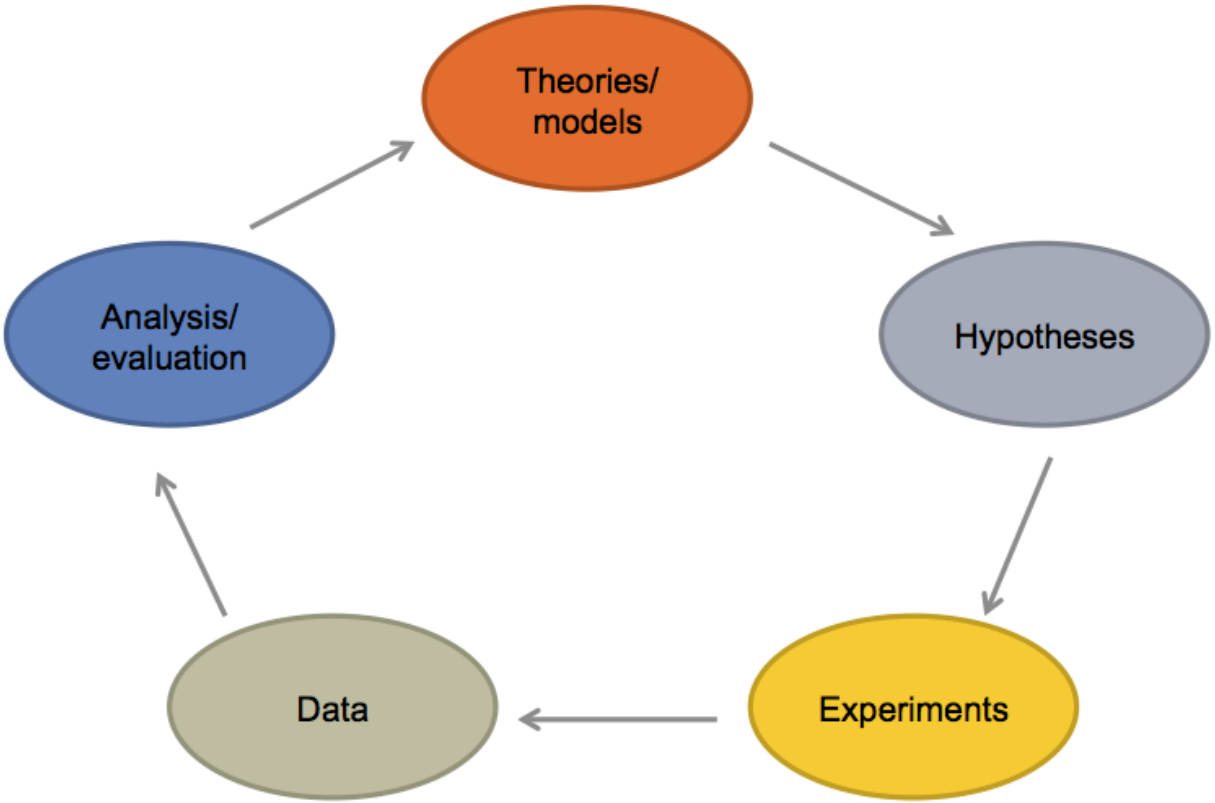
Step 5: Evaluate your evidence

- Your results might:
  - Refute the theory
  - Support the theory (but does NOT "prove" the theory)
  - Lead to revision of the theory
  - Lead to consideration of alternative theories

- There are *always* alternative explanations for the results
  - ***Parsimony***: Simpler explanations are preferred over more complex ones

Scientific Method

Step 6: Repeat



Psycholinguistics as a Science

How is psycholinguistics different from the physical sciences?

"Hard" sciences vs "soft" sciences

→ The hard sciences are relatively easy. It's the soft sciences that are hard!

- Human behavior is typically much more variable than most physical systems
- Often the variable of interest requires indirect measurement
  (and thus conclusions may require making inferences)

THE BASICS OF EXPERIMENTATION

What is an experiment?

