Introduction to psycholinguistic experiments:

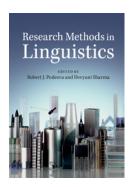
With a focus on reference resolution, attention and perception

Elsi Kaiser
University of Southern California

Workshop in experimental linguistics: from field to lab University of Tartu, April 11, 2019

Resources for further reading

- Edited by Robert J. Podesva & Devyani Sharma
- "Experimental paradigms in psycholinguistics (Chapter 8) by Elsi Kaiser
- "Experimental research design" (Chapter 7) by Rebekha Abbuhl, Susan Gass & Alison Mackey



What we can and cannot do today

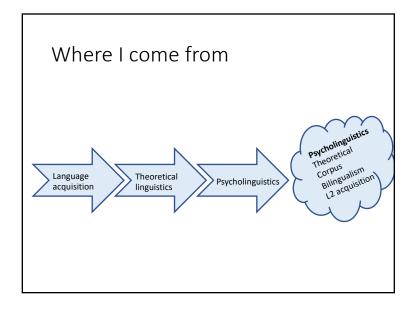
- Discuss some of the issues that come up regarding the design and implementation of experiments
- YES
- Differences between methods focusing on
 - processing load/ease of processing
 - Incremental interpretation
- YES
- Cover everything about how to run psycholinguistic experiments
- NO

Whom am I? Elsi Kaiser



- Department of Linguistics, University of Southern California
- Director of the USC Language Processing Lab
 - Reference resolution; information structure; interface(s) between discourse/syntax/semantics; perspective-taking, subjectivity
- PhD in Linguistics, MA in Psychology, BA in Germanic lang & lit, ling, French





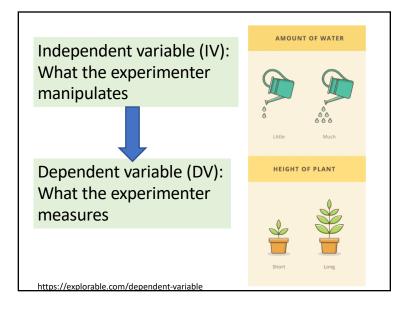


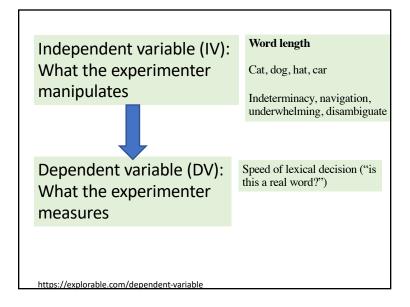
Topics we will cover today

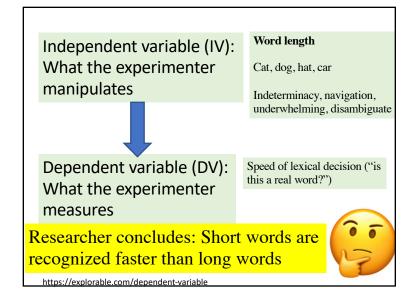
- General considerations
 - largely independent of method
- Choosing a method
 - Different methods tell you very different things
- Implementing your study
- Design and implementation considerations for specific methods

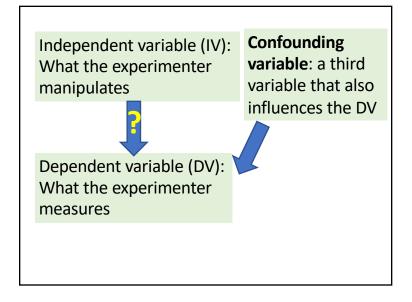
Quick: some basic terms
independent variable (factor)
dependent variable
confounding variable

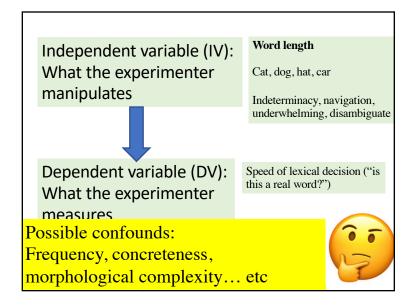
And is it dangerous to eat ice cream in New York City?

