Object Recognition: the Case for 2D Multiple Views

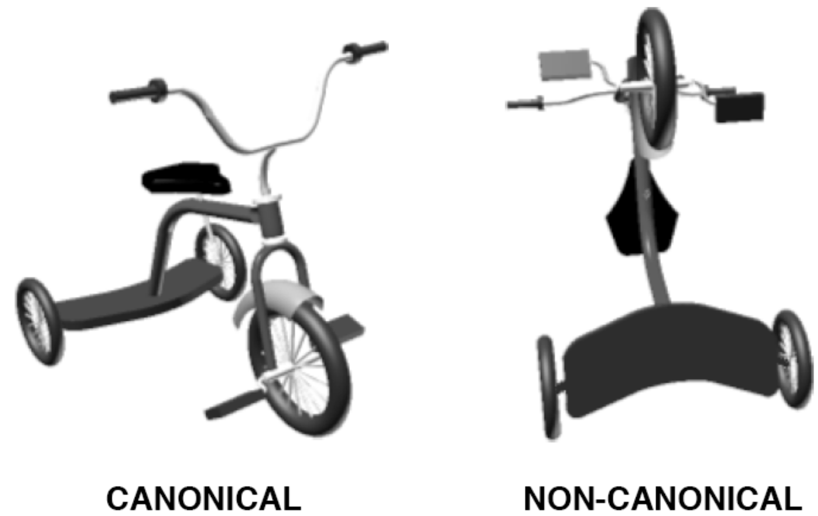
C. Andrew Burlingame

COS594B: Vision

February 25, 2008

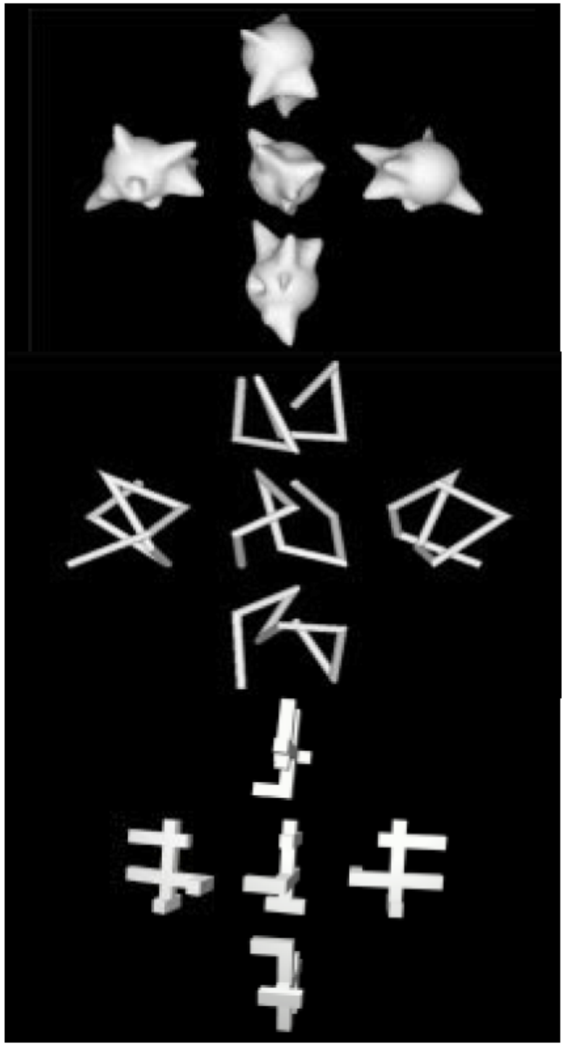
Elegant Geons Don't Fit the Data

- Geon model predicts no systematic affect of viewpoint.
- Reaction time (RT) and error rates (ER) both affected by viewpoint for certain object types.

