# Cognitive Foundations

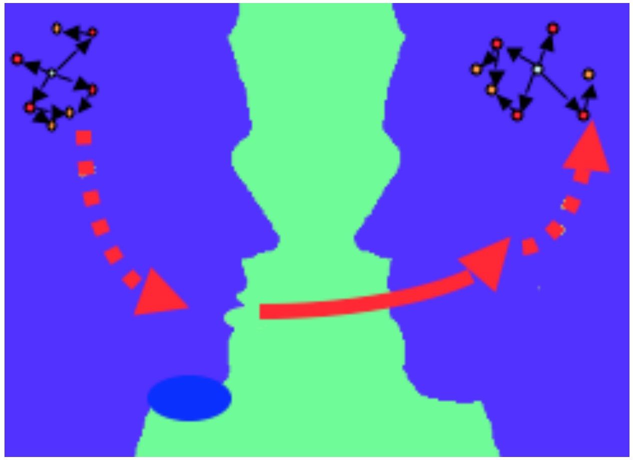
## Lecture 2: Experimental Methods (2)

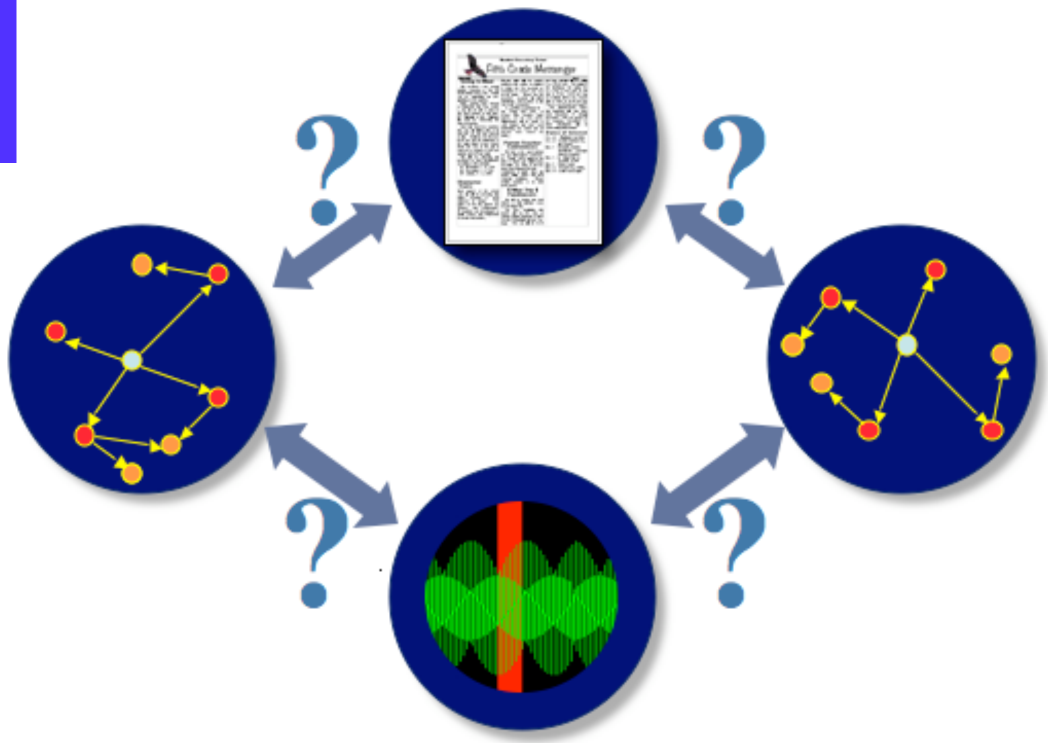
Foundations of Language Science and Technology