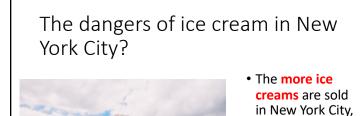




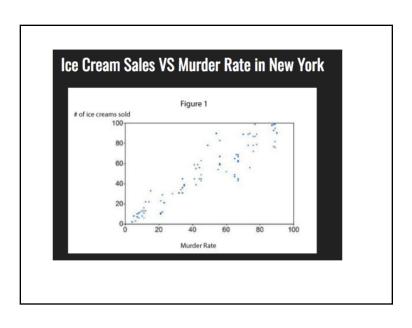


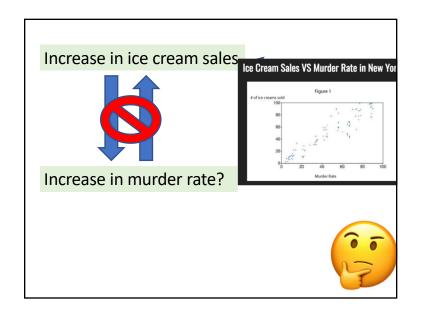






the higher the murder rate





No, that ice cream cone won't cause a murder

- Correlation does not equal causation
- Always be on the lookout for confounding variables
 - Is there something that could be influencing the dependent variable, other than the independent variable?
 - · Ask for feedback from others

2x2 within-subjects factorial design: Example from Kaiser & Trueswell 2008

- Finnish has multiple pronouns that can be used to refer back to humans. How is their use influenced by word order/topicality?
- Factor 1: Word order (SVO/OVS) 2 levels
- Factor 2: Pronoun type (hän/tämä) 2 levels
- Sentence completion task

	Personal pronoun hän (s/he)	Demonstrative tämä (this)
SVO	SVO. Hän	SVO. Tämä
OVS	OVS. Hän	OVS. Tämä

KISS and Goldilocks

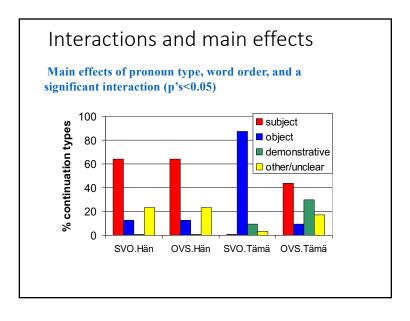
- 2x2, maybe 2x3
- not 6x6, 1x2
- Why have more conditions, test more things in one experiment?
 - Need enough datapoints per condition – study gets longer and people get tired
 - Stats get more complex
 - Interpretation of intearctions gets more complex



Analysis: What referent does hän/tämä refer to?

Sample item:

- Sentence 1:
 - · Niina was shopping at the grocery store.
- Sentence 2:
 - While waiting in line, she saw a cook with a white hat standing behind her.
- Sentence 3 (critical sentence):
 - SVO: 'The cook-SUBJ pushed a baker-OBJ standing at the back of the line.'
 - OVS: 'The cook-OBJ pushed a baker-SUBJ standing at the back of the line.'
- Sentence 4: (continuation prompt)
 - Hän / Tämä...



Questions Without easy answers

How many target items are necessary?

- It depends on the effect size!
 - Check recently-published work using similar methods
 - Very rough rule of thumb: 6-8 items per condition
 - 4 conditions: 4 x 8 = 32 targets (or more...)

How many filler items do I need?

- "Enough!"
- 1:1?
- 1:1.5?
- 1:2?
- How detectable/noticeable are your targets?
- How dangerous is it if people figure out what you are testing?
- Less fillers than targets? Bad idea....

heard that Ben and David rode their bikes to the beach. David buried the little girl with the shovel.	Who had the shovel?	David	The little girl
talked to Jack, who had to wear a suit to the basketball game. He was upset about the tie.	What was Jack upset about?	The score of the basketball game	The piece of cloth around his neck
Melissa walked over to Anna. She obviously needed some help with the math homework.	Who needed help with the math homework?	Melissa	Anna
understand that Rosie and Lisa went to a family barbeque for the Fourth of July. Rosie pinched the chef with the tongs.	Who had the tongs?	Rosie	The tongs
Charlotte brought her racquetball equipment with ner on the day of the trial. She wondered if she should bring it to the court.	Where would Charlotte put her equipment?	The space in the gym for racquetball	The site of her testimony

Amy punched Lisa. She knew that cheating on the geography test would get them in trouble.	Who knew they would get in trouble?	Amy	Lisa
Luke glanced at Andrew warily. He'd probably put toothpaste in the shampoo bottle again.	Who put the toothpaste in the shampoo bottle?	Luke	Andrew
Will and Greg went to the park to play fetch with their dog. Craig pointed at the dog with the stick.	Who had the stick?	Greg	The dog

- Targets can seem repetitive
 - Create repetitive subclasses of fillers?
- Targets are often 'hard'
 - Make sure fillers do not seem too easy
- Use some fillers as 'catch' trials to make sure people are paying attention
- If rating acceptability: Ensure that people see a range of acceptabilities across the course of the experiment

Providing instructions to participants

Sometimes the instructions are obvious!