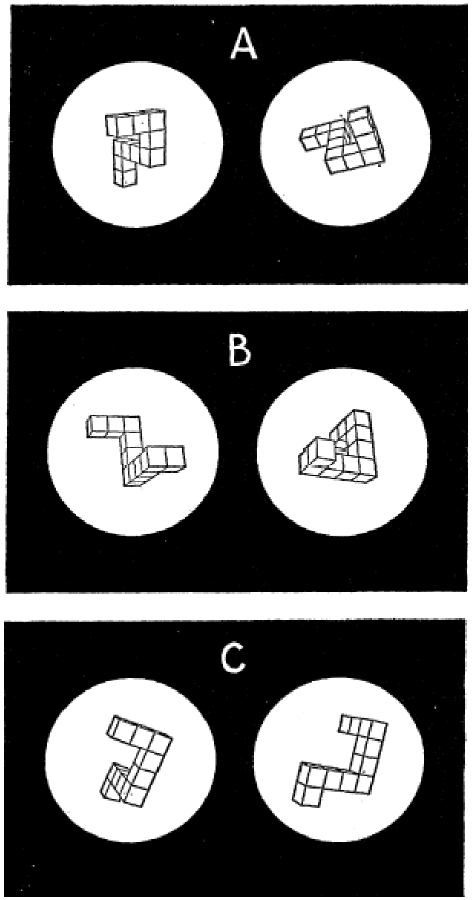


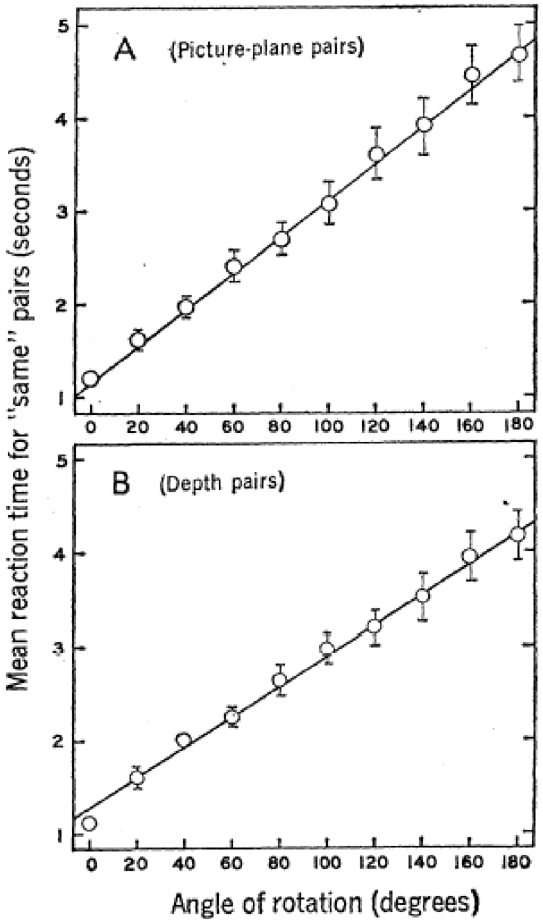
Inconsistent Results

- Little effect of viewpoint for very familiar objects
- Definite effect for novel objects
  - Effect decays with familiarity
- Explanation:
  - Handedness or “top/bottom” must be determined
  - View-invariant model built over time
  - Multiple-view model elaborated with time



Rotation for Handedness



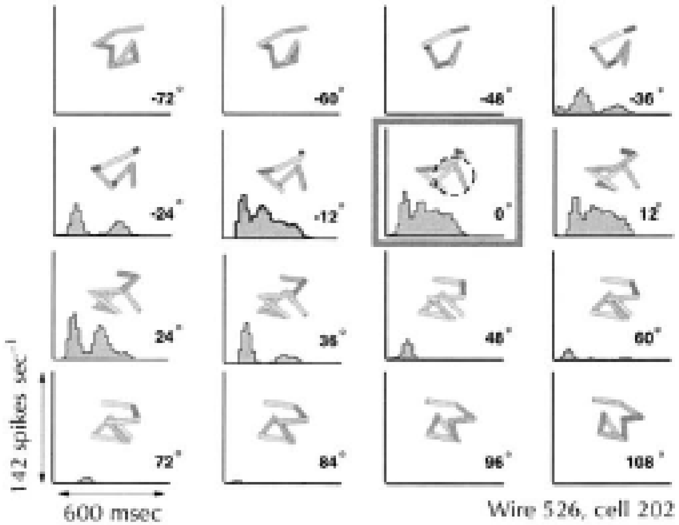


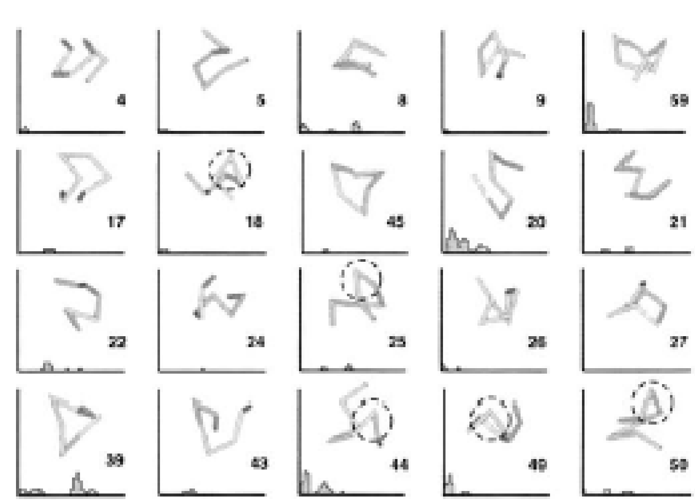
Shepard and Metzler, 1971

- Object must be rotated to "upright" to compare right and left.
- Normalization only necessary in handedness determination.
- Non-ethological studies.
- Surreptitious check for handedness.

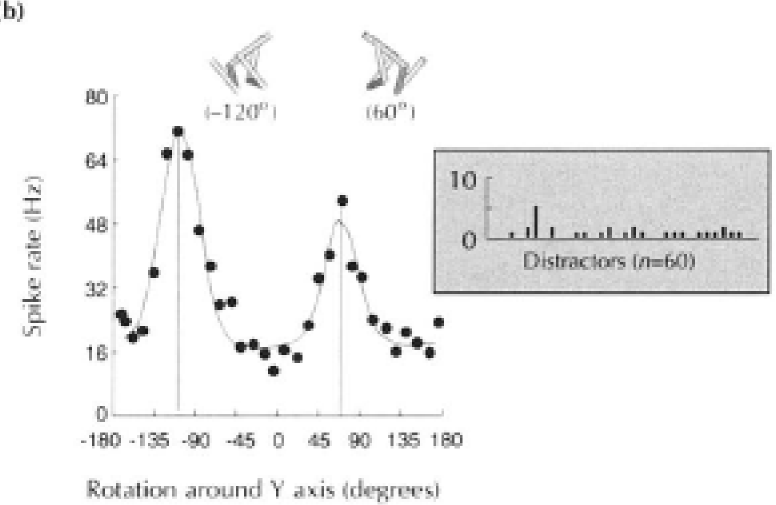
- Handedness determination established to involve angle-dependent normalization: "mental rotation."

Electrophysiological Aside





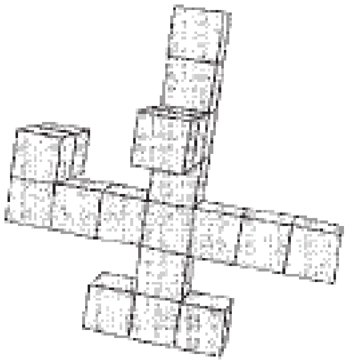
(b)



Tarr's Response

Tarr, 1995

- Goals:
  - Explicitly eliminate handedness from study.
  - Establish same normalization procedure used for handedness determination and object identification
- Problems:
  - Do these objects have geons?



