Methods in Psycholinguistics

Ben Weissman

COGS 4780

Overview

- Goals for today:
  - Discover various methods used in psycholinguistics research
  - Introduce basic principles of statistics
  - Discuss principles of experimental design

Pre-psycholinguistics

- Grammaticality judgments:

  What did the pig give to the donkey?

  What did the pig sleep to donkey?

Pre-psycholinguistics

- Grammaticality judgments:

  What did the pig give to the donkey?

  *What did the pig sleep to donkey?

Beginnings of psycholinguistics

- Cognitive psychology emerged in the 1960s as part of the broader approach of cognitive science
  - Mind is an information processor
  - *How* does the human mind get from input to output?

- Cognitive psychology aims to model a process
  - What are the parts of the process?
  - What is the order of the process?
  - How do the parts of a process interact?

Beginnings of psycholinguistics

- Psycholinguistics runs experiments to learn more about the way language works in the mind
- Typically done with language comprehension rather than production
- Begins with a hypothesis related to an existing model of language; experiment designed to test that hypothesis

Psycholinguistic methods

- Underlying assumption: direct relationship between processing speed and processing
  - The longer it takes to process something, the more work we're doing internally to process it
    - *Processing load*

- In most experimental methodologies, it isn't necessarily obvious what that processing is
  - Experimental design helps us figure it out

Psycholinguistic methods

- Simplest version:

- Design an experiment where half of the items are in Condition A and half of the items are in Condition B
- Run a statistical comparison between the responses to Condition A items and the responses to Condition B items

Psycholinguistic methods

- Typically we don't just have one independent variable (IV/predictor) of interest
  - Statistical methods would let us explore the effect of A vs B, the effect of 1 vs 2, and any interaction

2x2 design

	A	B
1	A1	B1
2	A2	B2

Quick statistics notes

- **Interactions** are relationships between predictors

- IV 1 – subject number (singular vs. plural)
- IV 2 – grammaticality (grammatical vs. ungrammatical)

- In a statistical model, we might want to study the effect of subject number and the effect of grammaticality, but we would also want to study the interaction between subject number and grammaticality
  - Maybe grammaticality affects singular items and plural items in a different way

Psycholinguistic methods

- 2x2 design:

- IV 1 – subject number (singular vs. plural)
- IV 2 – grammaticality (grammatical vs. ungrammatical)

The key to the cabinet was rusty
The key to the cabinets was rusty
The key to the cabinet were rusty
The key to the cabinets were rusty

Psycholinguistic methods

- Why might we not want a participant to encounter all four of the following sentences?

The key to the cabinet was rusty

The key to the cabinets was rusty

The key to the cabinet were rusty

The key to the cabinets were rusty

Psycholinguistic methods

- Why might we not want a participant to encounter all four of the following sentences?

A The key to the cabinet was rusty

B The key to the cabinets was rusty

C The key to the cabinet were rusty

D The key to the cabinets were rusty

- In a properly set up experiment, each item will have an A B C D version, each item will be seen the same number of times in each condition, and each participant will see equal numbers of A B C D items

Psycholinguistic methods

A The key to the cabinet was rusty
B The key to the cabinets was rusty
C The key to the cabinet were rusty
D The key to the cabinets were rusty
A The slogan on the poster was designed to get attention
B The slogan on the posters was designed to get attention
C The slogan on the poster were designed to get attention
D The slogan on the posters were designed to get attention
A The name on the billboard was of a prominent local politician
B The name on the billboards was of a prominent local politician
C The name on the billboard were of a prominent local politician
D The name on the billboards were of a prominent local politician

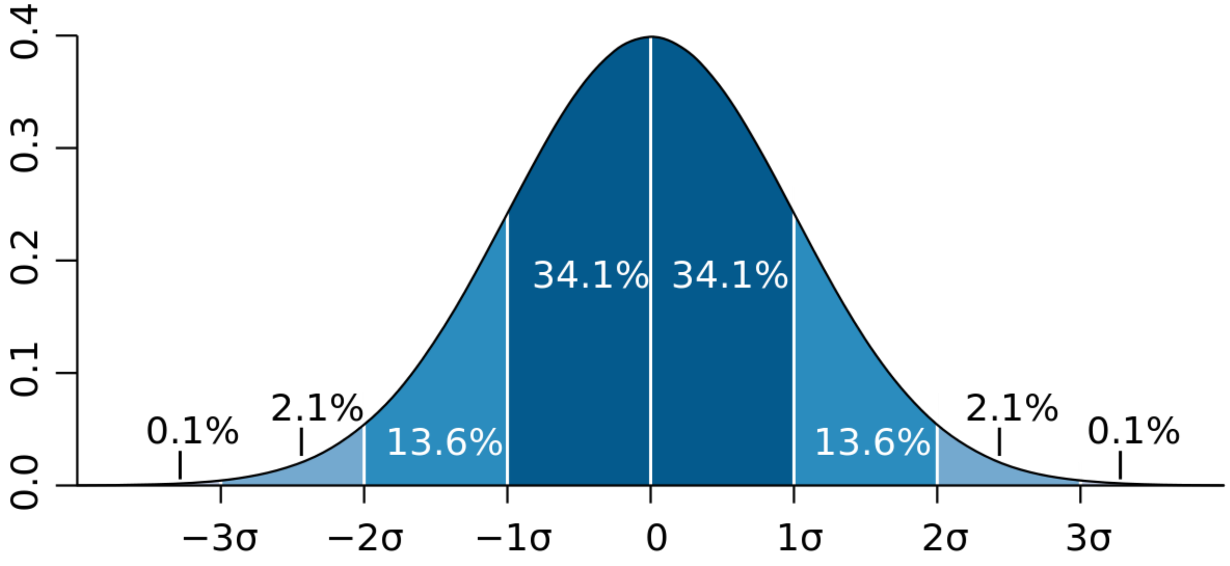
Psycholinguistic methods

- **Fillers** – items that appear in the experiment that are not directly relevant to the comparison in question

- Why might we want to use fillers?