Put the frog on the napkin in the box. (Tanenhaus et al. 1995)

The task is crystalclear, no risk of misunderstanding



From J Gerard

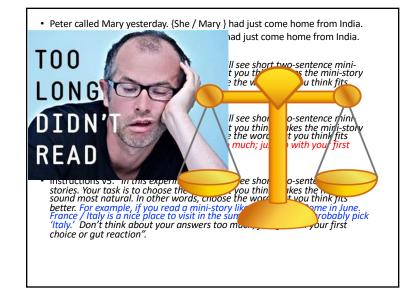
"If you do not explain to subjects what you want, each one takes his or her own interpretation and the results are meaningless" (Schütze 1996:132)

· Hill (1961)

- 10 participants (3 linguists, quite a few English professors)
- Instructions: "reject any sentences which were ungrammatical, and to accept those which were grammatical"
- No definition or explanation of these terms was given to participants
- "Those man left early." → all 10 participants rejected this
- "The child seems sleeping." → 6 participants rejected this
- · Some reasons for accepting; "it sounds like poetry"
- · Some reasons for rejecting: it did not start with a capital letter

Example with pronoun use

- Let's imagine that we want to do a forced-choice experiment to investigate pronoun use.
- The participants' task is to choose which form, out of the two given in brackets, is better.
- Peter called Mary yesterday. {She / Mary } had just come home from India.
- Peter called Mary yesterday. {He / Peter } had just come home from India.
- Instructions v1: "Choose the word which sounds best."
 - Pronouns vs. nouns? 'best' in what sense ?
- Instructions v2: "Choose the word which sounds best in the context of this sentence"
 - but we want people to consider sequence of sentences. (Prescriptive rules like 'Do not start a sentence with a pronoun'??)



University of Tartu 2019

- Peter called Mary yesterday. {She / Mary } had just come home from India.
- Peter called Mary yesterday. {He / Peter } had just come home from India.
- Instructions v3: "In this experiment, you will see short two-sentence ministories. Your task is to choose the word that you think makes the mini-story sound most natural. In other words, choose the word that you think fits better".
- Instructions v4: "In this experiment, you will see short two-sentence ministories. Your task is to choose the word that you think makes the mini-story sound most natural. In other words, choose the word that you think fits better. Don't think about your answers too much; just go with your first choice or gut reaction".
- Instructions v5: "In this experiment, you will see short two-sentence ministories. Your task is to choose the word that you think makes the ministory sound most natural. In other words, choose the word that you think fits better. For example, if you read a ministory like "Lisa visited Rome in June. France / Italy is a nice place to visit in the summer", you would probably pick 'Italy.' Don't think about your answers too much; just go with your first choice or gut reaction".

What is "Latin square" and why do we need it?

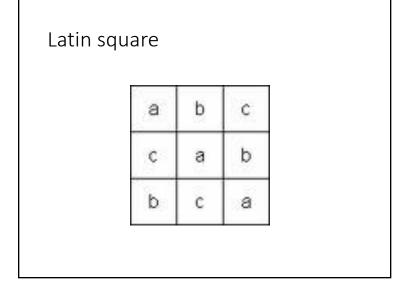
What kind of practice items? Are they necessary?

- How many practice items do you need? (order of choices?)
- Should participants be given feedback?
- Dangers of putting people in 'prescriptivist mode'
- a. Lisa and Tom went skiing. {He / Tom} is a very good skier.
- b. Lisa went skiing with Tom. {Tom / He} is a very good skier.
- c. Lisa and Tom went skiing. {The weather/the ocean} was beautiful.
- d. Mickey Mouse has big ears. {Are / is} they red?
- e. Andy tickled the frog with the napkin. {The frog / Andy} was holding the napkin.
- •
- If target items are ambiguous/ hard to answer, it can be a good to have some 'hard practice items too (but not hard in the same way...)
- Give you the opportunity to explain to the participant that if they aren't sure what the answer is, they should just go with their 'gut reaction.'

Latin Square

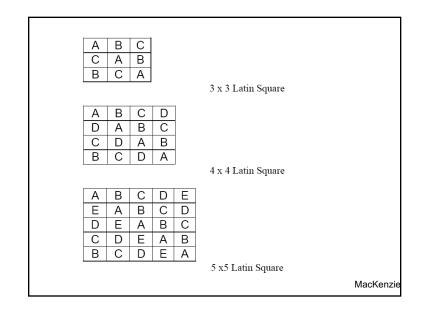
- The banker who the chairman informed invested a million in a start-up.
- The violinist who flattered the cellist played a piece from the symphony.
- The physician who consulted the cardiologist checked the files in the office.
- The physician who the cardiologist consulted checked the files in the office.
- The banker who informed the chairman invested a million in a start-up.
- The violinist who the cellist flattered played a piece from the symphony.