- Clearly defined base
- Subjects built and named objects
  - In both versions
  - Doing their best to allow for 3D model development

Tarr's Response

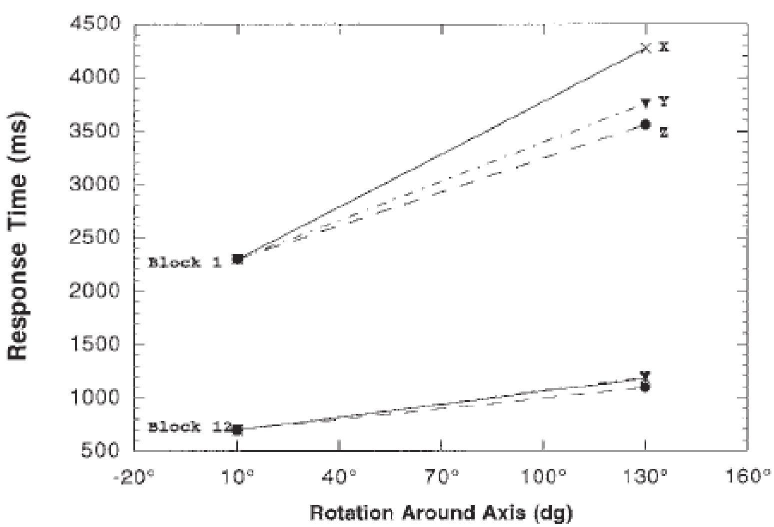
Tarr, 1995

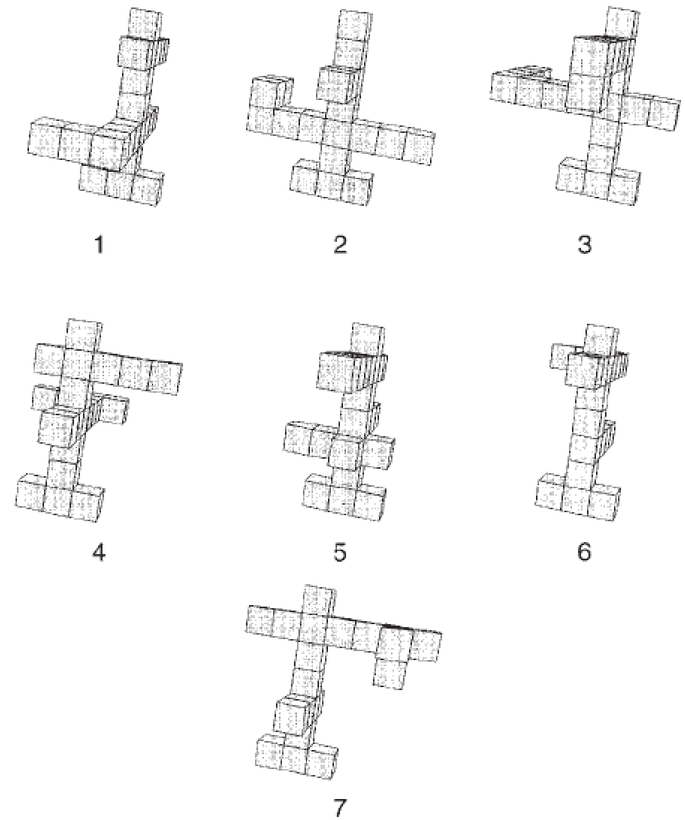
- General scheme of experiments
  - **Train:** Subset of test images shown on a specific orientation (10º off each axis
  - **Practice:** Subset of test images shown at an additional orientation (130º off axis)
    - Set's "multiple-view"
  - **Test:** Images (often containing distracters) shown at a variety of viewpoints

Tarr’s Baseline:

Establish Mental Rotation Effects

- Handedness Task
- Kip, Kef, Kor
- **Train:** 10º off each axis
- **Practice:** add 130º off axis

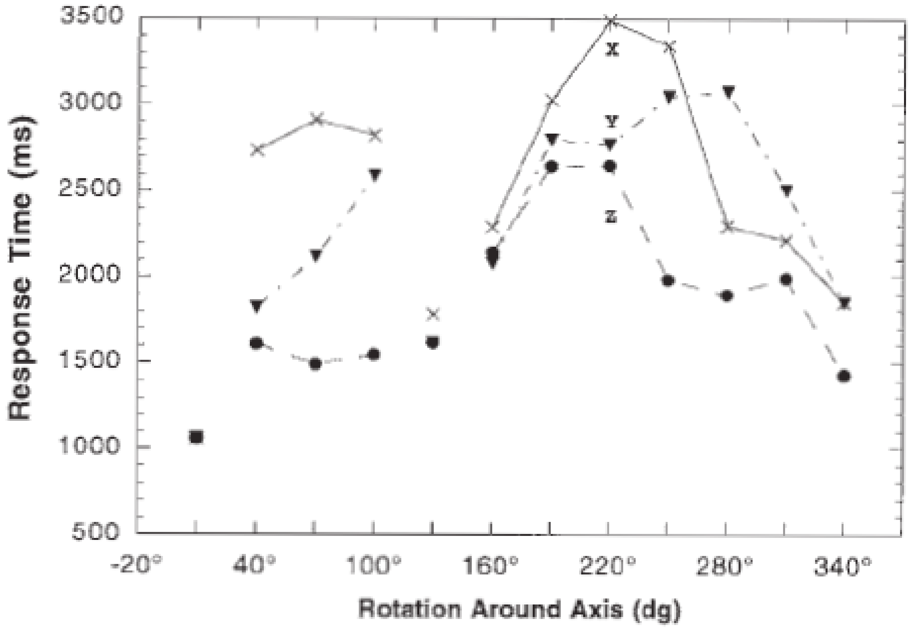




Tarr's Baseline:

Establish Correlates of Rotation

- **Test:** 11 viewpoints at 30º intervals about each axis
- **Results:**
  - May be explained by Rotation for Handedness
  - Shortest path rotation (usually)
  - Multiple View
  - Interpolation vs. Extrapolation

