Embodiment

CS786

September 23rd 2024

Recap: the mind constructs information

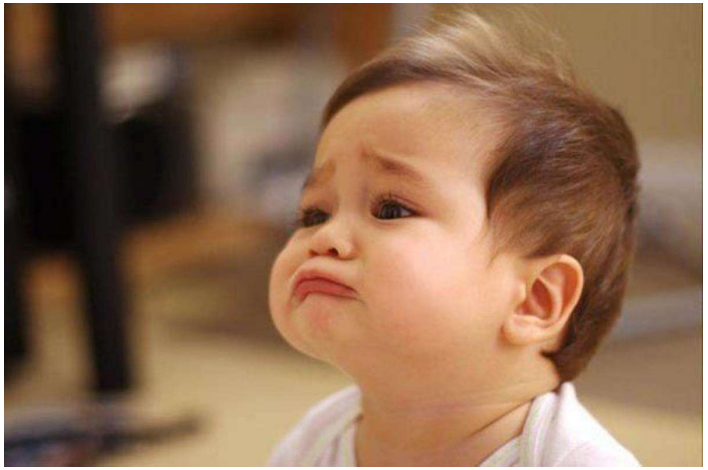
- Current neuroscience and AI fashions argue for the brain as a passive store of information
- The mind is not a passive receptacle
- It exists in a body with a long history and a complex present
- It constructs information based on being in the world (Heidegger)

The Debate on the Joy of Fish

- Zhuangzi and Huizi were enjoying themselves on the bridge over the Hao River. Zhuangzi said, "The minnows are darting about free and easy! This is how fish are happy.“
- Huizi replied, "You are not a fish. How do you know that the fish are happy?" Zhuangzi said, "You are not I. How do you know that I do not know that the fish are happy?“
- Huizi said, "I am not you, to be sure, so of course I don't know about you. But you obviously are not a fish; so the case is complete that you do not know that the fish are happy.“
- Zhuangzi said, "Let's go back to the beginning of this. You said, How do you know that the fish are happy; but in asking me this, you already knew that I know it. I know it right here above the Hao."

How do we know which dog is happy
And which is sad?

Where do cuteness judgments come from?



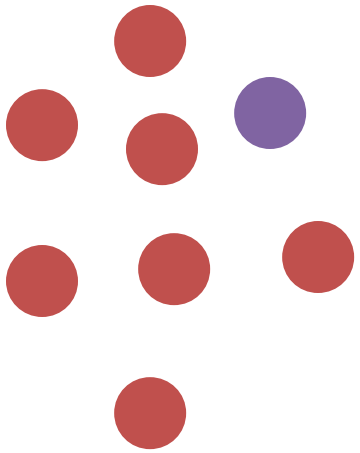
Embodied cognition

- The claim that cognition is deeply dependent on the fact that we inhabit bodies of flesh and blood
- Always in tension with efforts to abstract human-like ‘intelligence’ into machines
- Flavors of embodiment
  - Perceptual
  - Memory
  - Social

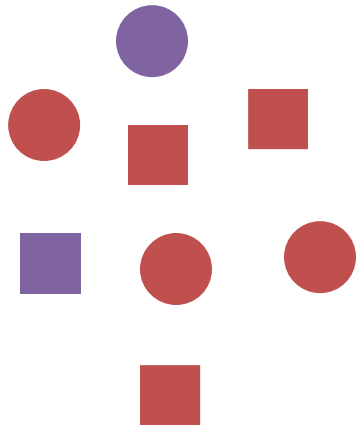
Embodiment in Action

VISUAL PRIORS

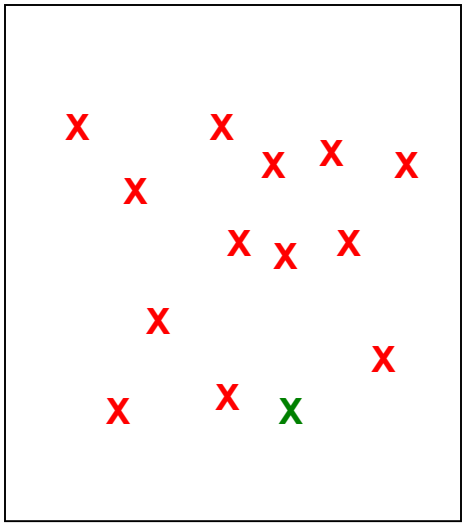
Visual search (easy)



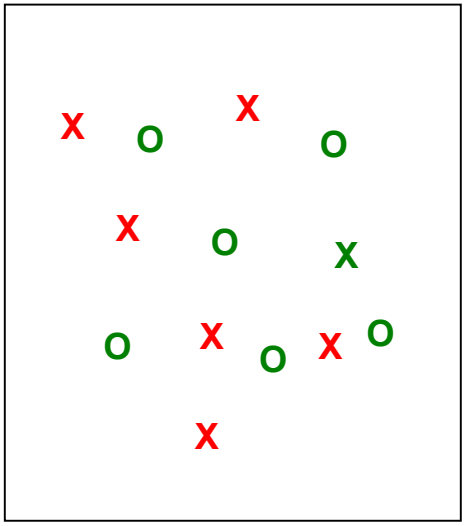
Visual search (hard)