Latin square

• "A Latin square is an n x n array, when each member of the array is one of the numbers from 1 to n. Each number being between 1 to n is written once in each row and column". (Marsh & Dunton)



Elsi Kaiser emkaiser@usc.edu

Latin square in experimental design

- Example: processing of RCs
- Factor 1: RC type
 - · Subject-extracted RC
 - Object-extracted RC
- Factor 2: Memory load
 - · One memory-noun
 - Three memory nouns
- 2 x 2 design = 4 conditions

	Subj- RC	Obj- RC
1-MN	а	b
3-MN	С	d

	Subj-RC	Obj-RC
1-MN	A= The physician who consulted the cardiologist checked the files in the office.	B = The physician who the cardiologist consulted checked the files in the office.
3-MN	C = The physician who consulted the cardiologist checked the files in the office.	D = The physician who the cardiologist consulted checked the files in the office.

Sample RCs

- ITEM #1
- The physician who consulted the cardiologist checked the files in the office.
- The physician who the cardiologist consulted checked the files in the office.
- ITEM #2
- The banker who informed the chairman invested a million in a start-up.
- The banker who the chairman informed invested a million in a start-up.
- ITEM #3:
- The violinist who flattered the cellist played a piece from the symphony.
- The violinist who the cellist flattered played a piece from the symphony.

Another example

Mommy will put away my car keys. [sing_1]

Mommy will put away her car keys. [sing_3]

Mommy and Daddy will put away our car keys. [pl_1]

Mommy and Daddy will put away their car keys. [pl_3]

agreement with "imposter" mouns, see e.g. Kaiser et al 2018

- How do you set up the conditions of your experiment?
- What does each participant get to see?

One option: each participant sees everything

Item-number	Condition	
Item1	Α	
Item1	В	
Item1	С	
Item1	D	
Item2	A	
Item2	В	
Item2	С	
Item2	D	
item3	A	

Avoiding repetition

- BUT We don't want one participant seeing all versions of an item
 - · Because that would be repetitive:
 - · Mommy will put away my car keys.
 - · Mommy will put away her car keys.
 - Mommy and Daddy will put away our car keys.
 - Mommy and Daddy will put away their car keys.
- BUT we want the all versions of an item to be seen by *someone*
- Use Latin Square

	List1	List2	List3	List4
Item1	Α	В	С	D
Item2	D	Α	В	С
Item3	С	D	Α	В
Item4	В	С	D	Α
Item5	Α	В	С	D
Item6	D	Α	В	С
Item7	С	D	Α	В
Item8	В	С	D	Α
Item9	Α	В	С	D
Item10	D	Α	В	С
Item11	С	D	Α	В
Item12	В	С	D	Α

	List1	List2	List3	List4
Item1	Α	В	С	D
Item2	D	A	В	С
Item3	С	D	A	В
Item4	В	C	B	A
Item5	Α	В	C	B.
Item6	D	Α	В	C
Item7	С	D	Α	В
Item8	В	С	D	Α
Item9	A	В	С	D
Item10	D	Α	В	С
Item11	С	D	Α	В
Item12	В	С	D	Α

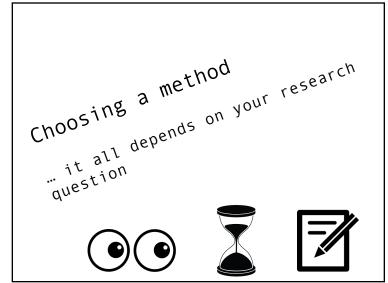
Latin Square

- Four conditions
- Twelve items
- Each participant sees 3 items in cond A, 3 items in cond B, 3 items in cond C, 3 items in cond D
 - (ideally at least 4 or 5 items or more per condition)
- Each item is seen by different participants in each of the four conditions

	List	:1	List2	List3	List4
Item1	Α		В	С	D
Item2	D		Α	В	С
Item3	С	Shou	ld I run my	experimer	t using
Item4	В	this e	xact set up	o, as showr	here?
Item5	Α		В	С	D
Item6	D		Α	В	С
Item7	С		D	Α	В
Item8	В		С	D	Α
Item9	Α		В	С	D
Item10	D		Α	В	С
Item11	С		D	A	В
Item12	В		С	D	Α

Potential problems

- Right now, every participant encounters targets in the order a,b,c,d
 - Possible consequences of exposure to one condition on how another condition is processed?
- Targets are in order (item1-item12)
 - Why might this also be undesirable?
- Better to randomize
 - Or pseudorandomize!
- Abbuhl et al (2014, ch 7 in 'Research Methods in Linguistics')



Two kinds of methods

Methods that look at processing load

Methods that look at interpretation





Load vs. interpretation

- Processing load: Is a particular structure or a particular interpretation especially hard to process?
 - Self-paced reading
 - Reading during eyetracking
- Incremental interpretation: What interpretations or parses do people consider before reaching their final interpretation?
 - Visual world eyetracking