Quick statistics notes

- **Mean –** numerical average of data collected in a sample
  - Expected value
- **Standard deviation** – a measure of variation in data
  - Average difference between any single observation and the expected value
- Normal distribution →



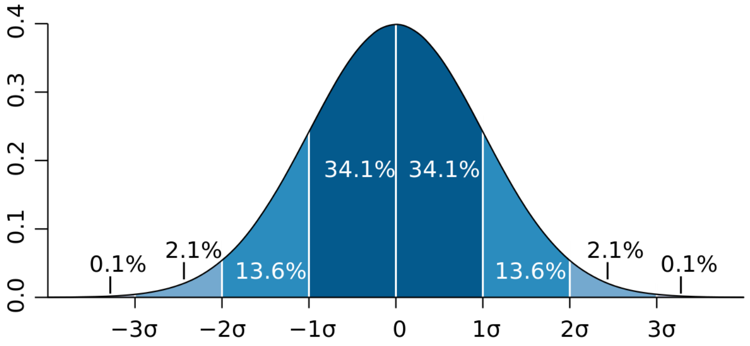
Quick statistics notes

- Many of the results we'll look at throughout the semester come from *regression* models – an attempt to use predictors (independent variables) to predict the value of an outcome (dependent variable)
- This sort of statistics is *inferential* – using mathematical models to make more general claims based on observed data
- We utilize what we have (observations, mean, variance, sample size, etc.) to infer about what really is the case
  - Our data is a sample of the entire population
  - What we have is the sample, what we want is the population
  - Data quality matters – good experimental design, large enough sample size, no confounds

Quick statistics notes

- In inferential statistics, we try to come up with an estimate of *parameters* – the numbers on which a statistical model is built
    - Often coefficients, estimates of the slope, the effect of a predictor on an outcome

Quick statistics notes



- In inferential statistics, we try to come up with an estimate of *parameters* – the numbers on which a statistical model is built
  - Often coefficients, estimates of the slope of the effect of a predictor on an outcome
- **Standard error** – the standard deviation of an estimate
  - SE can be used to represent the amount of uncertainty – it gets smaller as sample size gets bigger
- A **confidence interval** is a range of values of a parameter that is consistent with the data
  - A 95% confidence interval means that if we collected data this exact same way *n* times, our parameter estimate will be within ±2 standard errors of the *actual* parameter value 95% of the time
  - ***Not*** the probability that the parameter is within the confidence interval

Quick statistics notes

- Assume a *null hypothesis* that there is no effect present
  - Let's say that means the parameter is 0
- Coeffecient estimates are labeled **statistically significant** if they are at least two standard errors away from the null hypothesis (in this case 0)
- ***p*-value** is a metric of statistical significance; a *p*-value under .05 is said to be statistically significant

- If we flip 20 coins, we would expect a proportion of tails to be .5
  - Lets say we get 8 tails and 12 heads – .4 tails
  - Standard error = 0.11
  - Data is still within 2 standard errors of our null hypothesis value
  - This difference is not statistically significant
  - If we get 800 tails and 1200 heads…

Quick statistics notes

- Many problems with *p*-values
  - *p*-values are a function of sample sizes, so in many cases, a high enough sample size will get you statistical significance
  - Significance cutoff is arbitrary
  - Conflation of statistical hypotheses and scientific hypotheses
  - Often misinterpreted and misrepresented
- **however**
- *p*-values are extremely common and still can be valuable

- The potentially (likely) better approach to inferential statistics (Bayesian statistics) is still gaining popularity and thus not often found in existing research

Reaction Time

- One of the simpler methods to use is a general Reaction Time experiment
  - This can be used for many different purposes, including grammaticality judgments

Reaction Time

- One of the simpler methods to use is a general Reaction Time experiment
  - This can be used for many different purposes, including grammaticality judgments

The key to the cabinets were rusty

YES NO

Reaction Time

- One of the simpler methods to use is a general Reaction Time experiment
  - This can be used for many different purposes, including grammaticality judgments

The key to the cabinets were rusty

YES NO

- What are some potential issues with a test set up this way?

Reaction Time

- One of the simpler methods to use is a general Reaction Time experiment
  - This can be used for many different purposes, including grammaticality judgments

- **Counterbalancing** of left/right buttons across participants
- **Controlling** for sentence length/reading time
  - Must be even across conditions
    - Avoid confounds!!!!!!

Lexical Decision Task

- A Lexical Decision Task also measures response times and accuracy
- In a Lexical Decision Task, the prompt is to decide whether the item is a real word or not

Lexical Decision Task

- A Lexical Decision Task also measures response times and accuracy
- In a Lexical Decision Task, the prompt is to decide whether the item is a real word or not

grapefruit

YES NO

Lexical Decision Task

- A Lexical Decision Task also measures response times and accuracy
- In a Lexical Decision Task, the prompt is to decide whether the item is a real word or not
- In this case, accuracy typically isn’t all that interesting
  - (people know that “grapefruit” is a real word and “grakefruit” isn’t)
- But the reaction times can reveal effects of **priming** or other lexical processing
  - If the participant had recently seen related words, like “lemon” and “orange,” the response time to accept “grapefruit” as a real word will be significantly faster than if they had only seen unrelated words

Self Paced Reading

- Self Paced Reading also takes time as the measurement, but measures reading times on each word in a sequence
- Example
- Slower reading times taken as evidence of increased processing load
  - Unexpected words take longer to process than expected words
  - Ungrammatical words take longer to process than grammatical words

Self Paced Reading

- Any issues or considerations one would need to keep in mind in designing a SPR experiment?