Garance Paris

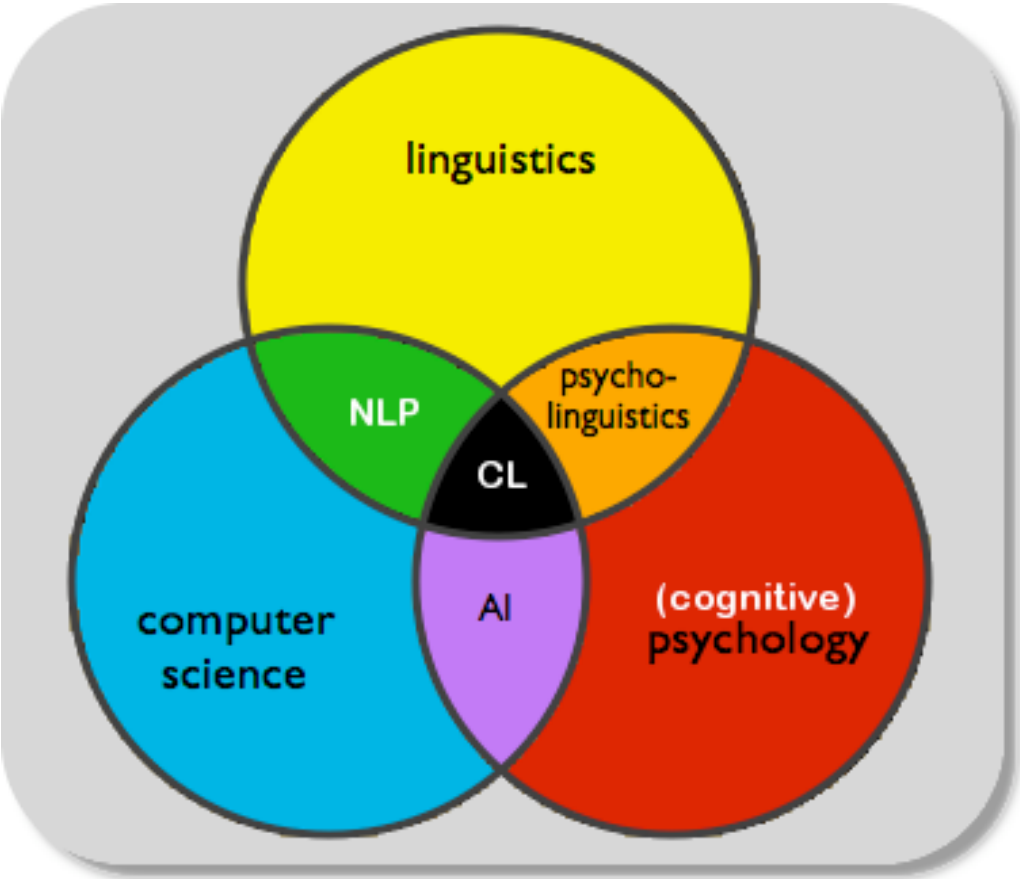
12 November 2008

# Review (1): The Miracle





# Review (2): An Interdisciplinary Field




The three motivations of computational linguistics:

- Theoretical motivations (linguistic & cognitive): Understand, check and improve linguistic and cognitive theories
- Practical motivation: Language technology applications

# Defining Language

- Language is specifically human
- Animal communication does not have the same properties
- Some features of human language:
  - infinite and "double-articulated", hierarchically organized
  - semanticity and arbitrariness
  - social/cultural phenomenon and learnable (bird songs are innate, but isolated children do not develop language)
  - spontaneous usage, creativity
  - ability to refer to things remote in time and place
  - meta-language, reflection, inner speech
  - ability to lie
  - ...

# Nativism vs. Empiricism

- Since 1950s-1960s ("The Cognitive Revolution"): First attempts to explain language processes (Chomsky)
  - Language is very complex, at least "context-sensitive" (type 1)
  - Distinction between competence and performance: Actual language data is very noisy and often ambiguous, but we can still deal with it in "real-time" (incrementally)
  - Therefore language skills must be in part innate ("principles")
  - This also explains universal properties of language
- Empiricism: Linguistic knowledge is acquired from experience with language and with the world
  - Assumptions are simpler
  - Machine learning is being used increasingly in computational linguistics, with at least some degree of success

# Fascinating...

- Language is extremely complex...
  - Speech streams include no boundaries to indicate where one word ends and another begins.
  - We understand stammering non-fluent politicians and non-native speakers. Incomplete and ungrammatical sentences are often no problem to interpret.
  - We deal with ambiguity all the time without breaking down. Computer parsers often maintain thousands of possible interpretations.
  - We have a vocabulary of about 60,000 words. We access somewhere between 2-4 words/second with an error rate of around 2/1000.
- Yet we understand it incrementally, in “real time”. We are so fast, we can even finish each others sentences!