Processing load

- As a person comprehends a sentence, processing load is different at different points in the sentence
 → measure the moment-by-moment fluctuations in the 'burden of work' by using on-line methods
- Often, but not always: If A takes longer to process than B, A involves a higher processing load than B



Processing load is always relative

- "Sentence X was processed slowly, which indicates a high processing load."
- · What's wrong with this claim?
- "Sentence X was processed more slowly than sentence Y, which suggests that X is harder to process than Y"
- What does it mean for something to be 'hard to process'?

considerations to keep in keep to methods

considerations to keep in methods

specific methods

reading reading reacking reading reacking reacking reading reacking reacking reading reacking reacking reading reacking rea

Sources of reaction time (RT) data:

- · Self-paced reading
- Speech rate
- · Reading during eye-tracking
- Speed of yes/no response
- · Speed of answer choice
- Auditory lexical decision*
- · Visual lexical decision*
- What about scales/rating tasks (Likert scale etc)?

Self-paced reading

• ---- ---- ---- ----

Self-paced reading

• John -----

Self-paced reading

• ---- thought ---- --- ---

Self-paced reading

• ---- that ---- ---

Self-paced reading

•--- ---- Bill ---- ---

Self-paced reading

• ---- owed --- ---

Self-paced reading

•---- him --

Self-paced reading

• ---- --- --- --- --- another -----

Self-paced reading

• --- chance -- ---

Self-paced reading

• ---- ---- to -----

Self-paced reading

•--- ---- -- solve --- ----

Self-paced reading

• --- --- -- --- the -----

Self-paced reading

---- ---- --- problem.

Self-paced reading

• John thought that Bill owed him another chance to solve the problem.

- (1c) The brother // of the mayor // who complimented himself // constantly // was // bothered by the reporter
- (1c) The brother // of the mayor // who complimented // himself // constantly // was // bothered // by the reporter

Self-paced reading

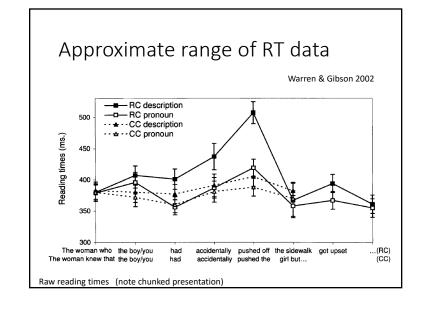
- (1a) The brother of the mayor who complimented himself smiled.
- (1b) The sister of the mayor who complimented himself smiled.
- (adapted from Swets et al.)
- → Any problems?

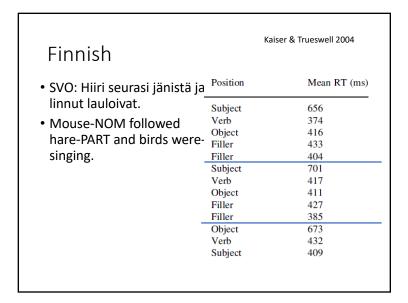
Designing sentences for reading time experiments

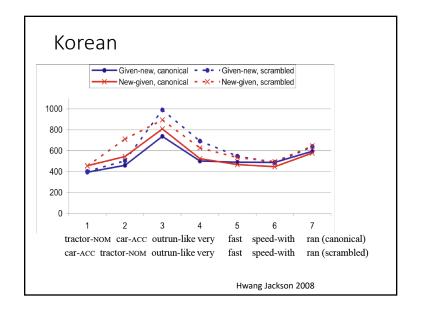
- For self-paced reading: Spillover regions
- (different for eye-tracking during reading)
- What else can influence reading time?
- Word length
- Frequency
- End-of-sentence wrap up effects

Analyzing reaction time data

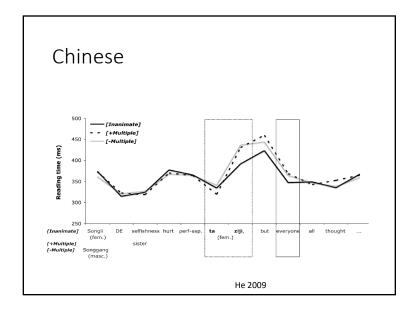
- What do the numbers look like?
- How to deal with outliers (extremes)?





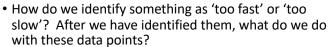


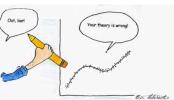
Elsi Kaiser emkaiser@usc.edu 20



Dealing with outliers?

- Very fast reaction times?
- Very slow reaction times?