- Longer words will take longer to read than shorter words
  - (this needs to be counterbalanced or accounted for)
- Effects aren't immediate – will extend into the **spillover** window
  - Effects that emerge in self-paced reading may show up on the target word *w* itself as well as *w*+1, *w*+2
- The slowdown doesn't necessarily tell us what's going on… just that something is going on
  - If we find a significant difference between two conditions, it still needs to be explained/interpreted

Self Paced Reading

The / key / to / the / cabinet / was / rusty

The / key / to / the / cabinets / was / rusty

What should we expect to observe in SPR results here?

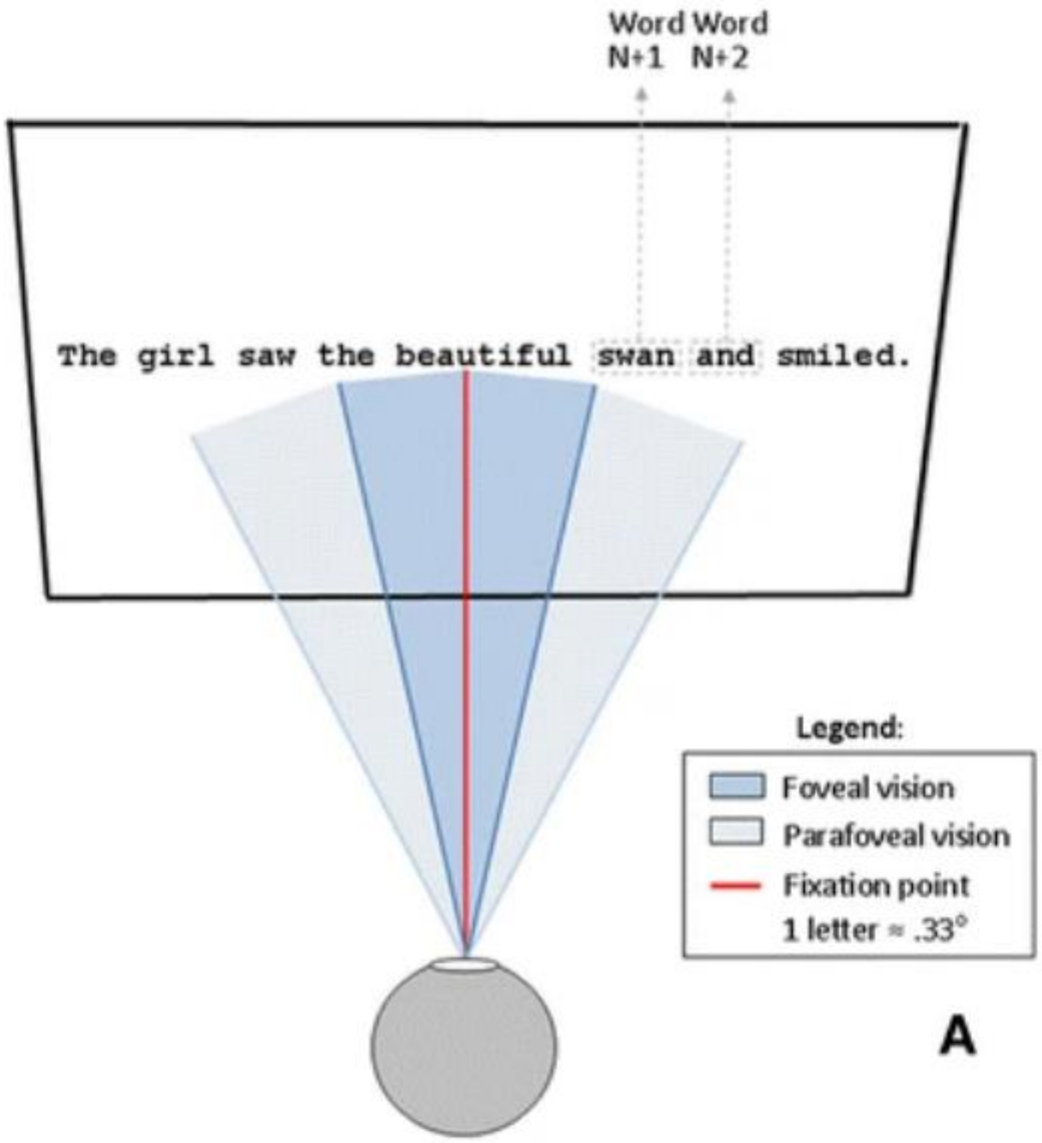
Eye Tracking

- Similarly, we can track eye movements as people read a sentence

- What sorts of eye movement measurements might we want to track?
  - Fixation duration
  - Number of fixations
  - Saccades
    - Regressions
    - Skips