A systematic investigation of the cause-and-effect relationship between two (or more) variables

- Given a hypothesis, we can manipulate some variables in order to observe their influence on other variables

An experiment involves

1. Manipulating some variable(s)
   - **Independent variables** (IV)
2. Observing the results of the manipulation on other variable(s)
   - **Dependent variables** (DV)
3. While controlling for all other factors
   - **Extraneous variables**

Independent Variables (IVs)

The manipulated variable(s) (aka, explanatory variable or factor)

- Variable that is being studied to see if it will influence behavior
- Called a ‘‘manipulated’’ variable because the experimenter has complete control over it

Each IV must have a minimum of two levels (aka, contrasts)

- The point of an experiment is ***comparison***
- At the very least, an experiment involves a comparison between two things

Conditions

- The combination of all levels of all of IVs

Exercise

Research Question

Are focused words faster to identify than non-focused words?

- IV = Focus
- Levels = focus, non-focus

Operationalizing Variables

- Must clarify what we mean by "focus": Syntactic? Prosodic? Semantic?
- How will syntactic focus be implementated: Clefting? Fronting? Other devices?

Dependent Variables (DVs)

The variable(s) that are measured by the experimenter (aka, response variables)

- The outcome of the experiment
- Called "dependent" because the outcome will depend on the IVs (assuming that the hypothesis is correct)
- Choice of DV is based on the task and the experimental method (e.g., Response time, accuracy, eye movements, brain response, etc.)

Research Question

Are focused words faster to identify than non-focused words?

- DV: We need a measure of word identification speed

  e.g., Response time, reading time, etc.

Important to choose an appropriate task and DV in order to be able to answer the research question

Example Experiment

Instructions

- You will see a list of scrambled words
- Try to unscramble as many words as you can
- You will have 3 minutes to complete the task

facebook
wikipedia
twitter
netflix
reddit
microsoft
linkedin
google

youtube
amazon
instagram
apple
tumblr
yahoo
pinterest
craigslist

Social networking sites

facebook
wikipedia
twitter
netflix
reddit
microsoft
linkedin
google

Non-Social networking sites

youtube
amazon
instagram
apple
tumblr
yahoo
pinterest
craigslist

Social networking sites

facebook
wikipedia
twitter
netflix
reddit
microsoft
linkedin
google

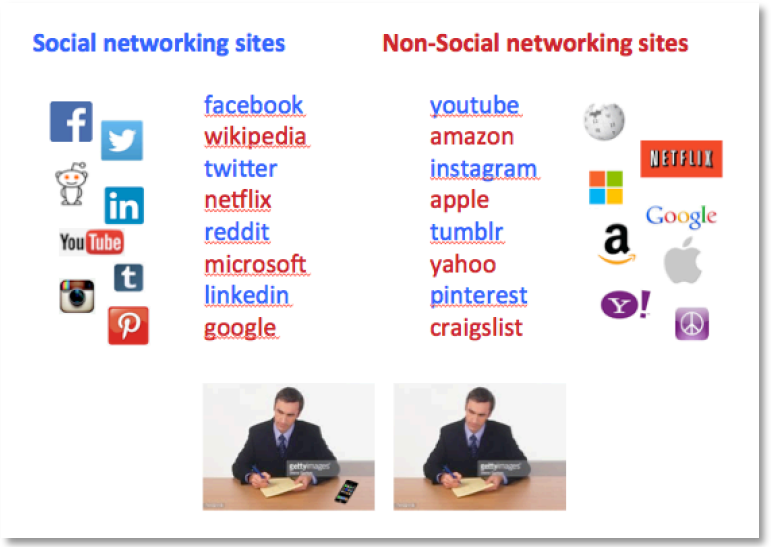
Non-Social networking sites

youtube
amazon
instagram
apple
tumblr
yahoo
pinterest
craigslist

Exercise

Hypothesis: People who feel more socially connected are more familiar with social networking websites than people who do not feel socially connected

- How many factors (IVs) were manipulated? How many levels?
- How many conditions were there?
- What was the DV? Why do you think it was chosen?



Ceiling and floor effects

Ceiling effect

- Participant scores are so high that no differences between conditions can be determined
- Occurs when the DV is so easy that everyone gets a high score

Floor effect

- All scores are extremely low because the task is too difficult, producing a failure to find any differences between conditions

To avoid floor and ceiling effects, pick levels of IV that result in mid-level performance on the DV

Extraneous Variables

Variables that are not of theoretical interest but which might influence behavior in some *systematic* way (can become a confound)

Confound: Co-varies with the IV and could provide alternative explanation of results

Extraneous variables must be controlled (held constant or balanced across conditions) or they may lead to incorrect conclusions