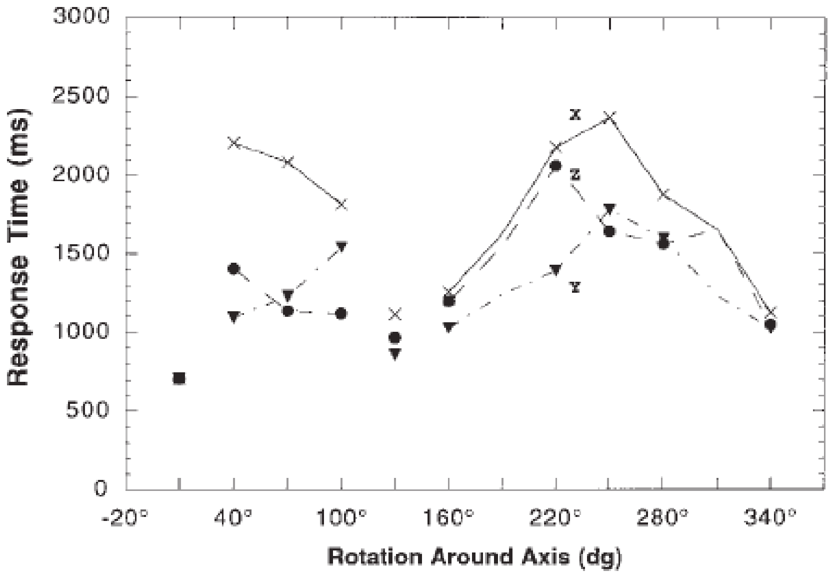


Tarr's Correlate:

Compare Rotation to Identification

- Identification Task
- Inverted objects did not appear
- Presence of Distracters

- Similarity to Exp 1 suggests same mechanism used in identification as handedness
- Failed to find any effects of visible feature set
  - Subjective evaluation of foreshortening



Tarr and Pinker, 1989

A Very Odd Result

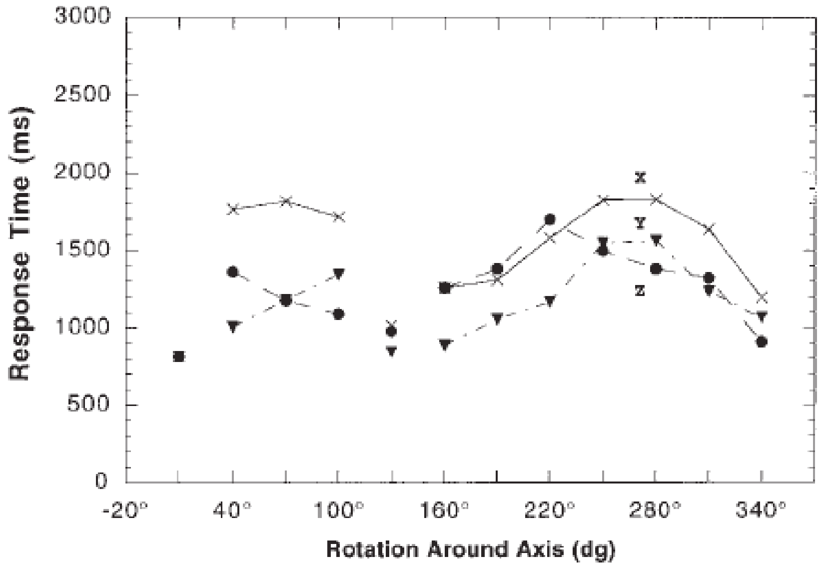
- 2D Objects
- Handedness explicitly irrelevant
  - Subjects trained on both orientations
  - Mirror pairs assigned same name
- Response time flat for all reversed images!
  - 180º rotation will always align
- With training in both orientations
  - Viewpoint variability recovered
- However, mirror image effects seen as evidence of invariant model

Tarr’s Invisible Hand:

Handedness Explicitly Removed

- Suppose models are invariant to viewpoint *and* handedness.
- Subjects may be “surreptitiously” determining handedness
- For 3D objects, rotation alignment would have to be in 4D.

- Two versions:
  - Learned both versions
  - Learned only standard version



- In both cases, images appear to have been normalized to the nearest learned orientation.
  - Even if that learned orientation was of a different handedness.