Eye Tracking

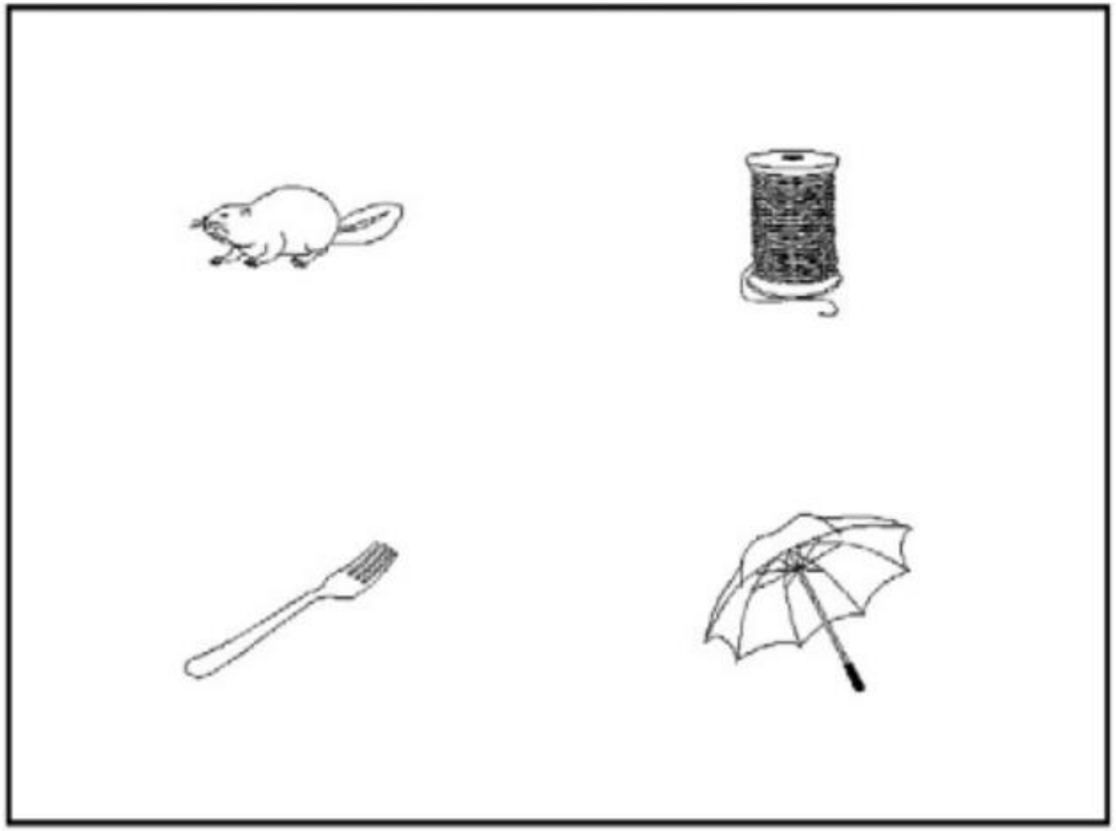
- When we read sentences, we get information from upcoming words – **parafoveal preview**
  - Fovea – part of our retina with the highest density of cells, where we have the highest visual acuity
- **Gaze-contingent** eye tracking will change upcoming words in between parafoveal preview and fixation