Potential Confounds

Type	Examples	Can be controlled via
Linguistic	Word frequency, word length, word predictability, verb biases, repetition, ambiguity, speech rate, etc.	Corpus analysis, norming pre-tests, balancing across conditions
Situational	Environment (time of day, noise, temperature, lighting, etc.); variability in the way instructions are given	Standardized procedures, standardized instructions
Participant	Language skill, reading experience, intelligence, mood, anxiety, gender, age, socioeconomic class, cultural group, etc.	Random assignment to conditions
Experimenter	Age, gender, accent, mood, etc.; unintentionally provide clues	Standardized procedures, naturalistic environment
Task Demands	Order effects, clues that give away experiment's purpose	Counter-balancing, naturalistic task

→ **If you can't control extraneous variables, be sure that they are not confounded with your IV**

Exercises

In the following studies, try to:

1. Identify the IV and number of levels
2. Identify the task and the DV
3. Consider whether other factors could influence the results

Traxler, Bybee & Pickering (1997)

Abstract

An eye-tracking experiment investigated whether incremental interpretation applies to interclausal relationships. According to Millis and Just's (1994) delayed-integration hypothesis, interclausal relationships are not computed until the end of the second clause […].

We investigated the processing of causal and diagnostic sentences […] that contained the connective *because*. Previous research […] has demonstrated that readers have greater difficulty processing diagnostic sentences than causal sentences. […]

Diagnostic

Heidi could imagine and create things because she won first prize at the art show.

Causal

Heidi felt very proud and happy because she won first prize at the art show.

Traxler, Bybee & Pickering (1997)

IV: Type of interclausal relationship

- Levels: Causal, Diagnostic
  (to obtain two readings, the first clause was manipulated)

Task and experimental method: Read, eye-tracking

DV: Eye-movements (reading time) – in which region did the differences occur

Predictions

- Delayed-integration hypothesis:
  Differences should emerge only at end of second clause
- Immediate-integration hypothesis:
  Differences should emerge before end of second clause

Traxler, Bybee & Pickering (1997)

How did they control for extraneous variables?

- The critical regions were held constant

Diagnostic

Heidi could imagine and create things because she won first prize at the art show.

Causal

Heidi felt very proud and happy because she won first prize at the art show.

Sturt, Pickering & Crocker (1999)

Abstract

Many theories of parsing predict that the difficulty of syntactic reanalysis depends on the type of structural change involved. […]

We report two self-paced reading experiments which demonstrate clear differences in the magnitude of garden path effects associated with different types of structural change.

NP/S

The woman saw the famous doctor had been drinking quite a lot.

NP/Z

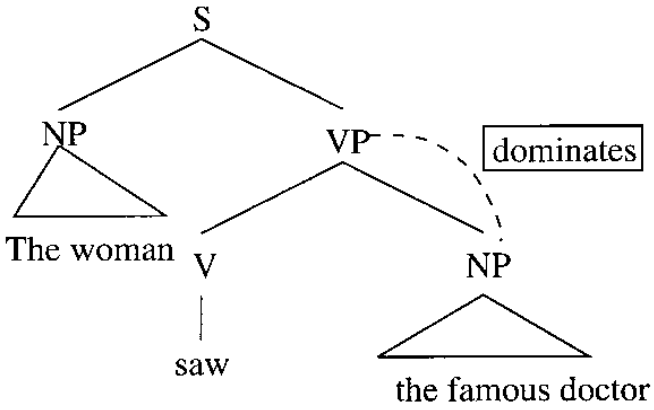
Before the woman visited the famous doctor had been drinking quite a lot.

Sturt, Pickering & Crocker (1999)

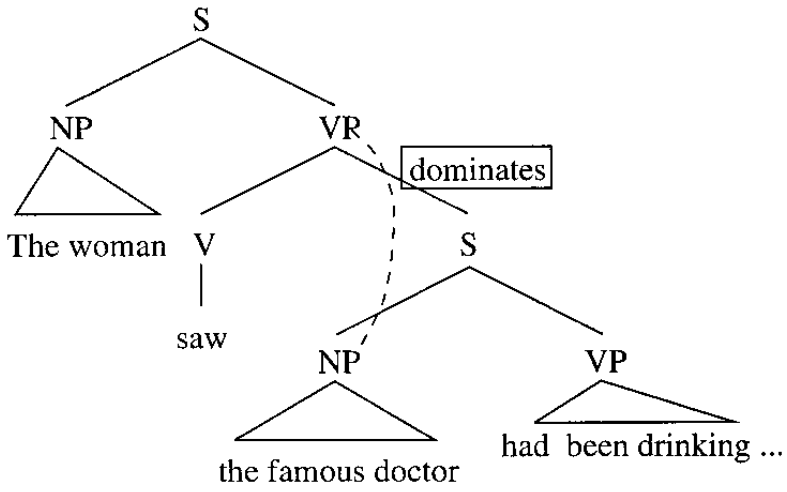
Types of Structural Change

NP/S The woman saw the famous doctor had been drinking quite a lot.