# Humans vs. Computers

- People:
  - are sensitive to context and adapt to circumstances
  - are accurate, fast, robust
  - process language incrementally
  - but have limitations on memory and work-load
- Computers:
  - can do some things better/faster than people: search 1000s of text, classify them, ...
  - can usually only do well very limited NLP tasks
  - can't do things people do trivially: build semantically rich, context-sensitive interpretations

# Natural Language vs. Programming Languages

- Ambiguity, malformed utterances:
  - Pervasive in natural language at all levels of analysis
  - We use context to disambiguate and often don't even notice the ambiguity or error
  - Programming languages must be unambiguous and cannot deal with malformations
- Natural Language is highly redundant
- Distinction between competence and performance does not apply to programming languages:
  - If a sentence is licensed by the grammar rules, it can be parsed, otherwise it cannot (including garden-paths sentences and center-embeddings)

# Where Data Comes in Handy

- Current challenge for NLP: Combination of deep and shallow processing
- How do humans do it?

# Different "Dimensions"

- Various levels of linguistics analysis
- Representation and knowledge, processing, acquisition language disorders
  - William's syndrom: IQ=50% but good language ability
  - Wernicke's aphasia: Speak fluently, but content does not really make sense + neologisms (e.g.:
    [...] but I have had that, it was ryediss, just before the storage you know, seven weeks, I had personal friends [...]"
  - Broca's aphasia: Normal IQ, comprehension ok, production non-fluent, few function words, no intonation
  - Language Specific Impairment: normal IQ, language appropriate, problem with grammatical morphemes, poor memory
- Comprehension vs. Production
- Written language vs. speech

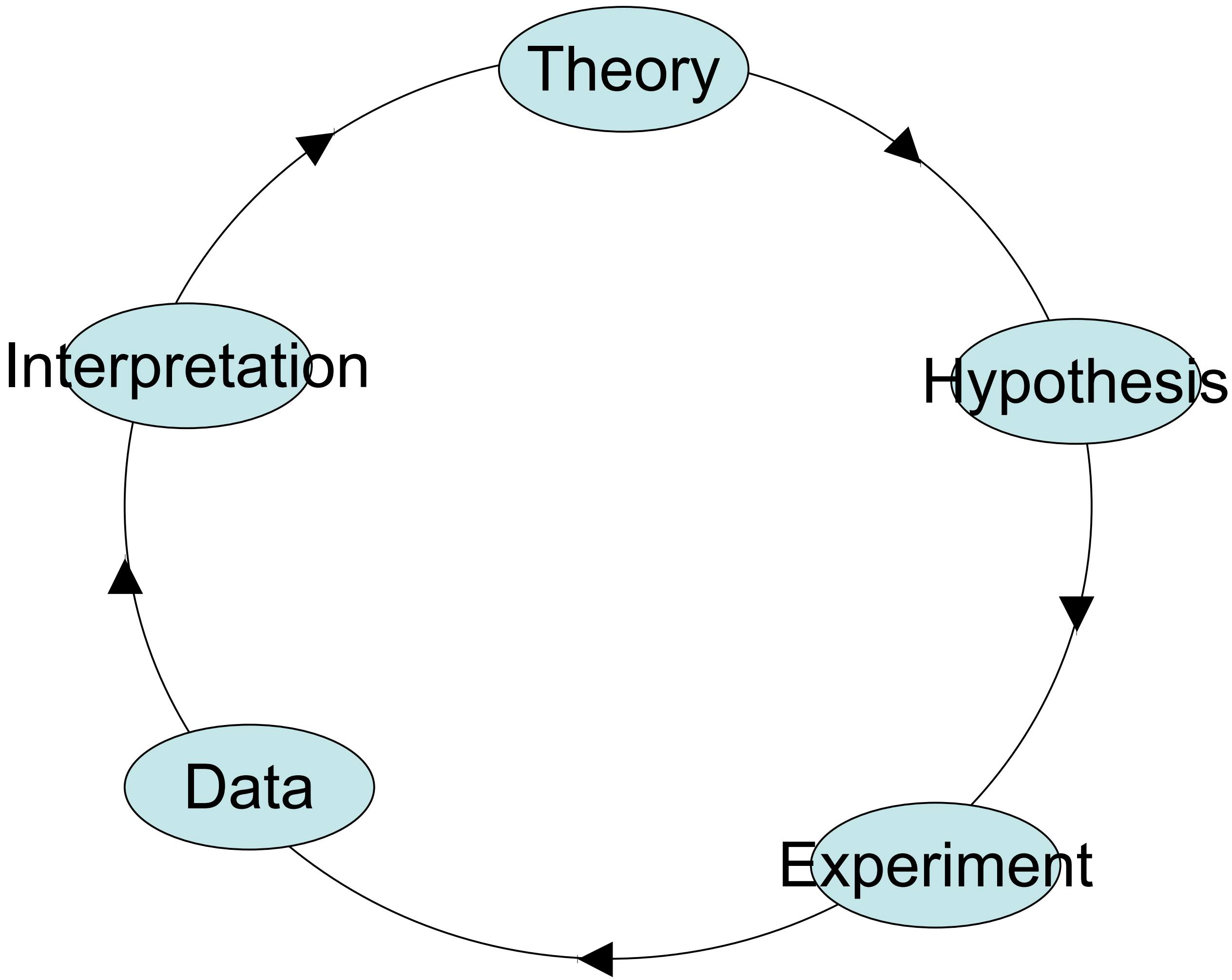
# Data, data, more data...