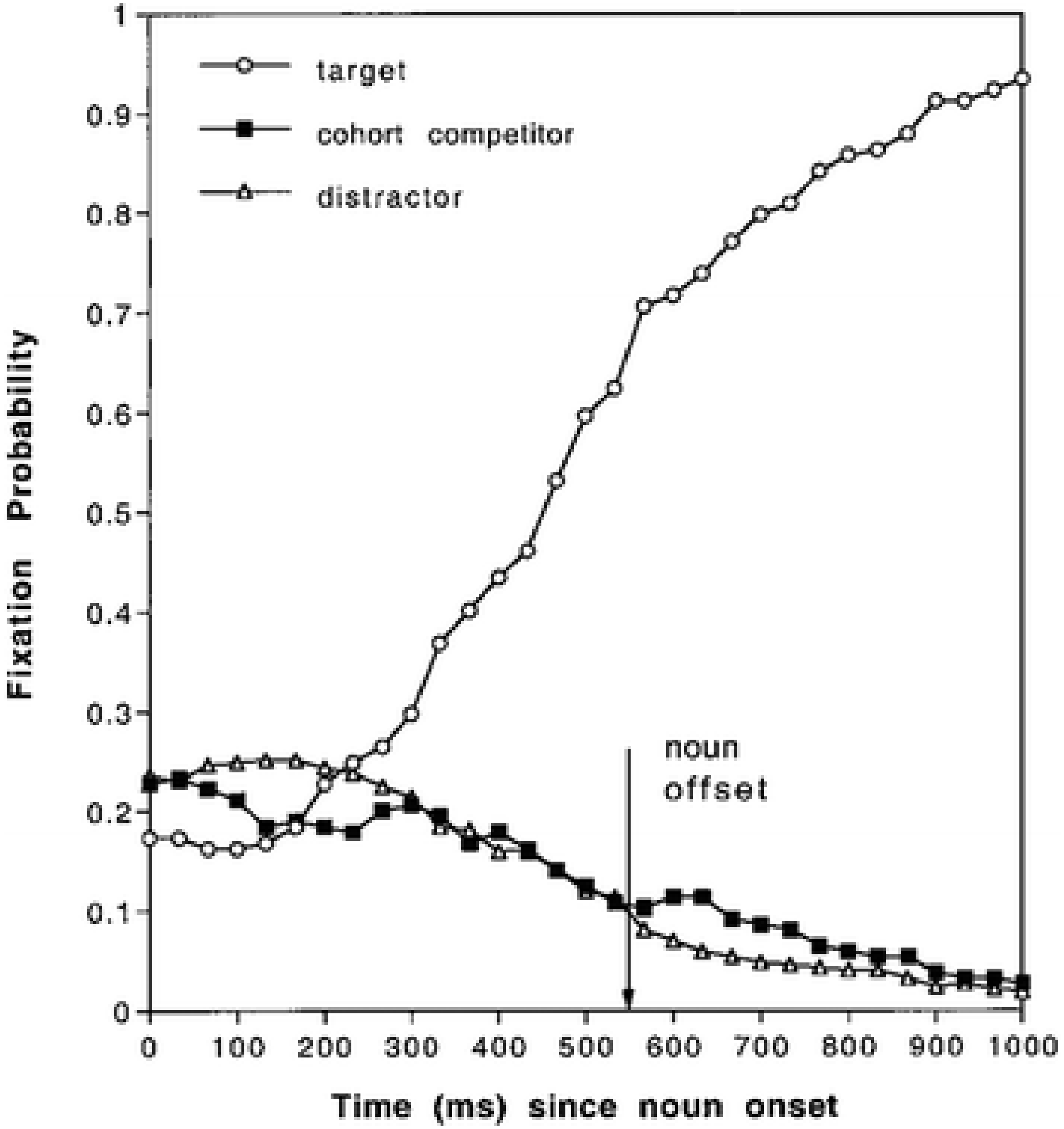


Visual World Eye Tracking

- Participants hear a sentence presented auditorily while eye movements are tracked

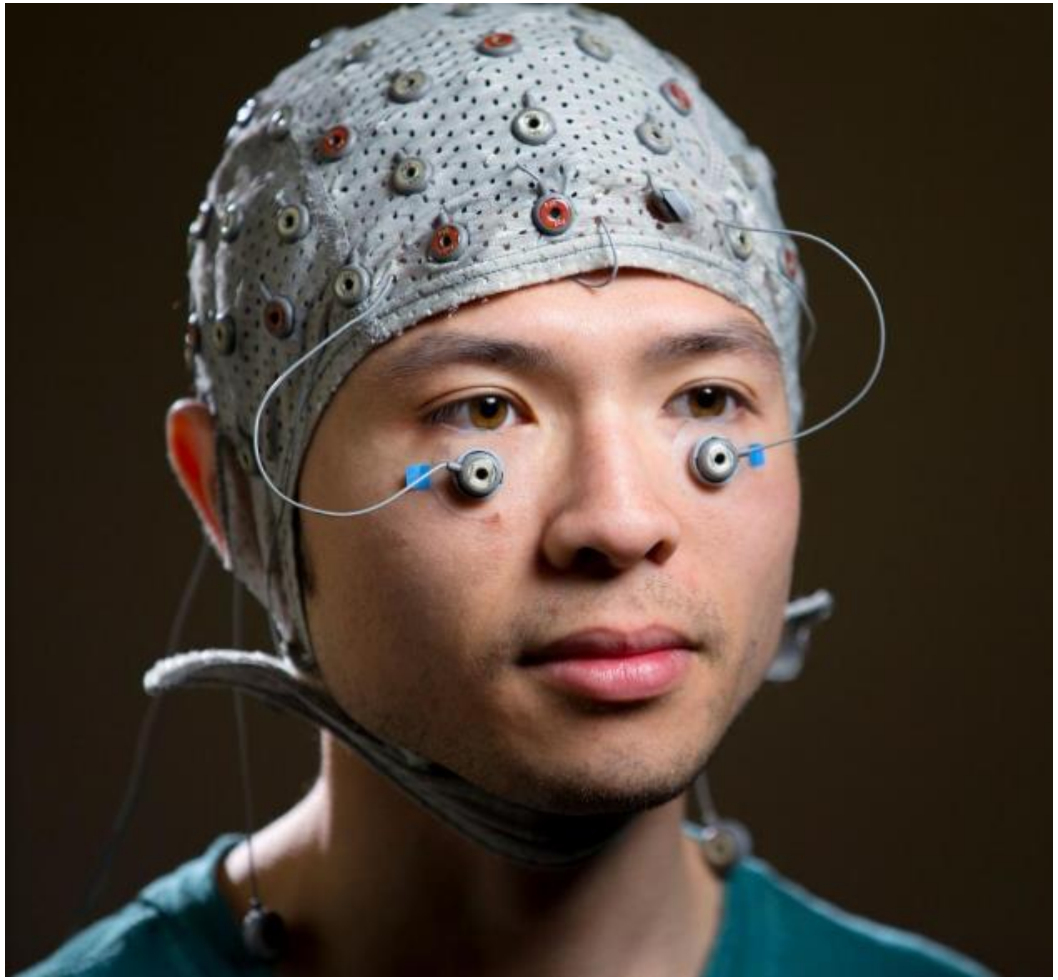
- Measure is proportion of looks at different items in the screen over time