Before Reanalysis



After Reanalysis

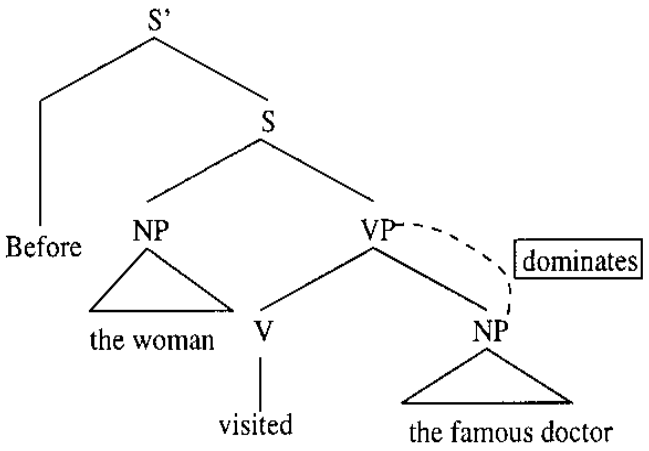


Sturt, Pickering & Crocker (1999)

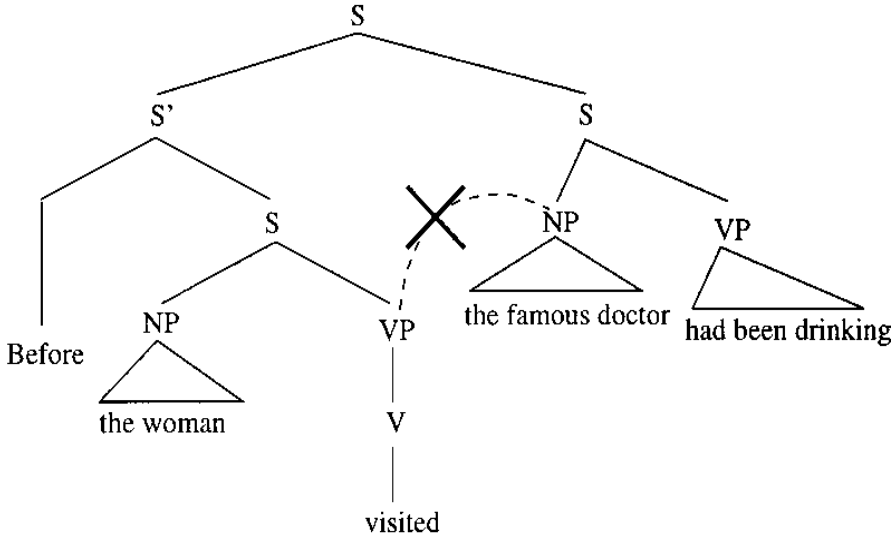
Types of Structural Change

NP/Z Before the woman visited the famous doctor had been drinking quite a lot.

Before Reanalysis



After Reanalysis



Sturt, Pickering & Crocker (1999)

IV: Type of structural change (verb subcategorization)

- Levels: NP/S, NP/Z
- NB: To obtain two readings, the first clause was manipulated

Task: Self-paced reading

DV: Reading time

Sturt, Pickering & Crocker (1999)

Pre-Tests for Possible Confounds

Verb Bias

- Assessed via corpus study
- Balanced across conditions
- Each item paired verbs which had similar preference for the NP-reading over alternative

Sentence Plausibility

- Assessed via judgment task
- NP-reading was equally plausible in each condition

NP/S verb	NP bias	NP/Z verb	NP bias
understood	.92	negotiated	.94
accepted	.93	polished	.93
recalled	.87	scratched	.91
heard	.89	packed	.89
confirmed	.81	typed	.86
maintained	.98	built	.97
forgot	.89	painted	.94
mentioned	.94	debated	.90
found	.94	lost	.90
announced	.91	investigated	.93
discovered	.71	watched	.68
noticed	.65	knitted	.67
saw	.97	visited	.98
acknowledged	.97	questioned	.99
remembered	.97	attacked	.97
remembered	.97	invaded	.95
read	.99	edited	.98
revealed	.79	washed	.77
revealed	.79	followed	.78

Sturt, Pickering & Crocker (1999)

Conditions

NP/S – Ambiguous

The Australian woman saw the famous doctor had been drinking quite a lot.