Bülthoff, Edelman, & Tarr, 1994

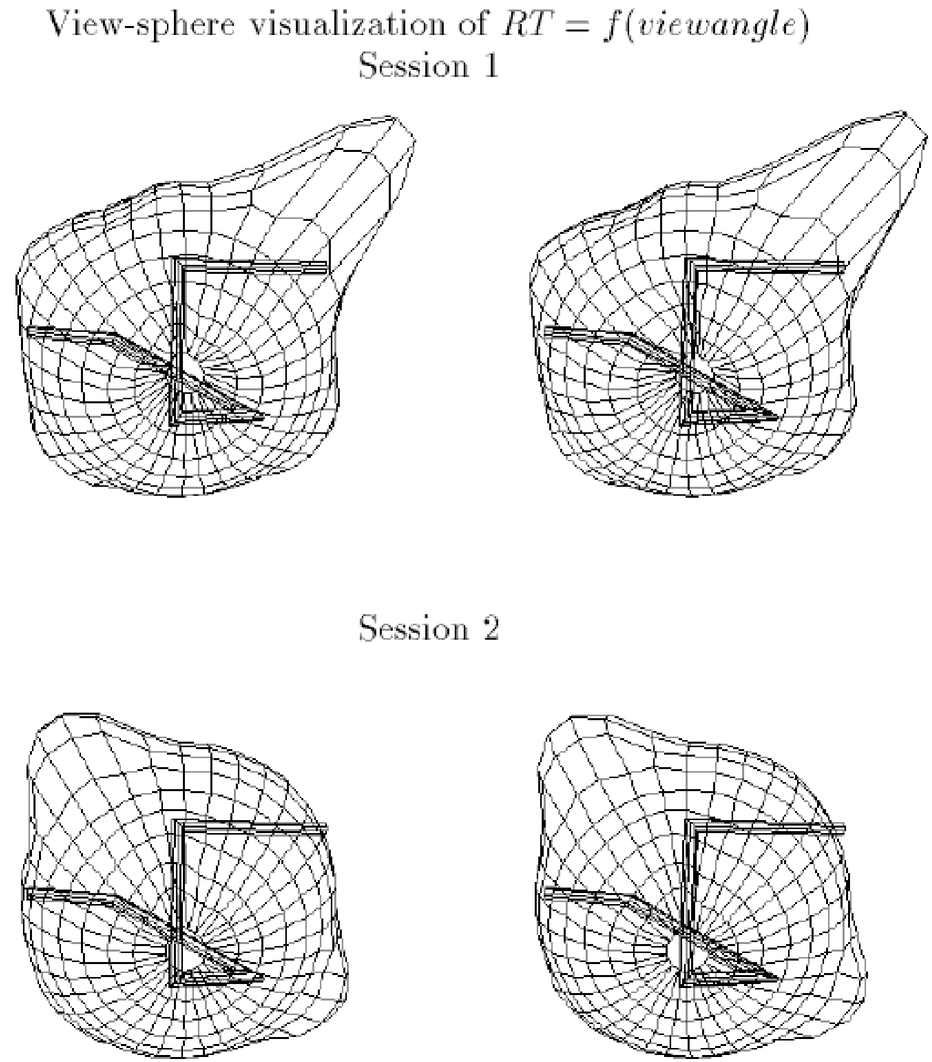
The Alpha and Omega, Now with Sprinkles!

Canonical Views

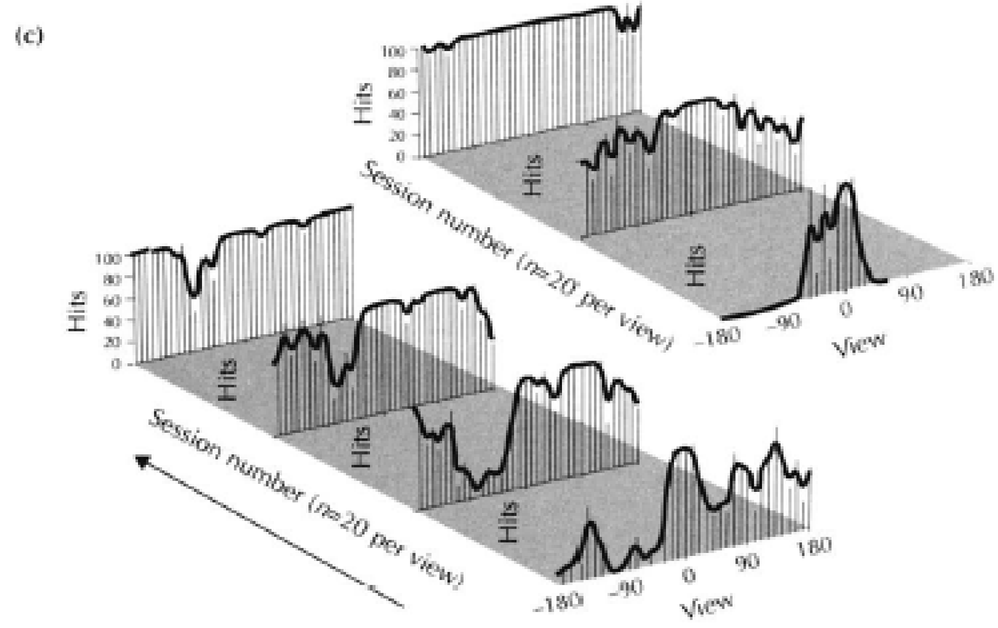
Some viewpoints are better than others.