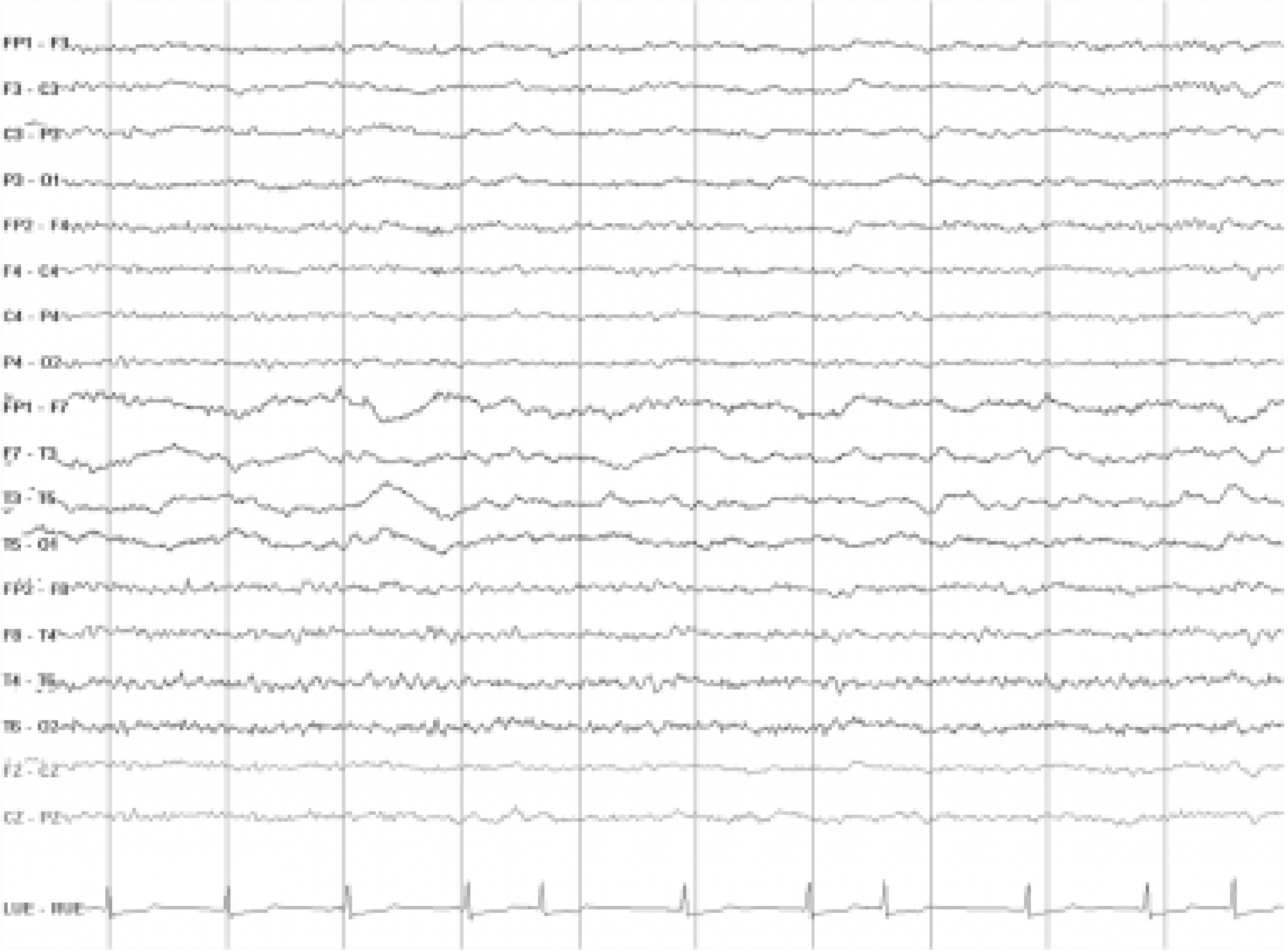



target
cohort competitor
distractor
noun
offset
Fixation Probability
Time (ms) since noun onset
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
100
200
300
400
500
600
700
800
900
1000

ERPs

- Collected using **EEG** - electroencephalography
- Electrodes monitor electrical activity produced by the brain in real time
- Can be measured over a period of time (like during sleep)
- Can measure real-time responses to stimuli – Event-Related Potentials (ERPs)
  - Excellent temporal resolution
  - Does not help with localization



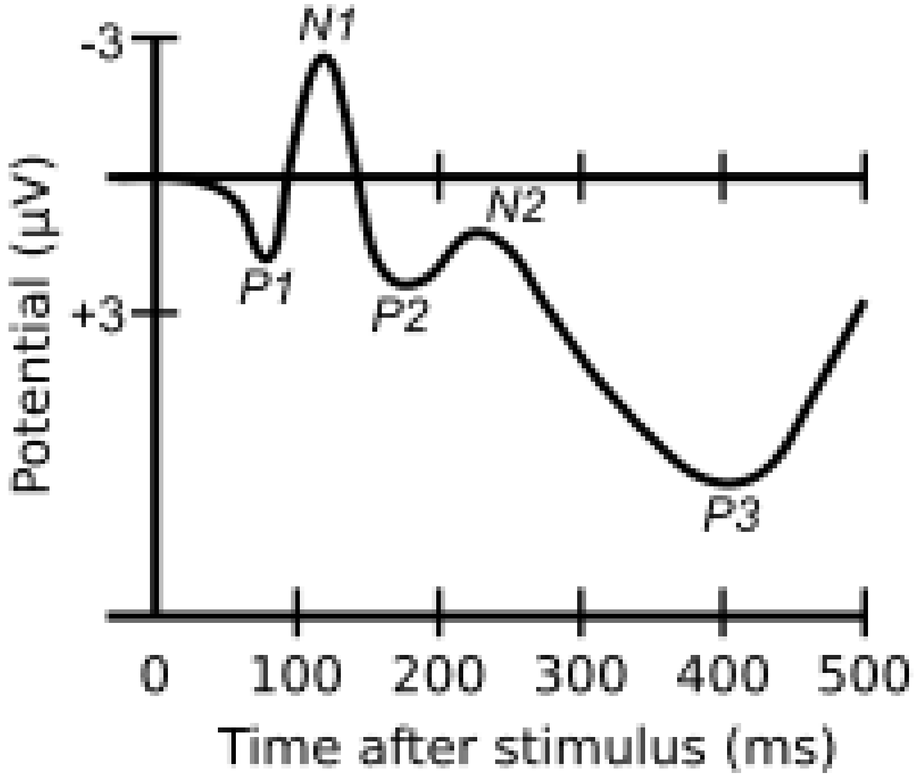

FP1 - F3
F3 - C3
C3 - P3
P3 - O1
FP2 - F4
F4 - C4
C4 - P4
P4 - O2
FP1 - F7
F7 - T3
T3 - T5
T5 - O1
FP2 - F8
F8 - T4
T4 - T6
T6 - O2
FZ - CZ
CZ - PZ
LUE - RUE