Visual search



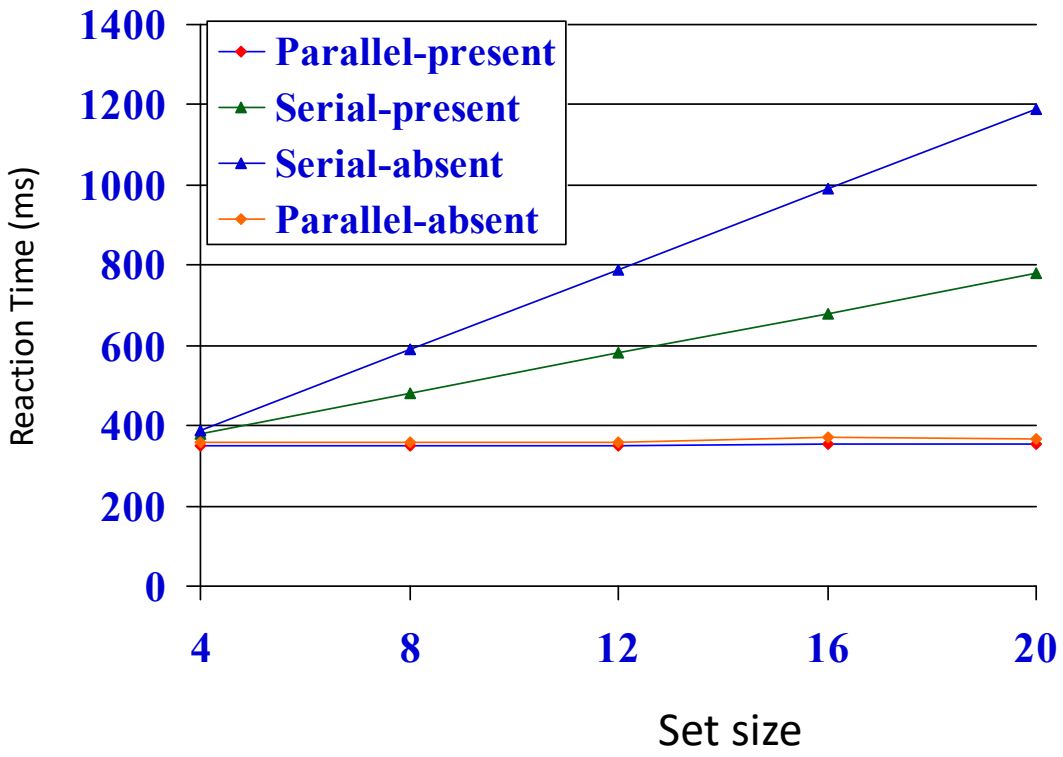
Feature search



Conjunction search

Treisman & Gelade 1980

“Serial” vs “Parallel” Search



Feature Integration Theory: Basics (FIT) Treisman (1988, 1993)

- Distinction between objects and features
- Attention used to bind features together ("glue") at the attended location
- Code 1 object at a time based on location
- Pre-attentional, parallel processing of features
- Serial process of feature integration

FIT: Details

- Sensory “features” (color, size, orientation etc) coded in parallel by specialized modules
- Modules form two kinds of “maps”
  - Feature maps
    - color maps, orientation maps, etc.
  - Master map of locations

Feature Maps

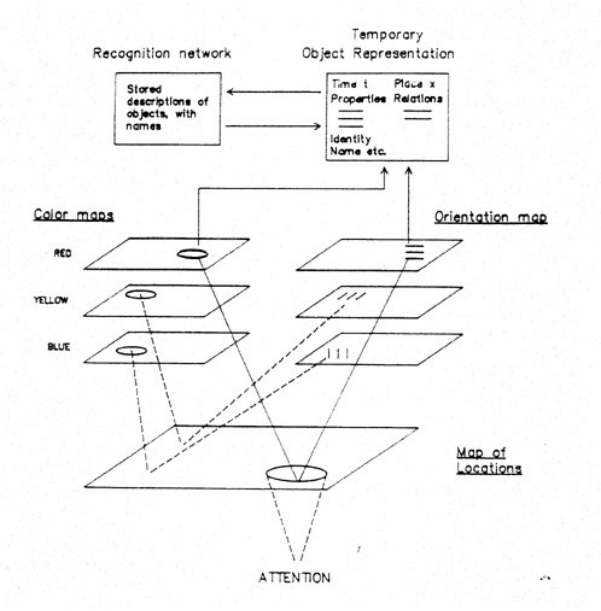
- Contain 2 kinds of info
  - presence of a feature anywhere in the field
    - there's something red out there…
  - implicit spatial info about the feature
- Activity in feature maps can tell us what's out there, but can't tell us:
  - where it is located
  - what other features the red thing has

Master Map of Locations

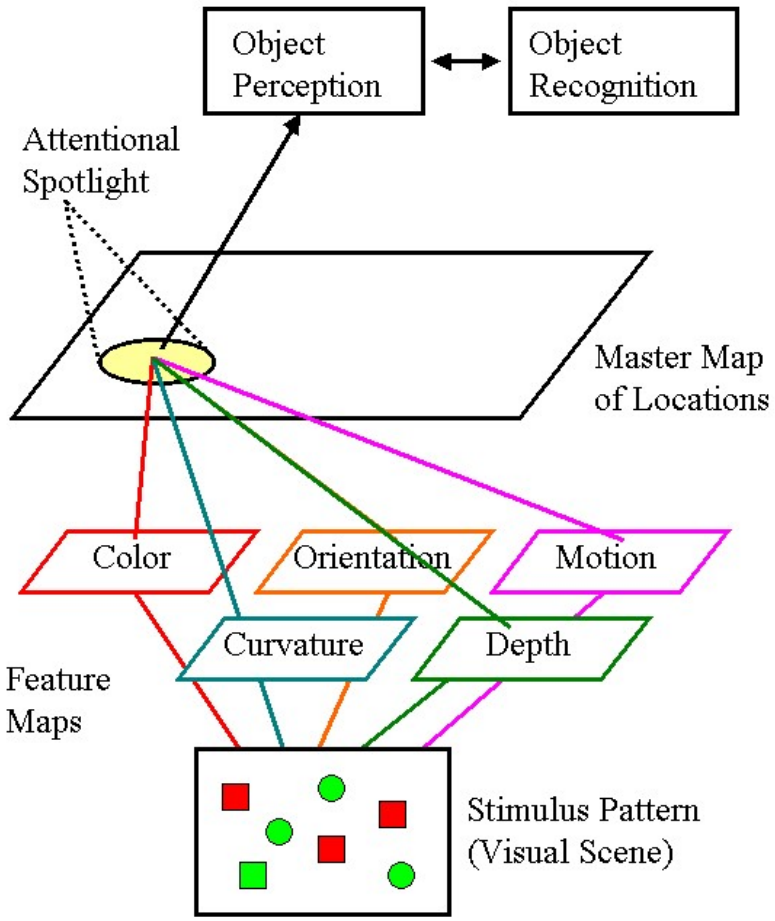
- codes where features are located, but not which features are located where
- need some way of:
  - locating features
  - binding appropriate features together
- [Enter Focal Attention…]

Role of Attention in FIT

- Attention moves within the location map
- Selects whatever features are linked to that location
- Features of other objects are excluded
- Attended features are then entered into the current temporary object representation