- Introspection (“arm-chair linguistics”) is extremely subjective
- Psycholinguistics is an empirical science: Theories are checked against data
- Two types of data collection:
  - Observation of natural data: corpus studies, collections of speech errors, long-term observation of what stages children go through in acquiring language, observation of your own behavior (e.g. garden-path effects), ...
  - More importantly: ***Experimental*** work

# What is an “Experiment”?

- Not just an attempt to see if something will work
- Systematic observation of a particular behavior under controlled circumstances
- Given a hypothesis, variation of a (single) factor to observe its influence on the way people comprehend/produce language
- Anything else that could influence the participants’ behavior is kept constant or otherwise controlled
- Therefore, if you observe a difference between conditions, it must be due to our manipulation

# The Research Cycle



# Some Research Questions

- How do people recognize words? What factors influence auditory and written word-recognition?
- How do people understand sentences?
  - How do they parse them? (top-down, bottom-up, ...)
  - Do ambiguous sentences take longer?
  - When there is an ambiguity, do people pursue both analyses concurrently or do they try one first and re-analyze? (Is the parser parallel or serial?)
  - When they make a mistake, how do they recover?
  - Why are some grammatical sentences difficult to understand?
- Do different levels of analysis influence each other or not, and how much / by what mechanism (modularity)?
- How do people produce language? What are the steps from concept to sound?
- How do bilinguals / 2nd language learners deal with several languages?