ERPs

- ERPs measure responses (in μV) to stimuli in real time
- Data is averaged over multiple trials and multiple participants to calculate the average response

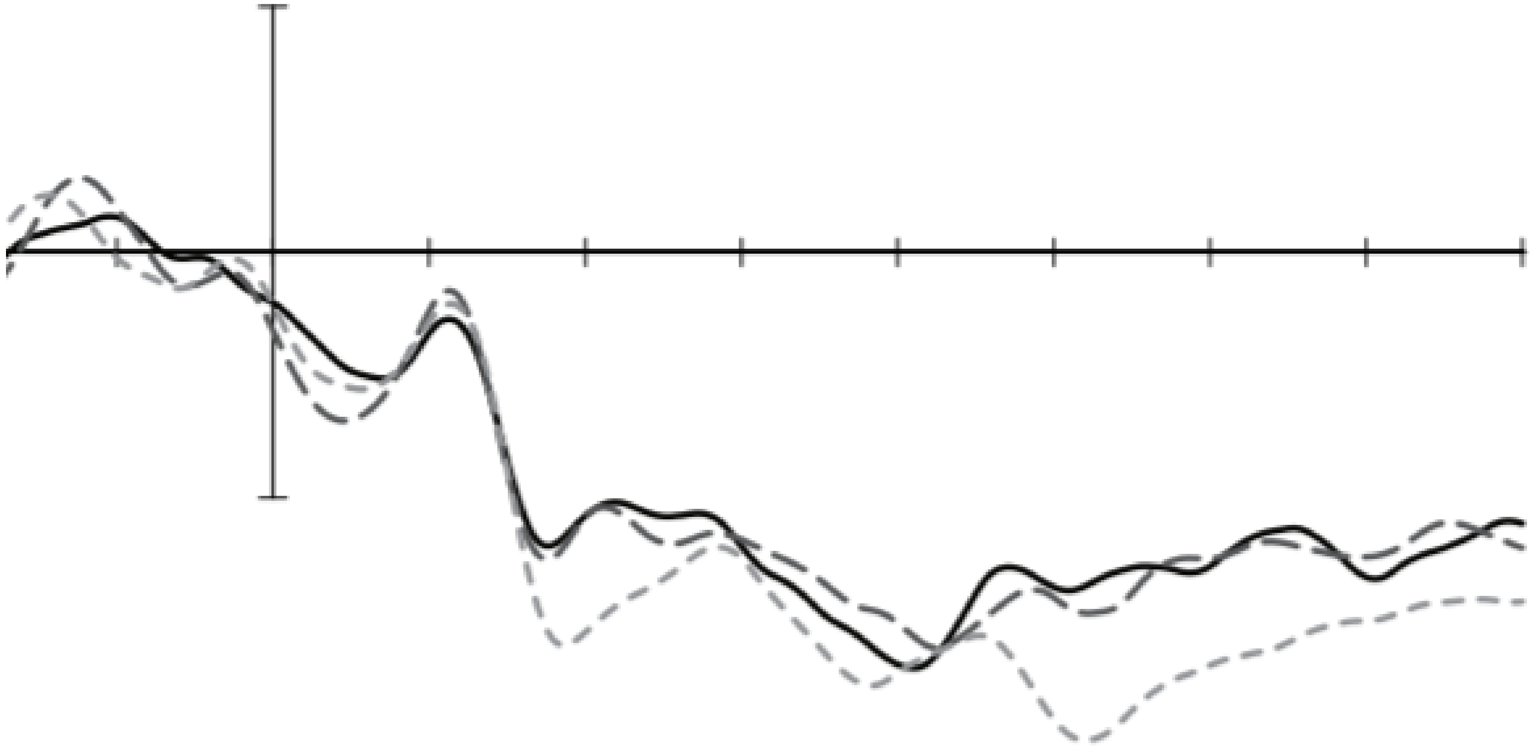
The cat chased after the **mouse**
The cat chased after the **potato**