Outliers

- Outlier: "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs." (Grubbs 1969)
- Step 1 : identify outliers
- Step 2: what to do with them? **Check recent papers*
 - Keep (do nothing)
 - Exclude
 - Adjust

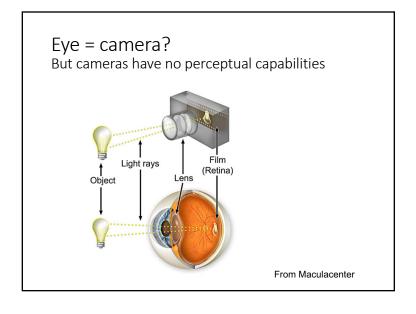
Grubbs, F. E.: 1969, Procedures for detecting outlying observations in samples. Technometrics 11, 1–21.

Two main types of eyetracking

- Visual world eye-tracking
- Eye-tracking during reading

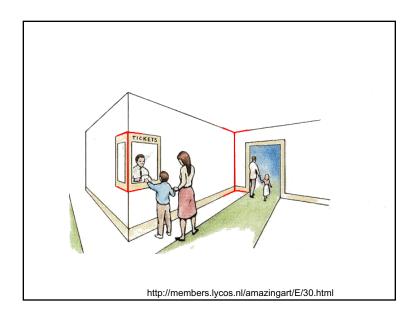


Kamide et al 2003 http://www.metrovision.fr/mv-po-notice-us.html



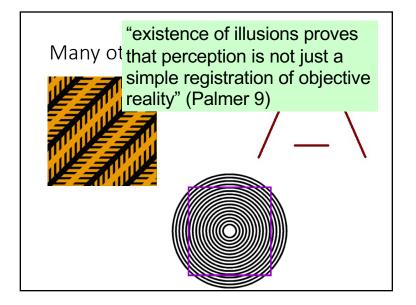
Role of perception

- Our eyes do not just 'take photographs' of what we see
- Importance of perception
 - "Actively constructed, meaningful model of the environment" (Palmer)
 - Perception is not always veridical ('truthful')
 - mostly "what you see is what you get" (evolutionary benefit, p.6)
- Vision is not a 'direct window into reality'



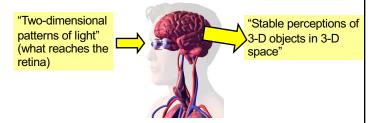
Ebbinghaus illusion http://members.lycos.nl/amazingart/E/30.html

Elsi Kaiser emkaiser@usc.edu



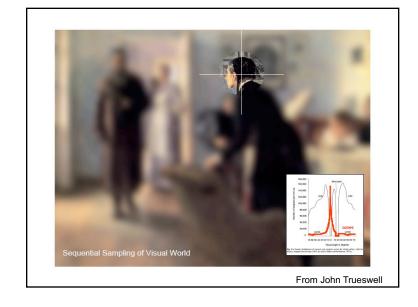
Role of perception – what the mind does with the input to the eye

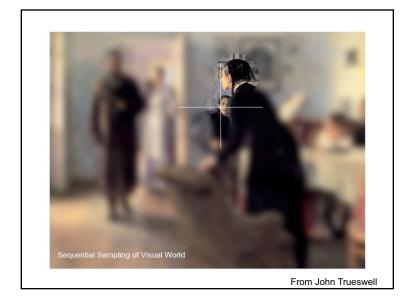
- Perception is fallible
- Not a "clear window onto external reality"
- Importance of the interpretative process

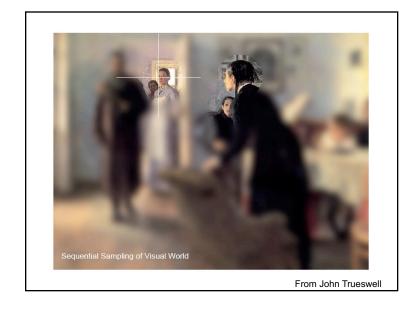


Functions of eye movements (see Palmer, 520)

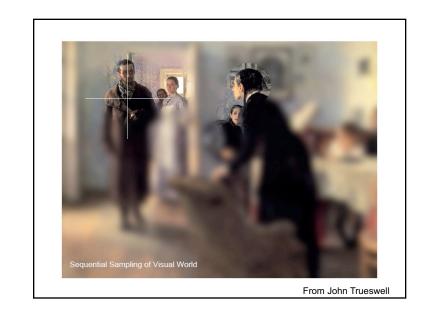
- Fixation: position target objects of interest on the fovea where visual acuity is highest
- Saccades:
- Tracking: to keep fixated objects on the fovea despite movements of the objects or the observer's head