# (Some) Psycholinguistic Paradigms

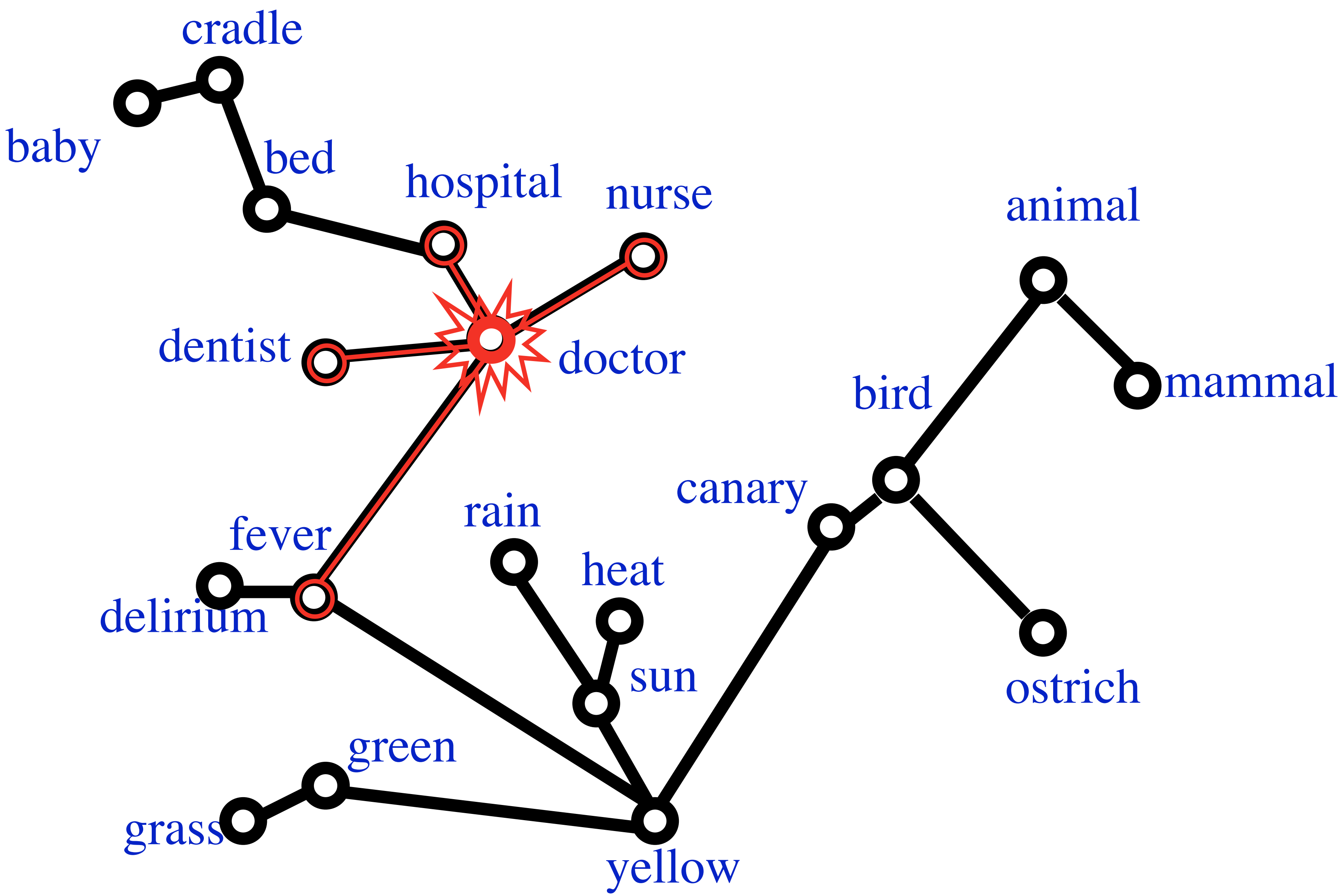
- Pen-and-Paper methods:
  - Rating studies, e.g. on a 7 point scale:
    - How similar are the words “water” and “rain”, “dog” and “puppy”
    - How grammatical is the sentence “T*he boy read the bread”*?
  - Sentence completion, e.g.
    - “The man raced the horse...”
    - “The child gave
- Nowadays on the web:

  http://www.language-experiments.org

# (Some) Psycholinguistic Paradigms

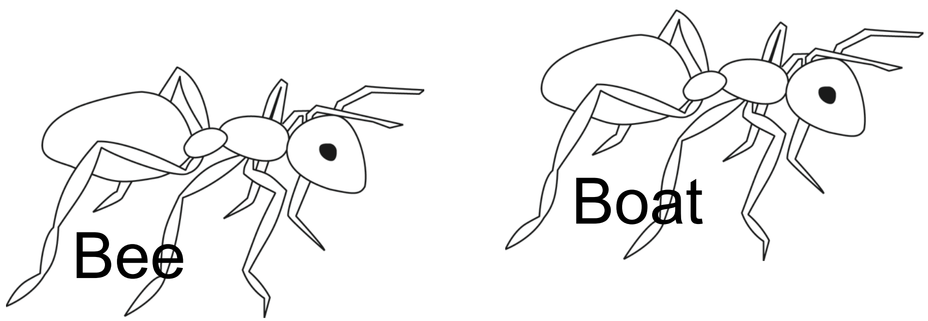
- Visual or auditory lexical decision
  - Stimuli: Words and pseudo-words (e.g. “poce”)
  - Task: Press yes if the stimulus is word, no otherwise
  - Demo: http://www.essex.ac.uk/psychology/experiments/lexical.html
  - Requires access to words in mental lexicon
  - Only word stimuli are analyzed
  - Properties of the words are manipulated (e.g. frequency)
- Priming
  - Show 1st stimulus (the “prime”)
  - Show 2nd stimulus (the “target”)
  - Depending on the 1st stimulus, reaction times to 2nd vary
  - E.g. Meyer and Schwaneveldt (1971): People are faster on “doctor” if preceded by “nurse” than if preceded by “butter”

# Spreading activation



# Paradigms (2)

- Cross-Modal Lexical Priming
  - Prime: spoken stimulus, Target: visual
- Phoneme-monitoring
  - Subjects listen to sentences or lists of unrelated words
  - Task: Press a button as soon as they hear a stimulus that contains the target sound
- Gating
  - Stimuli: Increasingly long segments of spoken words
  - Task: Guess what the word is
- Picture-Word Interference (production)



# Paradigms (3)

- Self-Paced Reading
  - Readers are presented with a blank sentence template
  - Each time a key is pressed, a word / phrase / segment is revealed
  - Latencies between key presses are measured

```
--- --- ---- -- --- ------- --- --------.
The man held -- --- ------- --- --------.
--- --- ---- at the station --- --------.
--- --- ---- -- --- ------- was innocent.
```

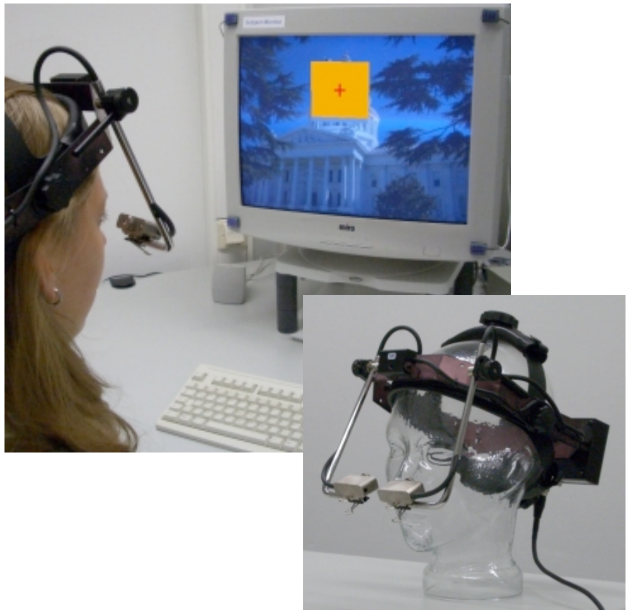
- Eye-tracking with written materials

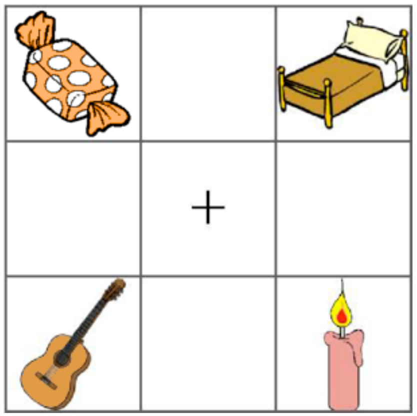
```
The man held at the station was innocent.
The man held at the station was innocent.
The man held at the station was innocent.
The man held at the station was innocent.
The man held at the station was innocent.
The man held at the station was innocent.
```