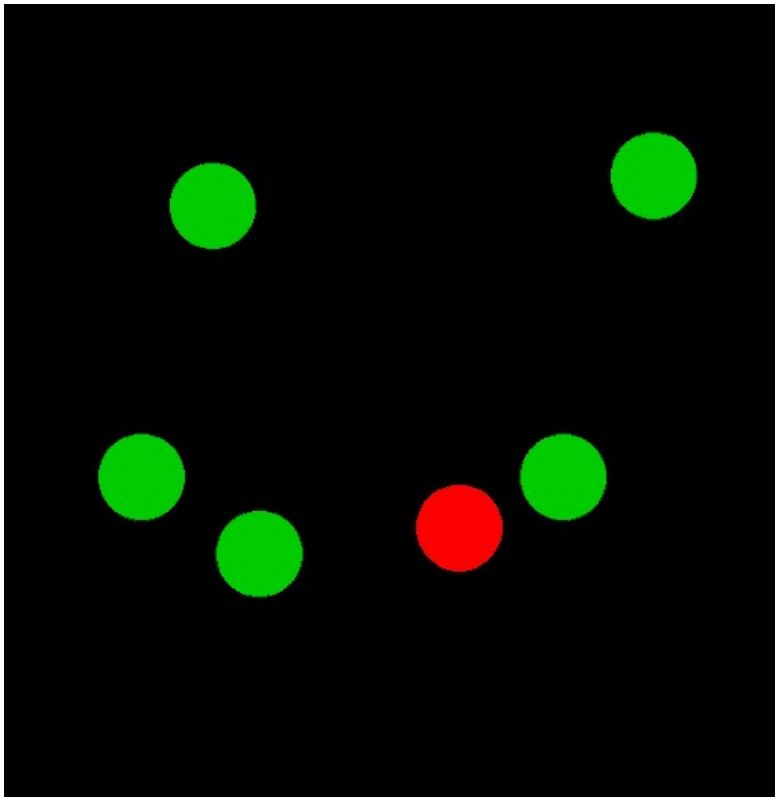
Feature Integration Theory (Treisman)