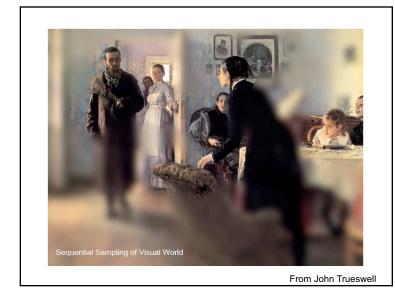


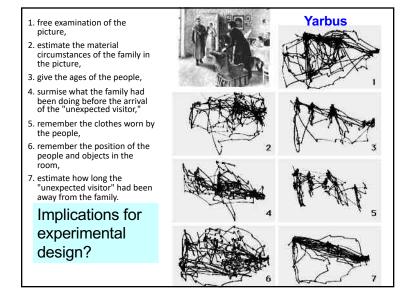


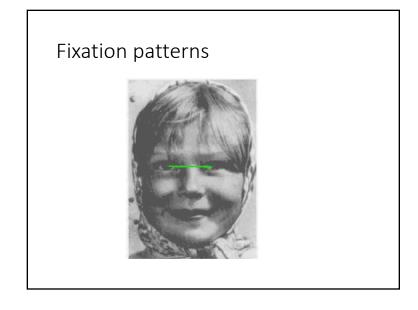




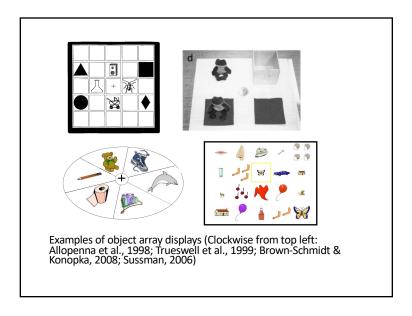


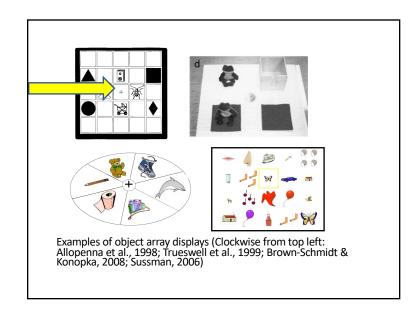


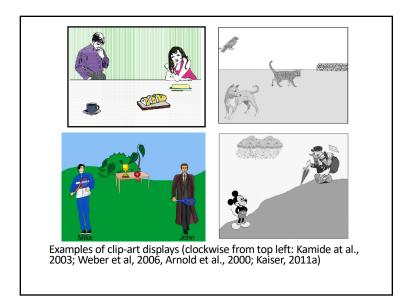




Designing displays for displays tracking displays tracking displays for devented eye-tracking visual world eye-tracking







- (6) Sample item [S = name | O = pro]
 - a. Mike did very well in last month's tennis tournament.
 - b. John congratulated him enthusiastically yesterday. (Critical sentence)
 - The prizes for the best-ranked tennis players were about to be announced, and
 - d. he was holding a new yellow tennis racket. (Test sentence)
 - e. Everyone was in a good mood that day.



Look-away in a neutral location before the critical word

Kaiser, Elsi (2011). Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus, *Language and Cognitive Processes*, 26, 1625-1666.



; images

of the critical objects.

Bias to look first to the top-left end to look from left to right.

er critical objects) should times in each of the four



objects raphic images?

When designing images

- Balancing the locations of the critical objects.
 - Languages read L-to-R: Bias to look first to the top-left corner of the screen, tend to look from left to right.
 - The target object (and other critical objects) should occur an equal number of times in each of the four quadrants. (or L vs R)
- Size and visual salience of objects
- Quality of clip art? Photographic images?
- Cultural differences?

A limitation: Need to find images

- What about things that are hard to depict?
- Solution: show words!
- McQueen & Viebahn 2007
- Holsinger (2011): idioms
- To kick the bucket
- To cut the mustard
- To hold your horses



Processing of idioms

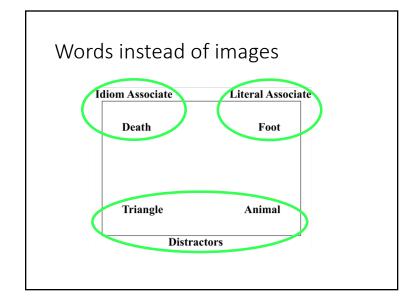
	Syntactically Available	Syntactically Unavailable	
Lexically Available	John kicked the bucket last Thursday	It was surprising to see someone as skilled as John completely miss when he kicked. The bucket full of orange slices was destroyed when he accidentally missed the ball.	
Lexically Unavailable	John kicked the pail last Thursday	It was surprising to see someone as skilled as John completely miss when he kicked. The bucket full of orange slices was destroyed when he accidentally missed the ball.	

Holsinger, E. (2011). *Cutting the mustard: An experimental investigation of idiomatic expressions in the lexicon*. University of Southern California Doctoral Dissertation.

Words instead of images

Death Foot

Triangle Animal



Tasks, dependent variable?