Magnitude of this effect tends to decay with time

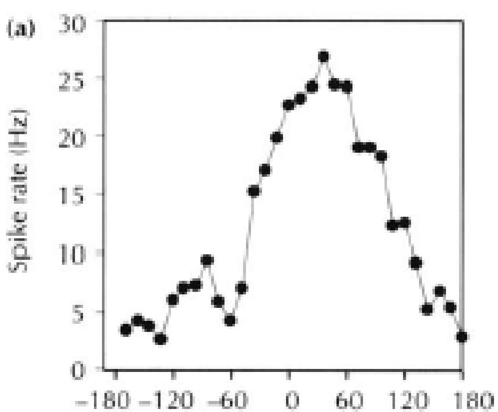
Monkeys and faces

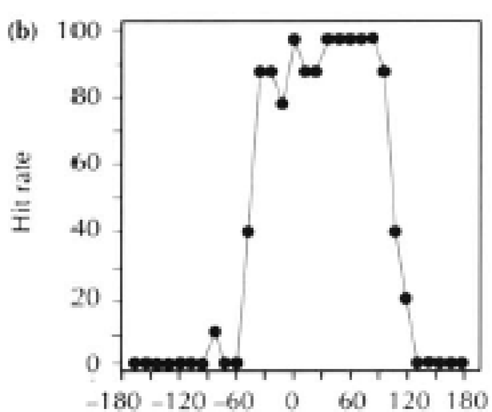


Electrophysiological Aside



Neural correlates of model development

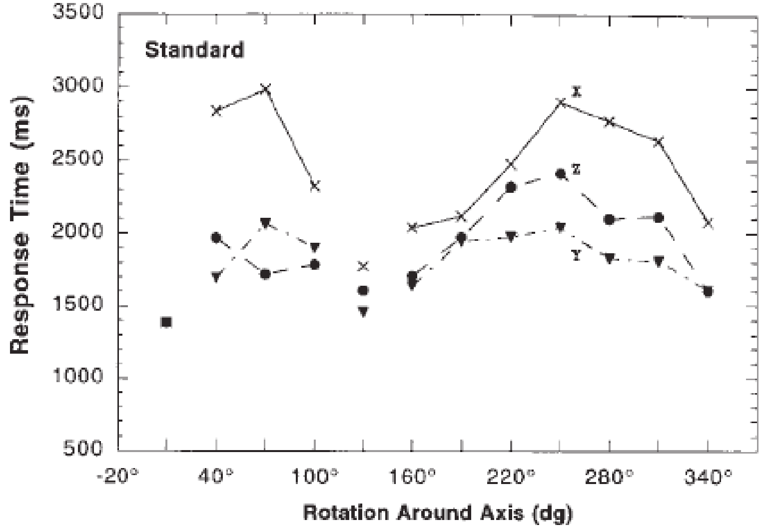


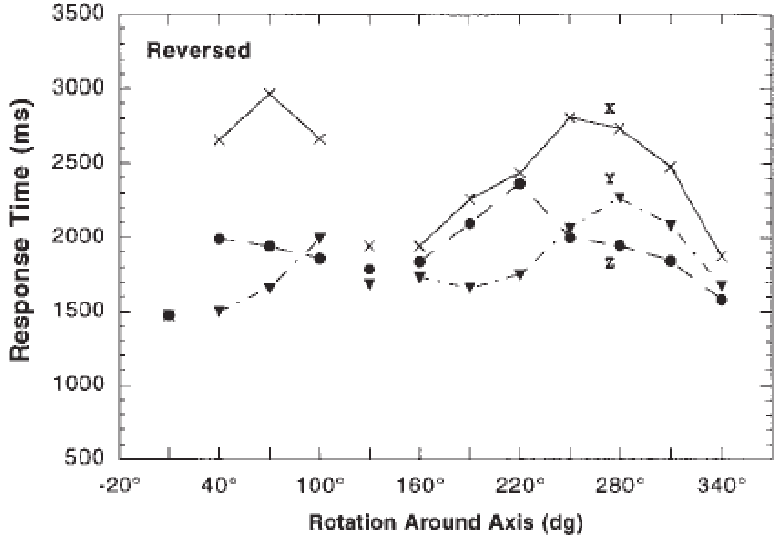


In the Familiar Limit:

Heavy Viewing

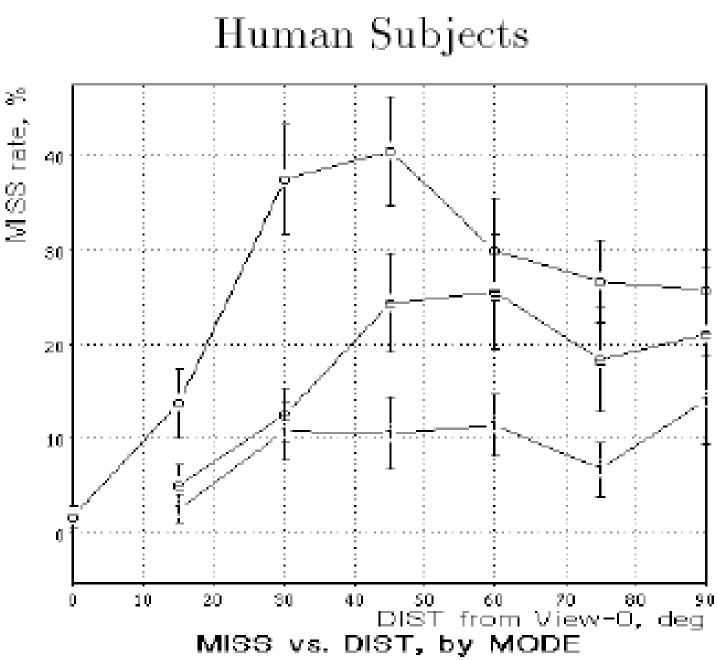
- Inverted images not shown until test phase.
- Inverted objects shown to be normalized to nearest familiar orientation.
- Evidence of handedness-invariant multiple-view model?