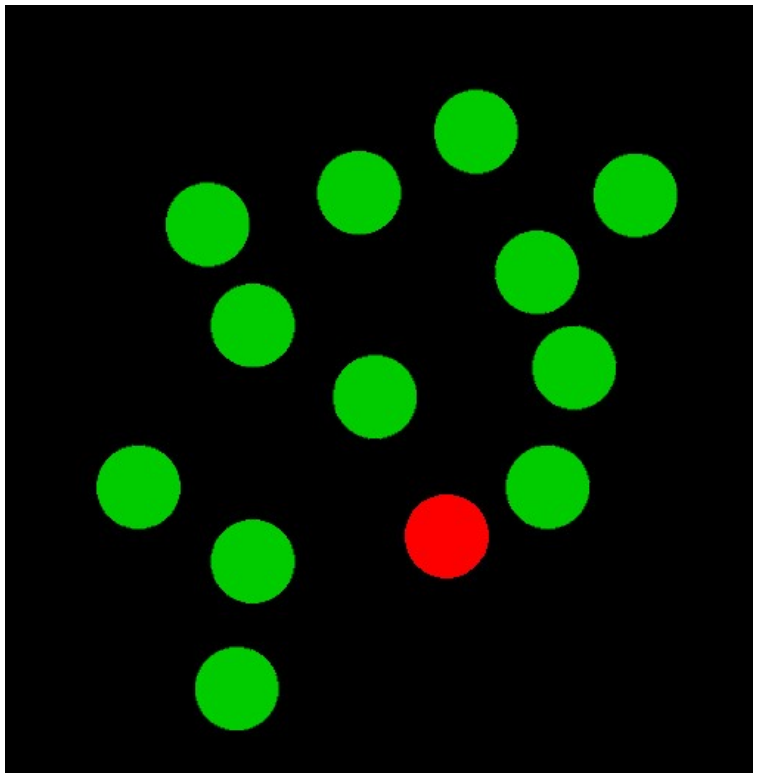
Evidence for FIT

- Visual Search Tasks
- Illusory Conjunctions

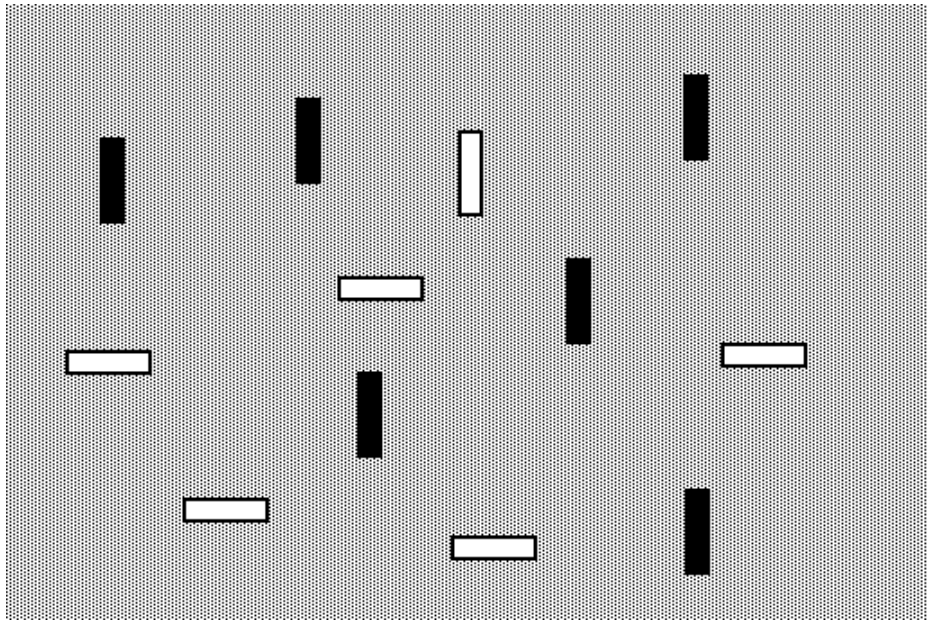
Feature Search: Find red dot



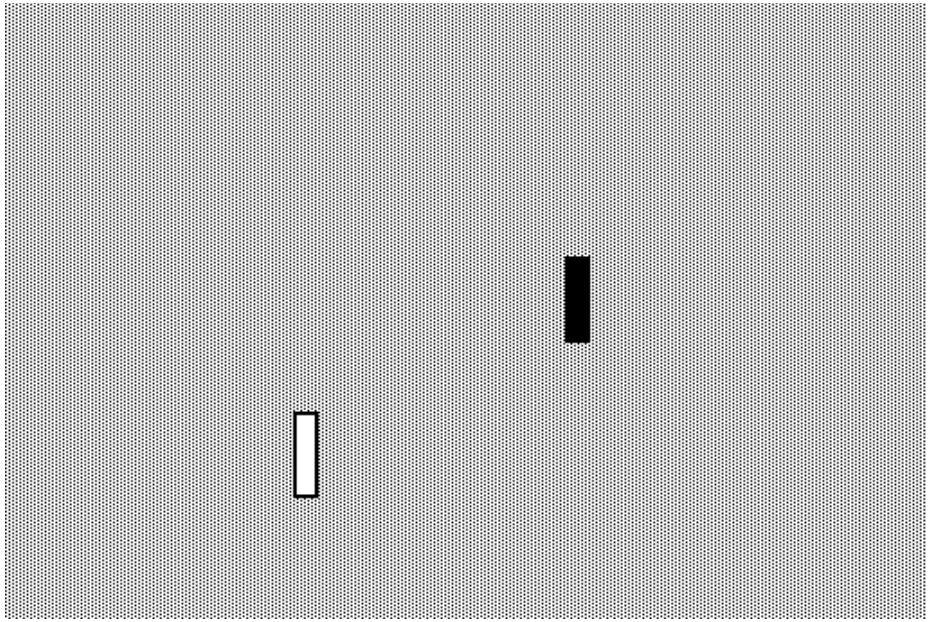
“Pop-Out Effect”