The child ate an **apple**
The child ate an **apples**



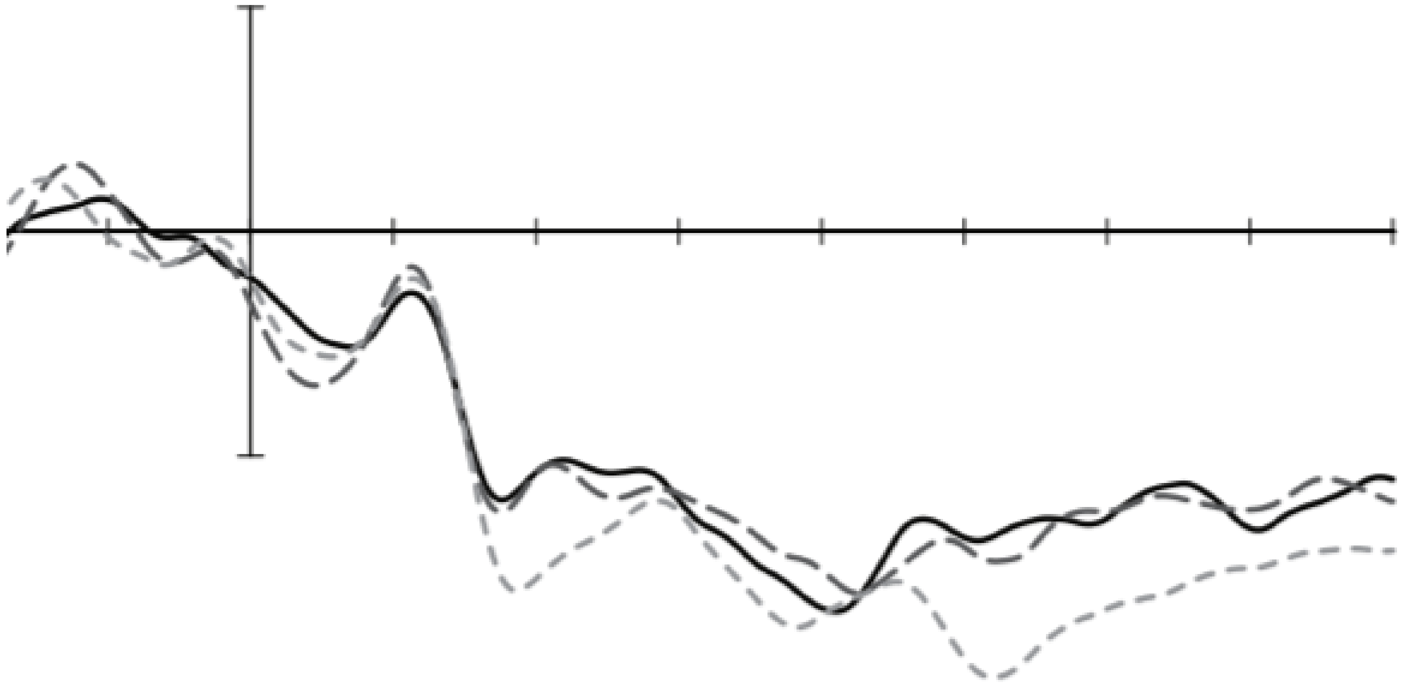
ERPs

- ERPs are plotted as voltage over time
- Negative voltage is plotted up, positive voltage is plotted down (don't ask)
- Different experimental conditions plotted as different colors or different types of lines



ERPs

- For any stimulus, there are typical negative and positive peaks – these are called **components**
- If two conditions differ significantly in a component, these are called **effects**



ERPs

- Two ERP components are especially relevant for language research
- To analyze these components, you would get the average voltage during a certain time window across all participants and all trials and run a statistical analysis comparing different conditions within the time window

ERPs

- Two ERP components are especially relevant for language research
- **N400** – a negative-going component that peaks around 400ms
  - (doesn’t have to be overall negative itself)
  - Typically appears over the middle/posterior sections of the scalp, slightly to the right (though this doesn’t mean that’s where the signal is coming from)
  - Associated with *meaning*, appears when processing a meaningful stimulus
    - Like a word

What are some other examples of meaningful stimuli?

ERPs

- Two ERP components are especially relevant for language research
- **N400** – a negative-going component that peaks around 400ms
  - (doesn’t have to be overall negative itself)
  - Typically appears over the middle/posterior sections of the scalp, slightly to the right (though this doesn’t mean that’s where the signal is coming from)
  - Associated with *meaning*, appears when processing a meaningful stimulus
    - Like a word
  - Component is enhanced when encountering something unexpected compared to expected
    - Expected – Unexpected difference = **N400 effect**

ERPs

- Two ERP components are specifically relevant for language research
- **P600** – a positive-going component that peaks around 600ms
  - Typically appears over the middle/posterior sections of the scalp right down the middle
  - **P600 effect** often found to grammatical errors and violations
  - Varying interpretations, sometimes suggested to be reflecting a revision process

- For both of these components, we don’t necessarily know what’s going on… but we know something is going on, and we know other cases in which they tend to appear

fMRI

- **Functional Magnetic Resonance Imaging** (fMRI) scans blood flow* in the brain
  - Localization of activity
  - *technically detecting differences between oxygen-rich blood and oxygen-poor blood (more oxygen suggests more activity)
- Like in EEG, there’s blood flow going on in the brain at all times
  - Experimental design allows for close comparison between conditions
  - *Relative* blood flow is what matters here

fMRI

- Using fMRI allows us to get a sense of where in the brain certain language processes may take place
    - **Voxel** analysis

- Simultaneous EEG/fMRI is rare but possible, allows for temporal and spatial resolution
    - Magnetoencephalography (MEG) also allows for temporal and spatial resolution
        - (much more expensive, therefore also rare)

Language Production

- Studying language production is much more difficult – why?

Language Production

- Studying language production is much more difficult
- A **corpus study** collects (lots of) instances of language use
  - Typically naturally-occurring, like social media posts or recordings of speech
  - Can be analyzed to observe development, speech errors, production patterns, etc.

- **Artificial Language Learning** teaches participants (elements of) an artificially created language and observes learning through production or comprehension

Quick Review

- Response time
- Lexical decision task
- Self paced reading
- Eye tracking
- Visual world
- ERPs
- fMRI
- Corpus study
- Artificial language learning

Practice

- Does metaphor take longer to process than literal language?
- Does being bilingual affect word recognition?
- Where in the brain do we process sarcasm?
- What happens when we encounter syntactic ambiguity?
- Does knowing a tonal language make you better at learning a new tonal language?