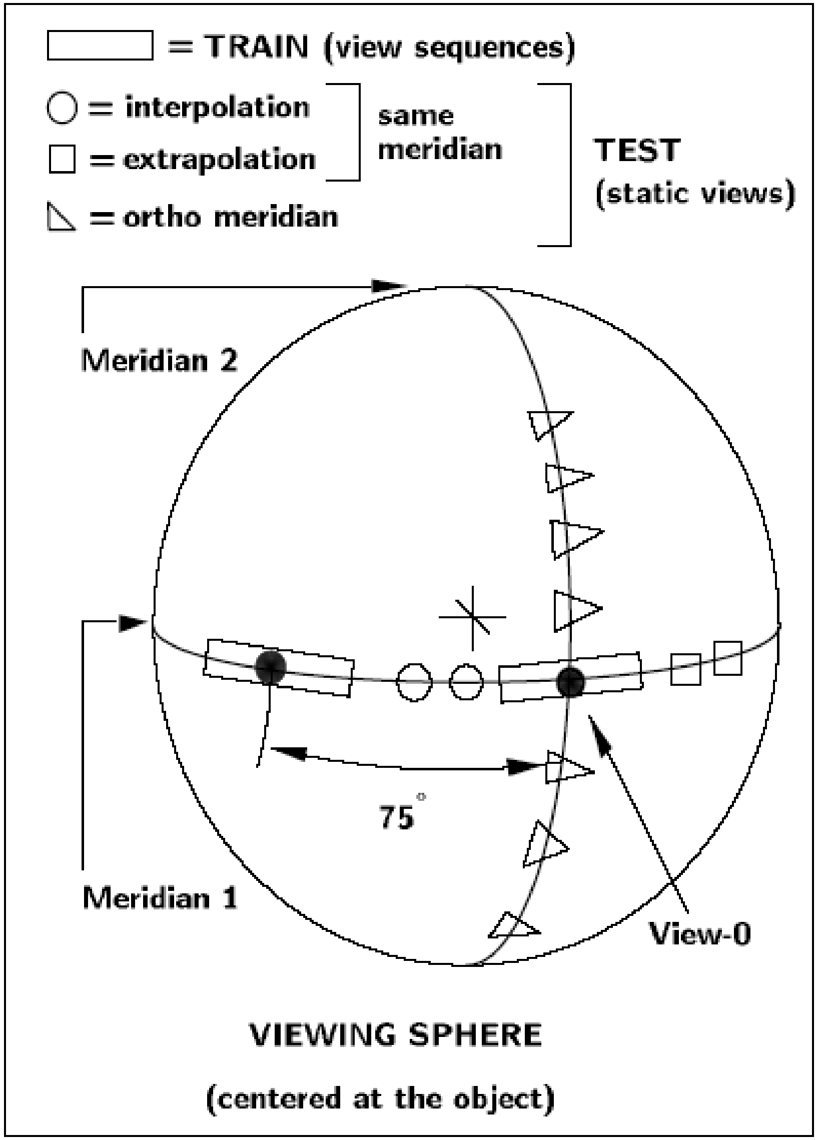




Object Models with Practice

- Given identification of familiar objects seems viewpoint independent
  - Does this imply development of an independent model?
  - Let's practice





Comparison with Models

- Clearly some form of normalization is not only extant but systematic.
- Is psychophysical data consistent with any particular normalization model?

- Ullman’s Method of Alignment: (Ullman, 1989)
  - A small number of orientation features used to align an object
  - Projection to 2D and comparison.
  - Expected Results:
    - Variable reaction time
    - Constant error rate

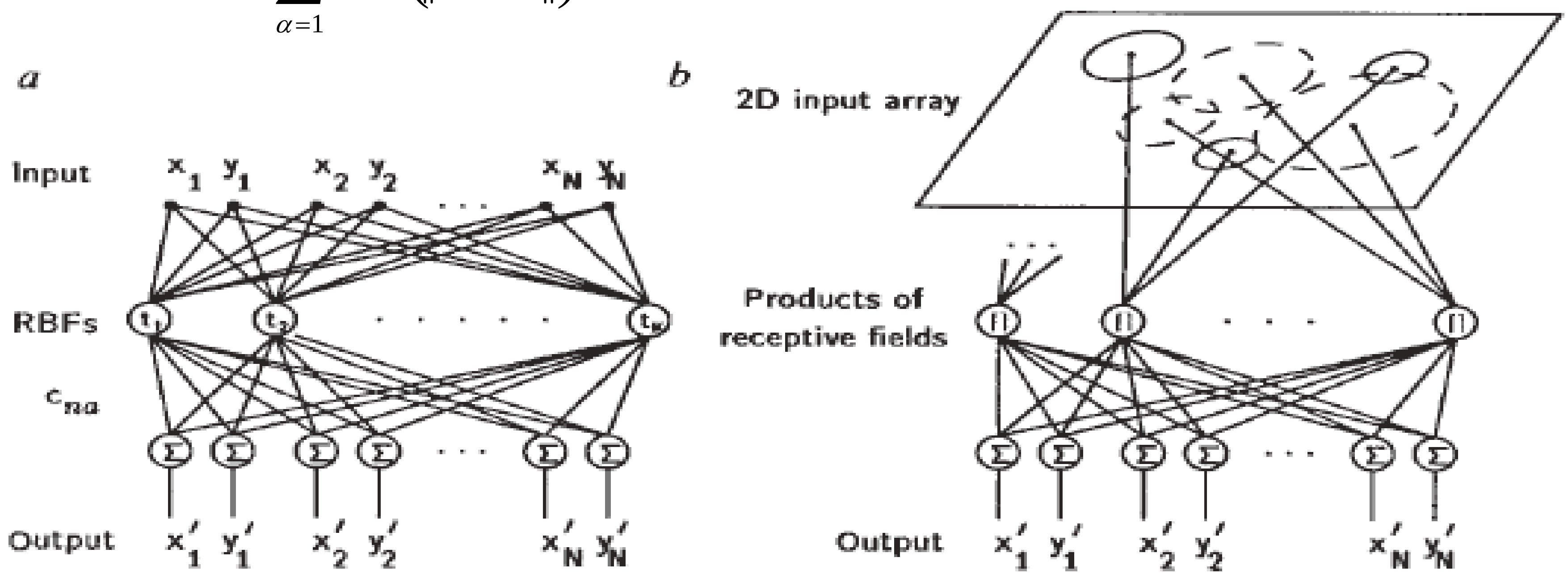
Comparison with Models

- Linear Combination of Views
  - Ullman and Basri, 1991
  - Any object point can be represented as a linear combination of the points of the same feature in a small number of 2D sample image representations.
  - Object is recognized if the test image lies in the subspace spanned by the “basis” views.
  - Expected results
    - Invariance in the subspace spanned by training views.

Comparison with Models

- HyperBF
  - Poggio and Edelman, 1990; Poggio and Girosi, 1990)
  - Output by threshold.
  - Most consistent with psychophysical data.
    - Somewhat complex performance variability

$$f(\vec{x}) = \sum_{\alpha=1}^{K} c_\alpha G\left(\left\|\vec{x} - \vec{t}_\alpha\right\|\right)$$