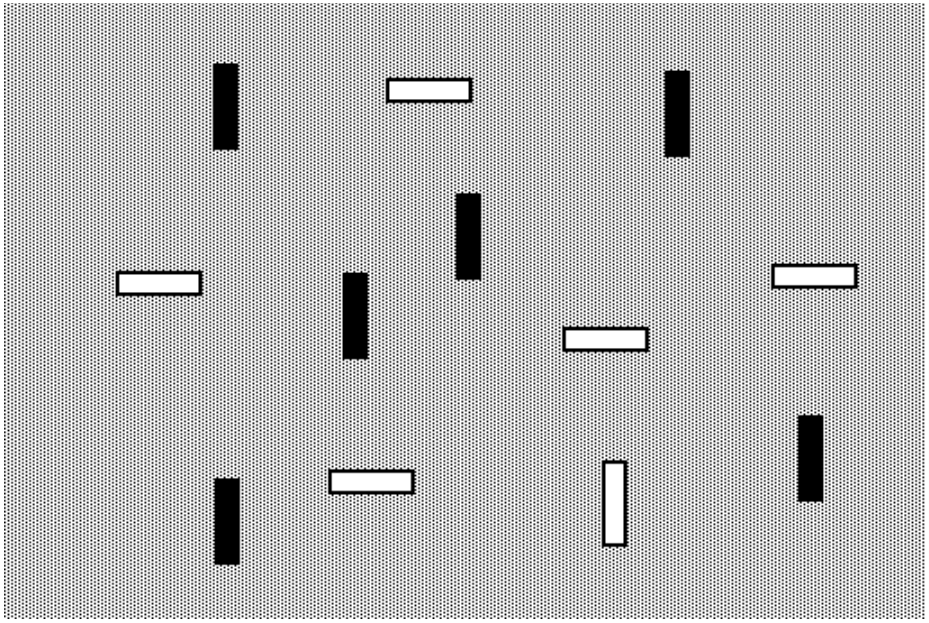
Conjunction: white vertical



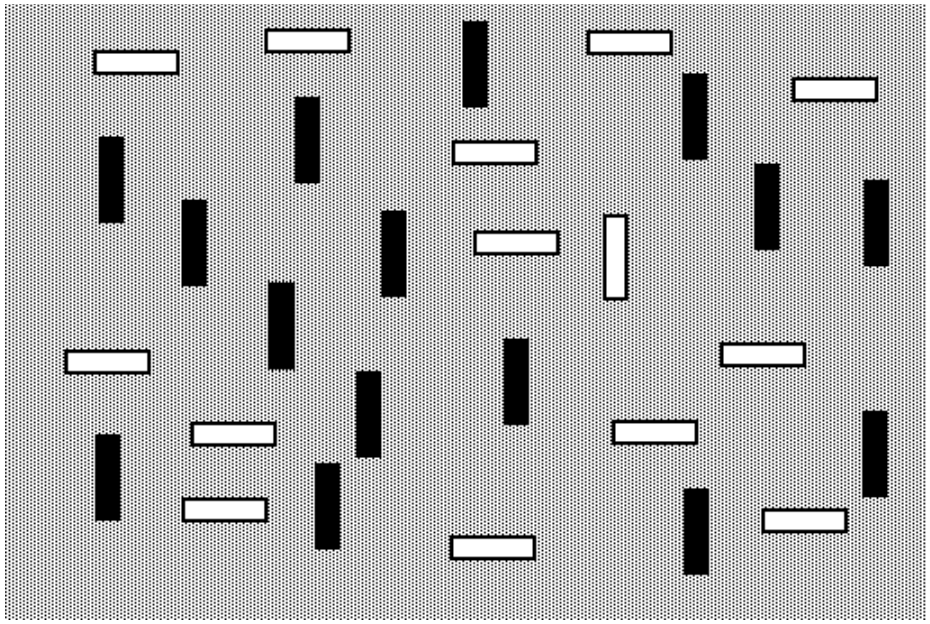
1 Distractor

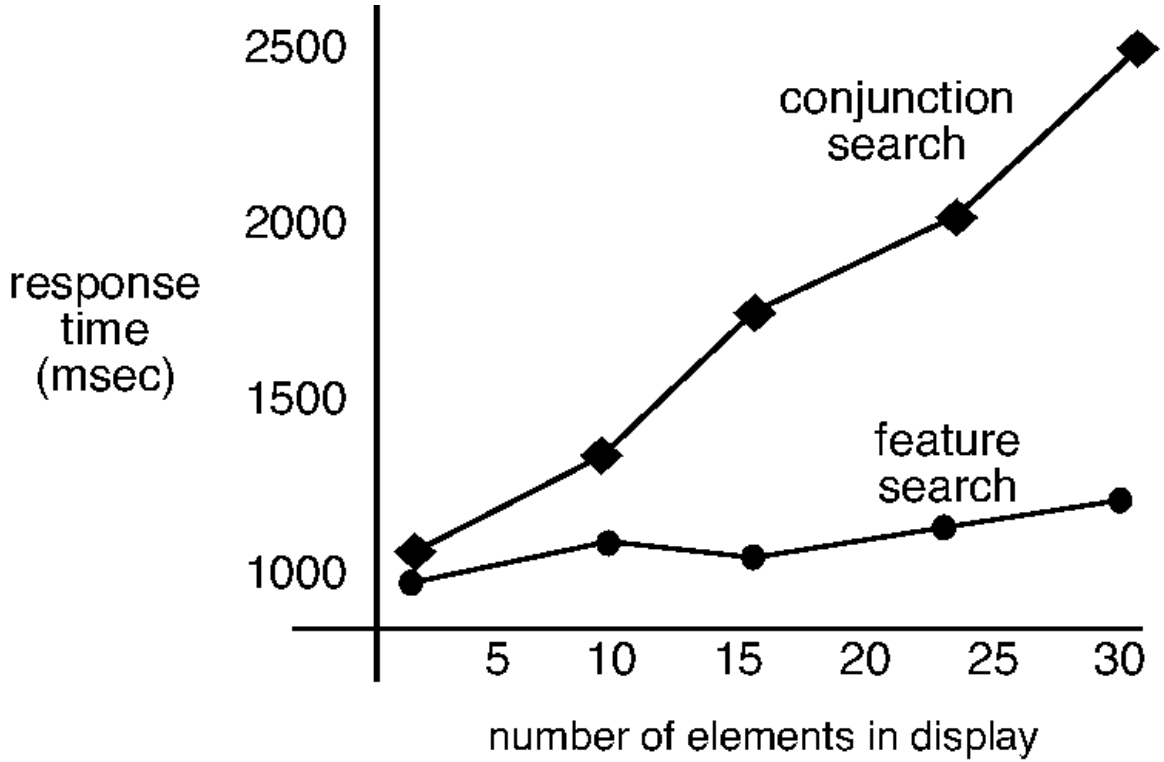


12 Distractors



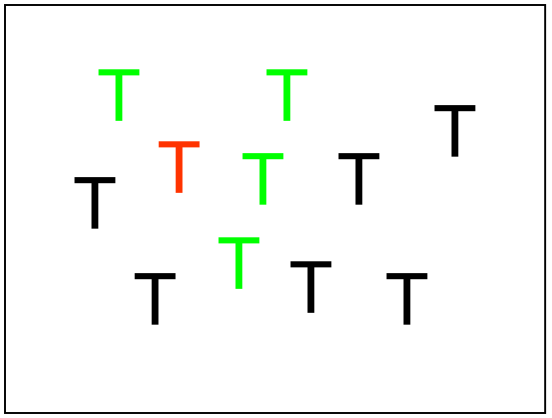
29 Distractors




2500
2000
1500
1000
response
time
(msec)
conjunction
search
feature
search
5
10
15
20
25
30
number of elements in display

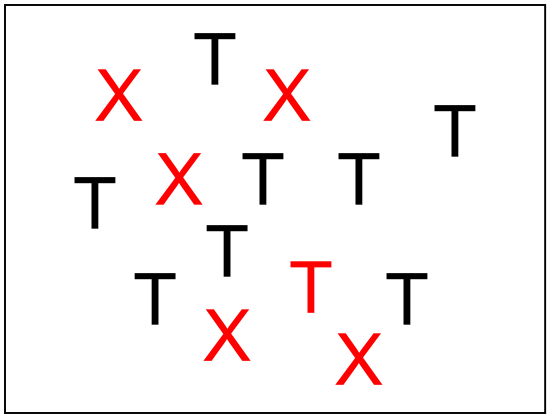
Feature Search

- Is there a red T in the display?
- Target defined by a single feature
- According to FIT target should "pop out"



Conjunction Search

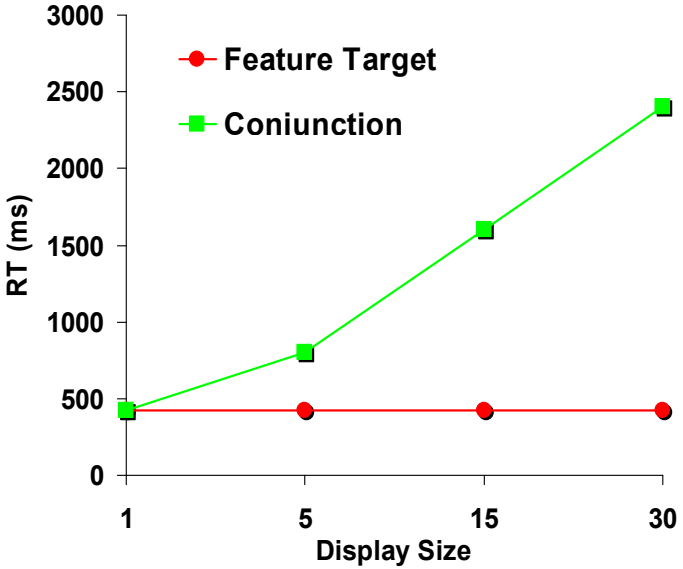
- Is there a red T in the display?
- Target defined by shape and color
- Target detection involves binding features, so demands serial search w/focal attention



Visual Search Experiments

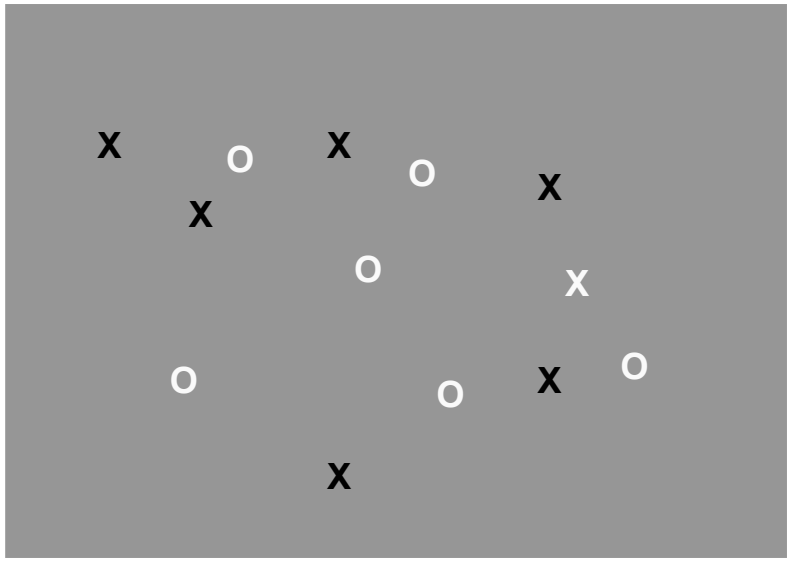
- Record time taken to determine whether target is present or absent
- Vary the number of distracters
- FIT predicts that
  - Feature search should be independent of the number of distracters
  - Conjunction search should get slower w/more distracters

Typical Findings & interpretation



- Feature targets pop out
  - flat display size function
- Conjunction targets demand serial search
  - non-zero slope

… not that simple...