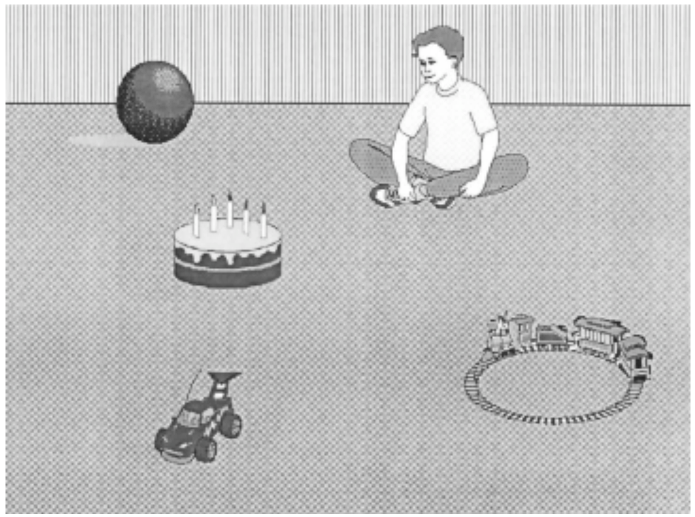
## Paradigms (4):

# Eyetracking in Visual Worlds:

- Show participants a scene / several objects
- Give them simple instructions to follow, e.g. “pick up the candy”, or have them listen to a description of the scene
- Eye-movements follow input at phoneme level or below
- People even ***anticipate*** if the structure of the sentence allows it

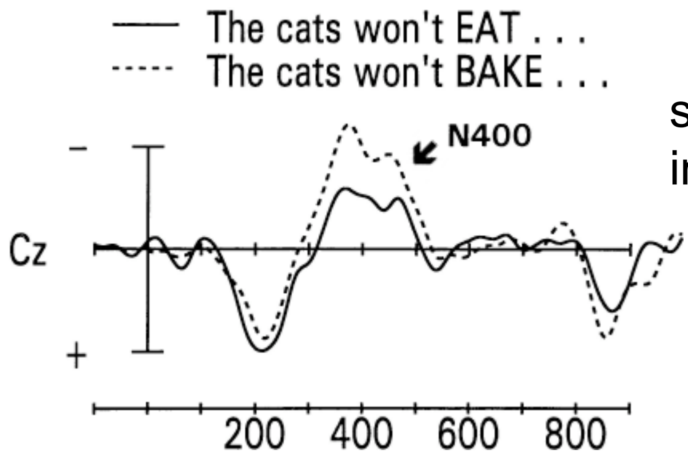






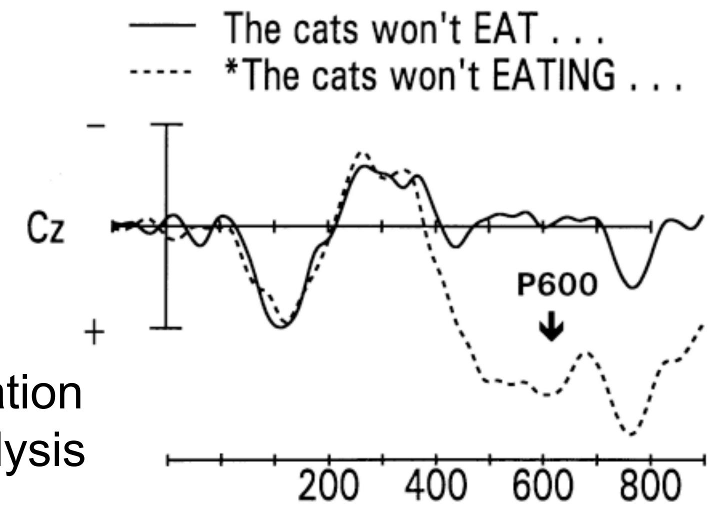
# Paradigms (5): Event-Related Potentials

- Subjects wear electrodes as for EEG
- They read sentences which are incorrect either semantically or syntactically
- The voltage change on the surface of scalp is measured and compared to correct sentences