NP/S – Unambiguous

The Australian woman saw that the famous doctor had been drinking quite a lot.

NP/Z – Ambiguous

Before the woman visited the famous doctor had been drinking quite a lot.

NP/Z – Unambiguous

Before the woman visited, the famous doctor had been drinking quite a lot.

Summary

IVs:

- **Ambiguity** (ambiguous, unambiguous)
- **Type of structure** (NP/S, NP/Z)

→ **2 X 2 design** – two factors, each with two levels

Task: Self-paced reading

DV: Reading time

Controlled variables

- Verb biases
- Plausibility of misanalysis

Types of Design

Single factor design

- One IV with two or more levels
  (e.g., Word length: 1, 2, or 3 syllables)

Factor A	
Cond 1	Cond 2

Factor A		
Cond 1	Cond 2	Cond 3

Factorial design

- Multiple IVs, each with two or more levels
  - 2x2 design
  - 3x2 design
  - 2x2x2 design
  - 3x2x2x2 design (difficult to interpret!)

	Factor A	
Factor B	Cond 1	Cond 2
	Cond 3	Cond 4

	Factor A		
Factor B	Cond 1	Cond 2	Cond 3
	Cond 4	Cond 5	Cond 6

Between-Subjects Design

Each participant is tested in only a subset of conditions

Group A	Group B
Cond 1	Cond 2

Advantage

- Participants are less likely to guess the purpose of the experiment

Disadvantages

- Larger number of participants needed
- Differences between conditions could reflect ***individual differences*** between groups

Individual Differences

People vary along multiple dimensions

- "Naturally defined" variables (not manipulated by experimenter)
  - Language skill, reading experience, intelligence, mood, anxiety, gender, age, socioeconomic class, cultural group, etc.
- If groups differ, we cannot determine whether results are due to our manipulation or to differences between groups

Solution: Create equivalent groups

- Random assignment of participants
- Confirm that individual differences are balanced

Within-Subject Design

Each participant is tested in each condition (also called ***repeated measure design***)

Advantages

- More control on individual differences
- Less subjects are needed

Disadvantages

- Participants are more likely to guess purpose of experiment
- ***Carry-over effects, order effects***

Carry-Over Effects

Occur when testing under one condition affects how participants behave in another condition

The Australian woman saw the famous doctor had been drinking quite a lot.

The Australian woman saw that the famous doctor had been drinking quite a lot.

- If you present participants with very similar sentences such as (a) and (b) (in this order), they may be faster to read (b) because they remember (a)

Solution

- ***Counterbalancing*** (e.g., ***Latin Square***)

Order Effects

Presenting conditions in the same order to all participants may affect performance

- **Practice effects**
  Participants get better at a task the more they become familiar with it

- **Fatigue effects**
  The longer participants do a task, the more tired they will get and the less attention they will give

- **Catching-on**
  The longer participants do a task, the more likely they will figure out the manipulation

Counterbalanced Lists (Latin Square)

Ensures that each item is tested equally often in each condition and each participant "sees" an equal number of items in each condition

Simple 1x3 design: N items (sets of sentences) in 3 conditions

	List 1	List 2	List 3
Item 1	Cond 1	Cond 2	Cond 3
Item 2	Cond 2	Cond 3	Cond 1
Item 3	Cond 3	Cond 1	Cond 2
Item 4	Cond 1	Cond 2	Cond 3
…			
Item n	Cond 3	Cond 1	Cond 2

→ Participants are randomly assigned to lists

Hiding the Manipulation

Include Filler items (e.g., sentences with different structure)

	List 1	List 2	List 3
Item 1	Cond 1	Cond 2	Cond 3
Filler 1	Filler 1	Filler 1	Filler 1
Filler 2	Filler 2	Filler 2	Filler 2
Item 2	Cond 2	Cond 3	Cond 1
Filler 3	Filler 3	Filler 3	Filler 3
Filler 4	Filler 4	Filler 4	Filler 4
Item 3	Cond 3	Cond 1	Cond 2
…			
Item n	Cond 3	Cond 1	Cond 2

- Number of fillers depends on experiment
- Rule of thumb: Use at least twice the number of items