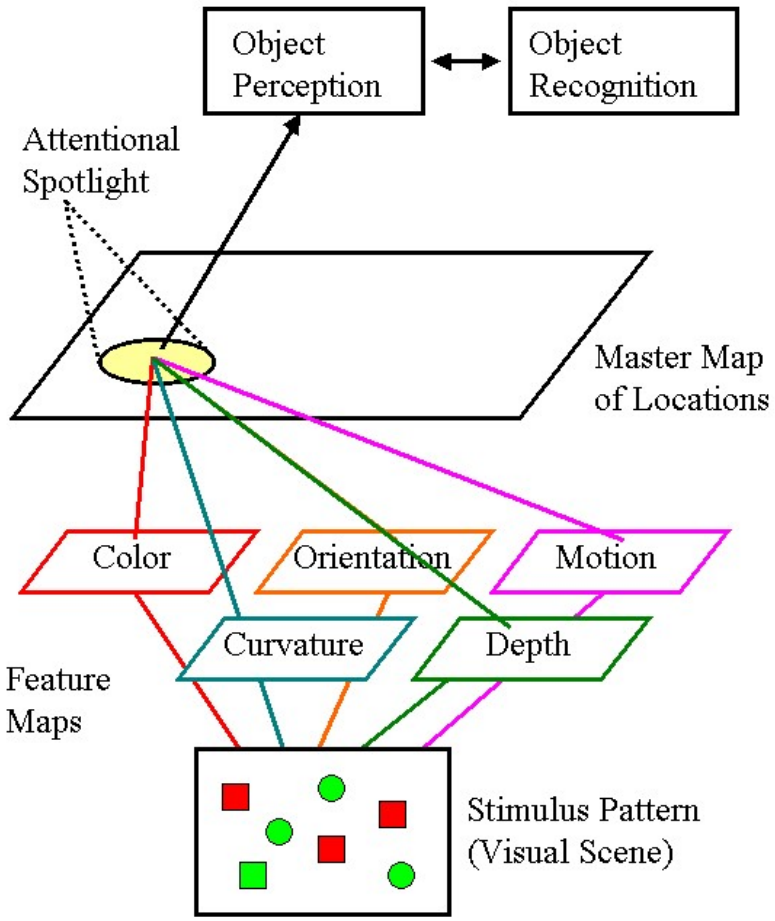
easy conjunctions - -

depth & shape, and movement & shape

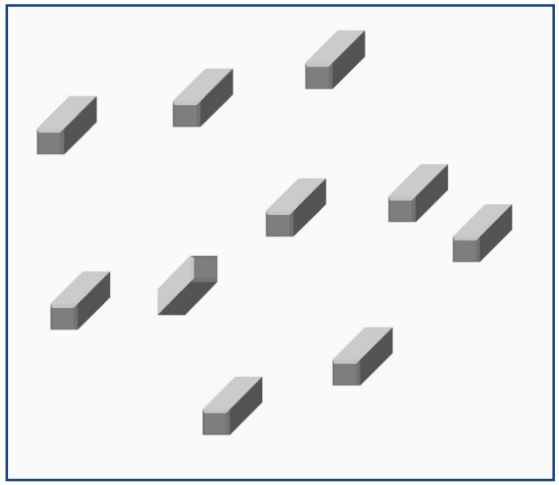
Theeuwes & Kooi

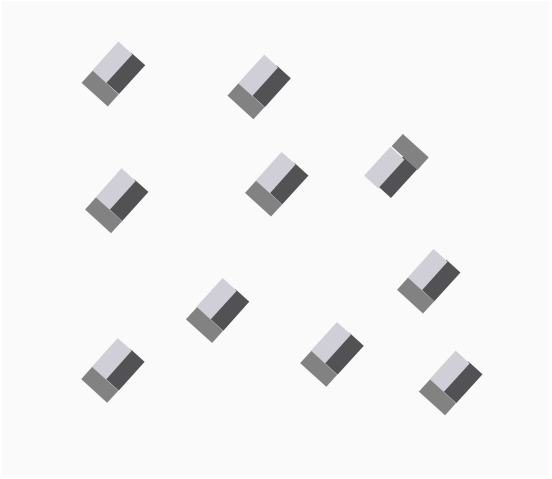
(1994)

Feature Integration Theory (Treisman)