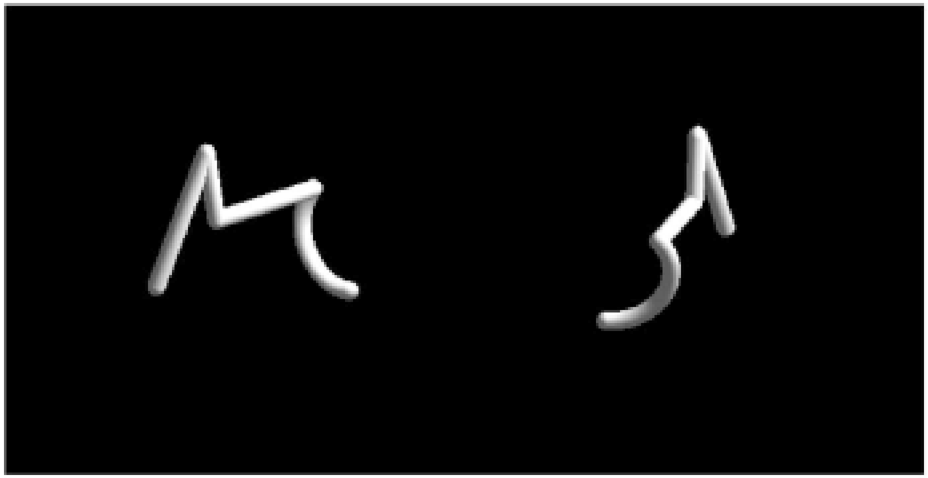


Synthesis:

Foster and Gilson, 2002

- Same/Different task
- Objects defined by normalized:
  - Number of elements
  - Length of elements
  - Curvature of elements
  - Angle of join
- "Different" pairs only varied by one attribute.
- Discriminability:

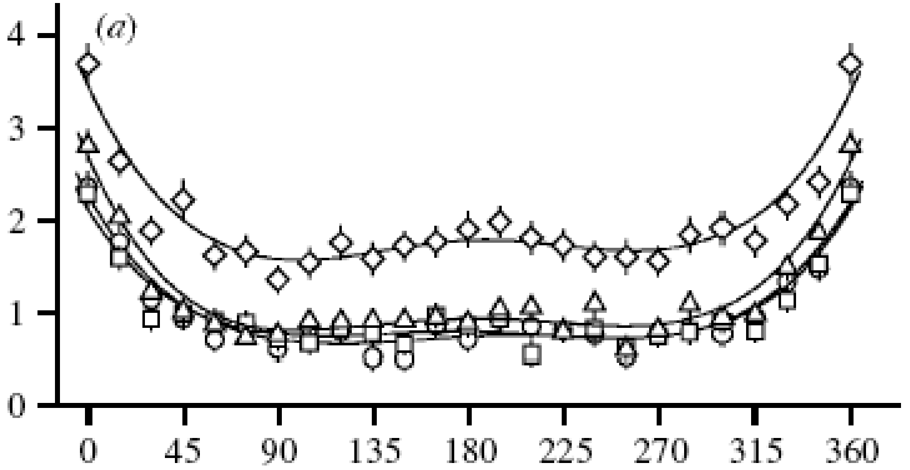
$$d' = z(HR) - z(FAR)$$



Results

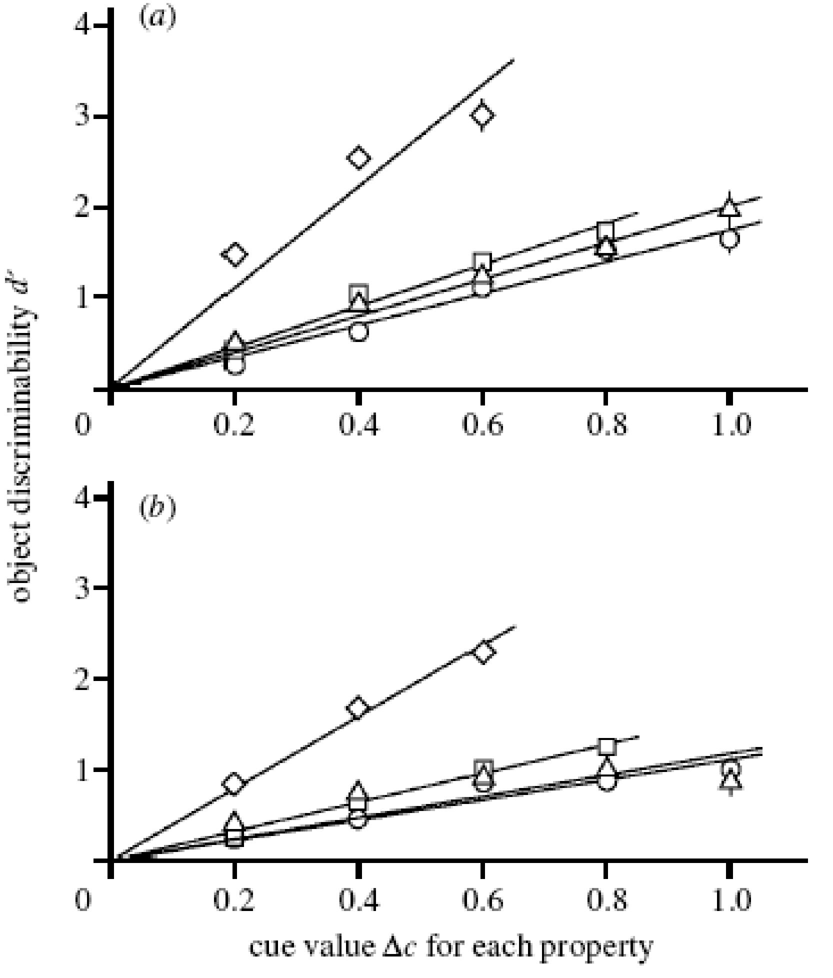
Foster and Gilson, 2002

- Linear dependence of discriminability on cue value



- Additivity of discriminability

$$d' = [k_i + f(\theta)]\Delta c$$



Summary

- In the end, both sides agree
  - A change in viewpoint will result in viewpoint costs
    - Small in some cases
  - Invariant structural properties important for generalizing across viewpoint
  - Data supporting both sides has been replicated many times
    - Can no longer argue opponent’s results are a special case
- Moving on, we try to understand how both types of analysis combine to provide robust object recognition