Dealing with Order Effects

Randomization

- Randomize the order of items so that conditions are not predictable

Pseudo-randomization

- Adjust the „true“ random order to avoid potential issues:
  - Begin list with 2-3 filler items
  - Remove „runs“ (more than X number of same condition in a row)

Reversed Lists

- Avoids any potential order effects that may remain

Breaks

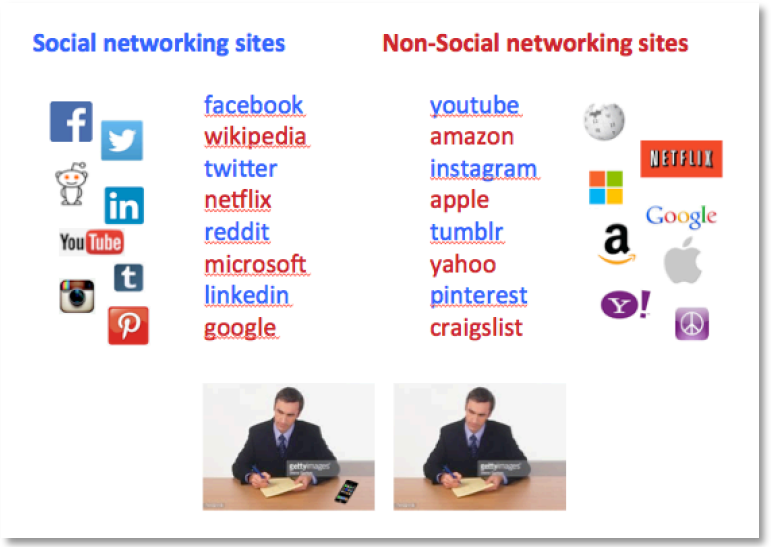
- Minimizes fatigue

Exercise

What was the design of this study?

- Depends on number of factors (IVs) and within- vs between-subjects
  - 2x2 design
  - With one within-subjects factor
  - And one between-subjects factor

	Factor A (within)	
Factor B (btwn)	Cond 1	Cond 2
	Cond 3	Cond 4



Summary

General Design Principles

1. Formulate your research question clearly
2. Choose appropriate independent and dependent variables
3. Control or balance all other factors
4. Use a within-subject design whenever possible
5. Counterbalance and (pseudo)randomize your materials

Agenda

Today

- The basics of experimentation

Next Monday

- Introduction to the course experiment
- **Have read**: Van Berkum et al. (2008)

Next Wednesday

- Methods and measures

Reading for Monday

The Neural Integration of Speaker and Message

Jos J. A. Van Berkum[1,2,3], Danielle van den Brink[3,4], Cathelijne M. J. Y. Tesink[3,4], Miriam Kos[3], and Peter Hagoort[1,3,5]

Abstract

■ When do listeners take into account who the speaker is? We asked people to listen to utterances whose content sometimes did not match inferences based on the identity of the speaker (e.g., "If only I looked like Britney Spears" in a male voice, or "I have a large tattoo on my back" spoken with an upper-class accent). Event-related brain responses revealed that the speaker's identity is taken into account as early as 200–300 msec after the beginning of a spoken word, and is processed by the same early interpretation mechanism that constructs sentence meaning based on just the words. This finding is difficult to reconcile with standard "Gricean" models of sentence interpretation in which comprehenders initially compute a local, context-independent meaning for the sentence ("semantics") before working out what it really means given the wider communicative context and the particular speaker ("pragmatics"). Because the observed brain response hinges on voice-based and usually stereotype-dependent inferences about the speaker, it also shows that listeners rapidly classify speakers on the basis of their voices and bring the associated social stereotypes to bear on what is being said. According to our event-related potential results, language comprehension takes very rapid account of the social context, and the construction of meaning based on language alone cannot be separated from the social aspects of language use. The linguistic brain relates the message to the speaker immediately. ■

INTRODUCTION

Reading for Monday

Focus on the following questions and come prepared to discuss them

1. What was the theoretical basis/motivation for the study?
2. What was the specific research question?
3. What was the Independent Variable(s) and levels?
4. What was the Dependent Variable(s)?
5. What was the task (what did the participant have to do)?
6. Was it a within-subjects or between-subjects design?
7. What steps were taken to eliminate potential confounds and order effects?
8. What potential confounds were not controlled for?
9. What did they find?
10. What did they conclude based on these findings?