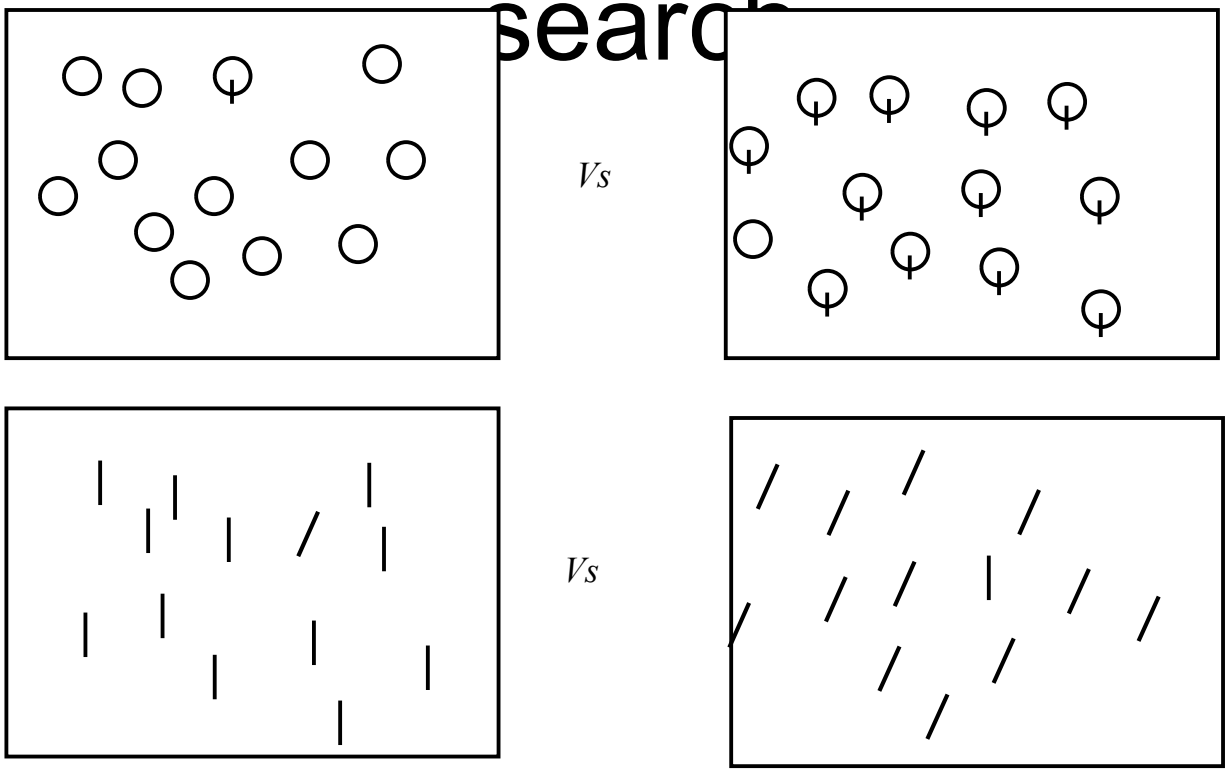
Problems for FIT accounts of vision





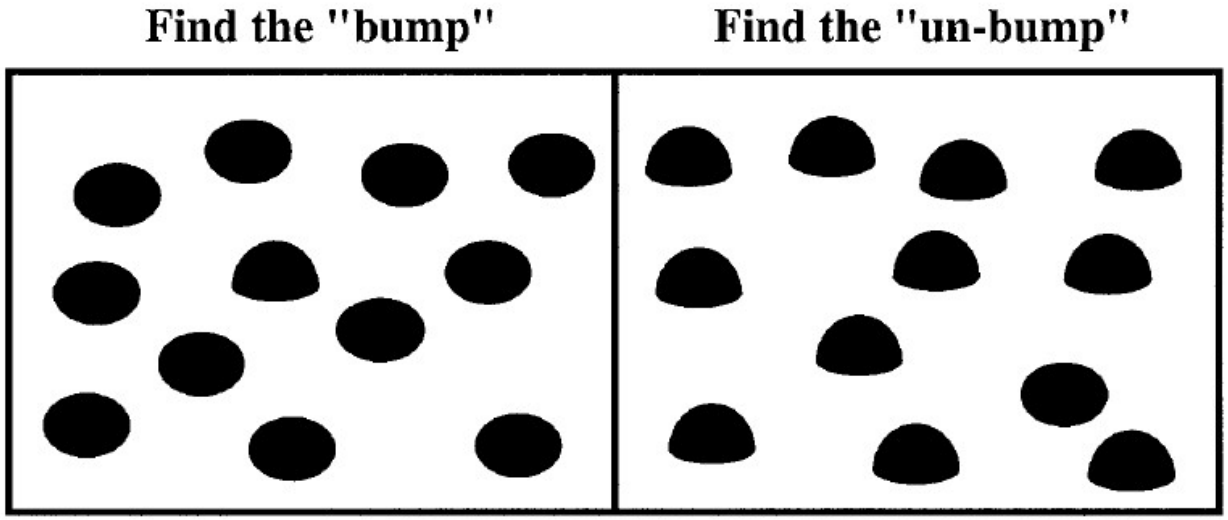
- Search is very fast in this situation only when the objects look 3D - can the direction a whole object points be a “feature”?

Asymmetries in visual search



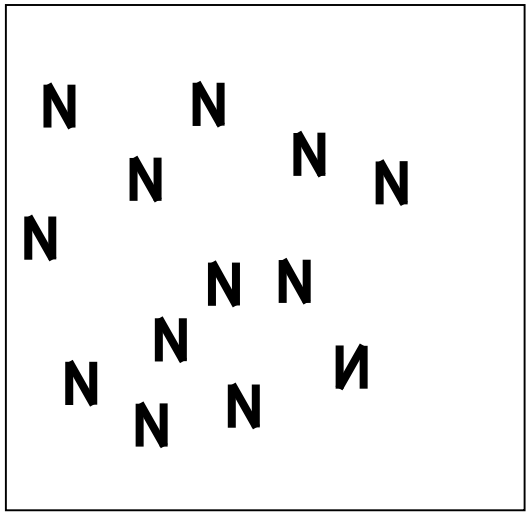
- the presence of a “feature” is easier to find than the absence of a feature

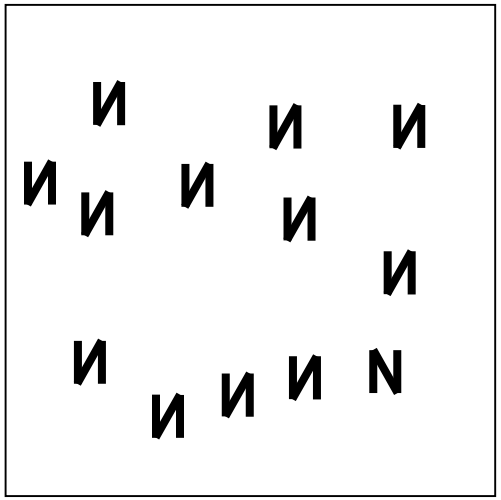
Kristjansson & Tse (2001)



- Faster detection of presence than absence - but what is the “feature”?