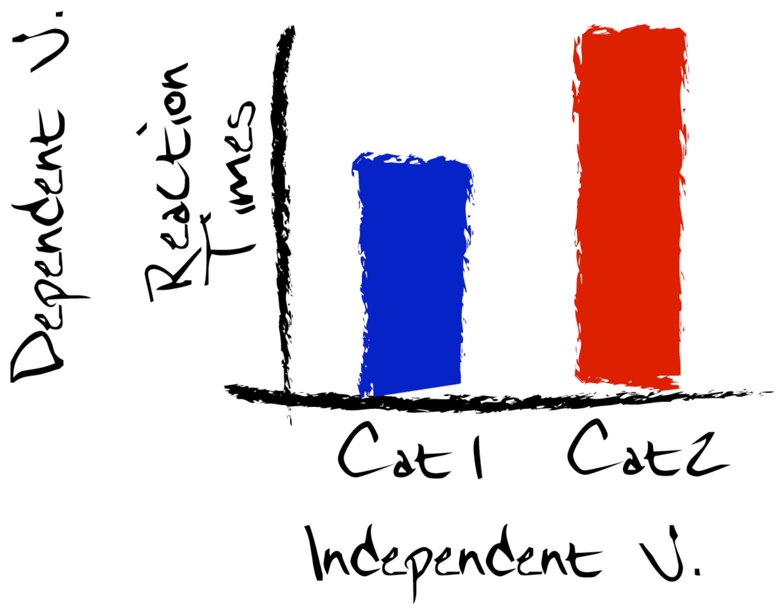

semantic integration

syntactic disambiguation and re-analysis



# Two Types of Variables

- The independent variable is the variable that you manipulate; it may have several “levels”
  - e.g. word length, frequency, semantic relationship, ...
- The dependent variable is the one you measure
  - e.g. reaction times, number of errors, proportion of looks to an object, voltage on brain surface, ...
- If you find a difference in your dependent variable, you say that you found an effect of the independent variable



# On-line and Off-line

- Off-line measures: Return only the end product of the process
  - Pen-and-paper methods
  - Lexical decision
  - ...
- On-line measures: Allow observation of the process as it unfolds
  - Gating
  - Self-paced reading
  - Eyetracking, ERPs

# No IV manipulation = No Experiment

- Example: Does sleep deprivation affect reaction times?
  - Deprive one group of people of sleep and then measure their RTs
  - Compare to a control group
- IV manipulation: sleep deprivation
- If we find a difference (and the groups were similar) we can draw a conclusion about a causal relationship: Sleep deprivation ***affects*** RTs
- The same people in reversed condition would likely have produced similar results

# No IV manipulation = No Experiment

- Bad example: Do smart people react faster?
  - Divide people into two groups: one smart, one dumb
  - Measure RTs.
- We are not manipulating the IV. Subjects are not assigned to one group randomly.
- We can't make any causal claim because other factors could be correlated with intelligence (motivation, attention to the task, etc.)

# No IV manipulation = No Experiment

- Give people a number of sentences to read and record their reading times or their comprehension
- Based on the data, try to group the sentences in groups of similar types and try to infer backwards what characteristics lead to the reading time patterns or comprehension patterns
- This isn't an experiment!
  - Nothing manipulated beforehand
  - Grouping of sentences after the fact (***post-hoc)***
- No strong conclusions can be drawn
  - Only speculations about the cause
  - There may be correlations but no causal link

# The Ideal Case

- Manipulate the IV and hold all other variables constant
- Nearly impossible, especially with human participants
  - different skills, IQ, experiences, and genes
  - how well they slept last night, how much they ate for lunch,...
- Instead: Avoid systematic confounds
  - Make sure there is no systematic assignment of subjects to conditions and no systematic differences in the sets of materials you use (use of databases/corpora and/or run pretests, then evenly distribute the effects of confounding factors)
  - To reduce subject variance, use same subjects in both conditions: ***within-subjects***
  - Counterbalance presentation
  - Control for order effects: Rotate through possible alternatives

# That's it for Today!

**Thanks to Berry Claus, Matt Crocker, Alissa Melinger, Andrea Weber, and others, who provided slides for me to work from :-)**