- Passive listening
 - DV: fixations ('looks')
- Listen and carry out a task
 - DV: fixations
 - · Offline DV: click data
- · What kinds of tasks?
 - Task-as-instruction: Put the frog on the napkin in the basket
 "If there is a mistmatch between picture and story click on
 - "If there is a mistmatch between picture and story, click on where the error is" (Kaiser 2011 etc)
 - "Click on the last-mentioned person or thing" (Kaiser 2009 etc)

Click on where the mistake is

- (6) Sample item [S = name | O = pro]
 - a. Mike did very well in last month's tennis tournament.
 - b. John congratulated him enthusiastically yesterday. (Critical sentence)
 - The prizes for the best-ranked tennis players were about to be announced, and
 - d. he was holding a new yellow tennis racket. (Test sentence)
 - e. Everyone was in a good mood that day.



Kaiser 2011

Elsi Kaiser emkaiser@usc.edu

Click on where the mistake is

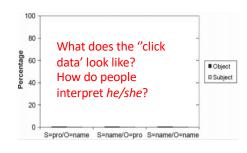
- (6) Sample item [S = name/O = pro]
 - a. Mike did very well in last month's tennis tournament.
 - b. John congratulated him enthusiastically yesterday. (Critical sentence)
 - The prizes for the best-ranked tennis players were about to be announced, and
 - d. he was holding a new yellow tennis racket. (Test sentence)
 - e. Everyone was in a good mood that day.



Both people are holding tennis rackets, but neither is YELLOW. Who would you click?

Kaiser 2011

- (7) Sample item [S = name | O = name]
 - b. *Greg* congratulated *John* enthusiastically yesterday. (Critical sentence)
 - The prizes for the best-ranked tennis players were about to be announced, and
 - d. he was holding a new yellow tennis racket. (Test sentence)
 - e. Everyone was in a good mood that day.



Kaiser 2011

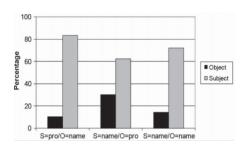
(6) Sample item [S = name | O = pro]

- a. Mike did very well in last month's tennis tournament.
- b. John congratulated him enthusiastically yesterday. (Critical sentence)
- The prizes for the best-ranked tennis players were about to be announced, and
- d. he was holding a new yellow tennis racket. (Test sentence)
- e. Everyone was in a good mood that day.
- (5) Sample item [S = pro/O = name]
 - a. Greg is always very supportive of others.
 - b. *He* congratulated *John* enthusiastically yesterday. (Critical sentence)
 - The prizes for the best-ranked tennis players were about to be announced, and
 - d. he was holding a new yellow tennis racket. (Test sentence)
 - e. Everyone was in a good mood that day.

Kaiser 2011

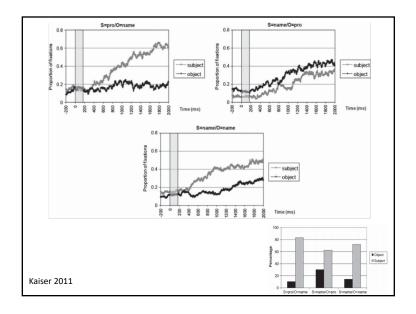
(7) Sample item [S = name | O = name]

- b. *Greg* congratulated *John* enthusiastically yesterday. (Critical sentence)
- The prizes for the best-ranked tennis players were about to be announced, and
- d. he was holding a new yellow tennis racket. (Test sentence)
- e. Everyone was in a good mood that day.



Kaiser 2011

Elsi Kaiser emkaiser@usc.edu



What about corrective focus?

(8) Speaker A: I heard that Greg congratulated Mike enthusiastically yesterday.

Speaker B:

- a. No, that's not quite right.
- b. (i) He congratulated John. (Critical sentence: [SVO.Object = focus])
 (ii) John congratulated him. [SVO.Subject = focus]
 (iii) It was John that he congratulated. [Cleft.Object = focus]
 - (iii) It was John that he congratulated. [Cleft.Object = focus] (iv) It was John who congratulated him. [Cleft.Subject = focus]
- c. The prizes for the best-ranked tennis players were about to be announced, and
- d. he was holding a new yellow tennis racket. (Test sentence)
- e. Everyone was in a good mood that day.

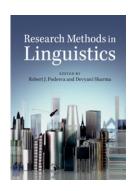
See Kaiser (2011), LCP for details!

Taking stock

- Basic structure of experiments
 - Targets, fillers, instructions, practice
 - Latin Square
- Methods that measure processing load
 - · Reading times
 - outliers
- Methods that measure interpretation
 - Visual attention
 - · Design considerations

Resources for further reading

- Edited by Robert J. Podesva & Devyani Sharma
- "Experimental paradigms in psycholinguistics (Chapter 8) by Elsi Kaiser
- "Experimental research design" (Chapter 7) by Rebekha Abbuhl, Susan Gass & Alison Mackey



Thank you!

- Elsi Kaiser
- emkaiser@usc.edu