Familiarity and asymmetry

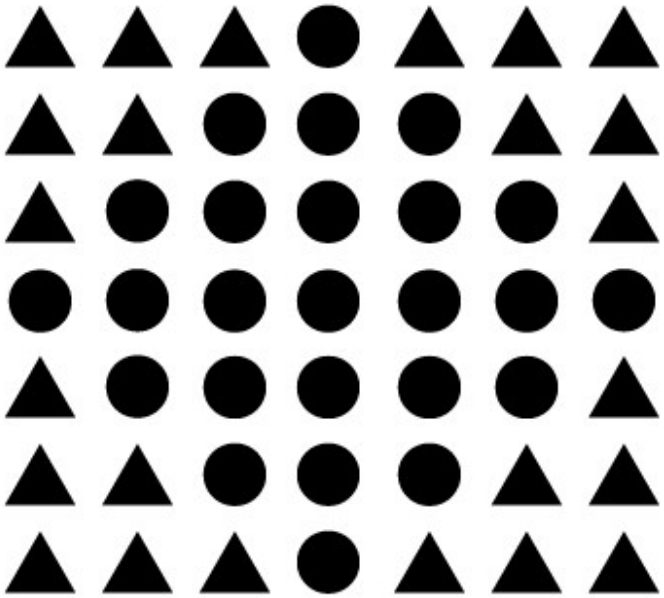




asymmetry for German but not Cyrillic readers

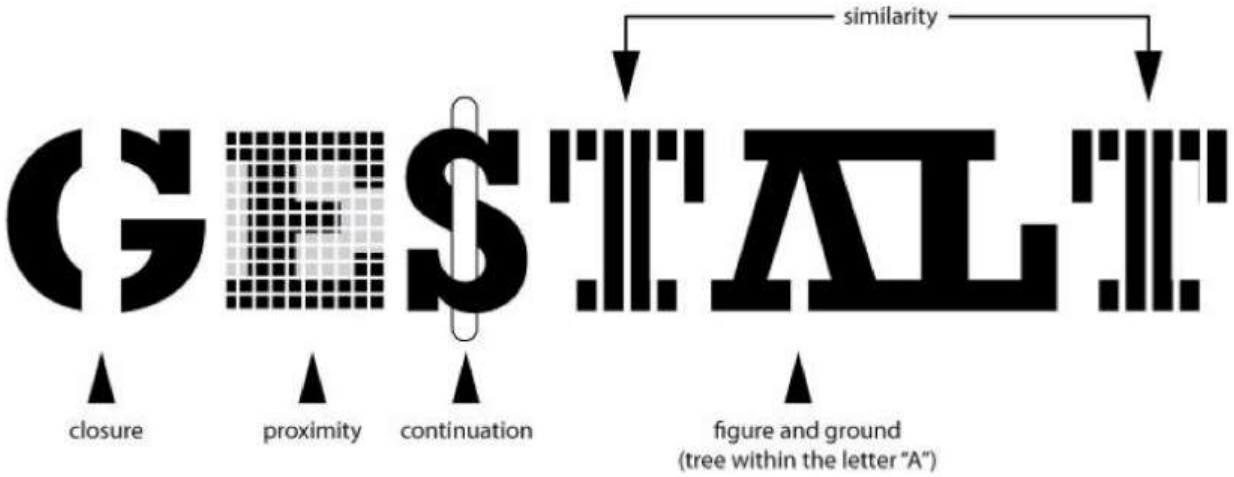
Gestalt effects

Pragnanz



Perception is not just bottom up integration of features. There is more to a whole image than the sum of its parts.

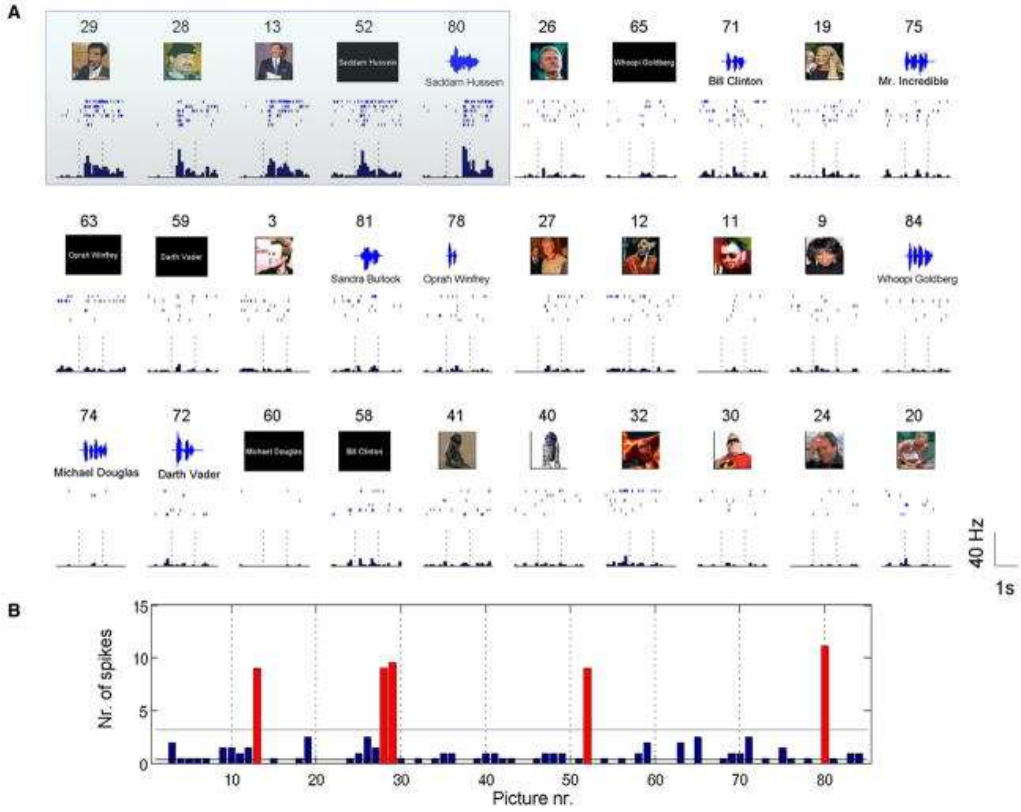
Not understood computationally



Principles are conceptually clear; DCNNs can learn them, but translation is missing

https://arxiv.org/pdf/1709.06126.pdf

Concept-selective neurons



https://www.newscientist.com/article/dn7567-why-your-brain-has-a-jennifer-aniston-cell/

Summary

- Classic accounts of visual perception have focused on bottom-up integration of features
  - Consistent with data from visual search experiments
  - Inconsistent with phenomenological experience in naturalistic settings
- Top down influences affect visual perception
  - But how?
  - We will see one promising modeling strategy today