Digital baby: innate structures and knowledge

Shimon Ullman

Digital baby: innate structures and knowledge

Nativism – Empiricism

Innate part

Empiricists

Aristotle 384 BC
Locke 1632

Nativists

Plato 428 BC
Descartes 1596

Deep Nets

‘Digital Baby’

Innate capacities

→

197	178	172	127	165	173
218	204	202	196	193	194
215	198	186	180	187	184
218	199	203	195	191	156
167	170	134	193	106	110
95	157	160	204	168	151
192	197	203	197	187	175

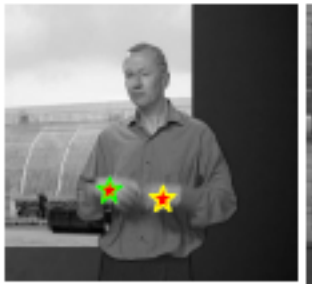
→

219	188	218	204	202	196
190	235	215	198	186	180
163	223	218	199	203	195
210	224	167	170	134	193
226	179	95	157	160	204
216	193	192	197	203	197
218	221	204	203	186	218

→

184	113	118	105	117	82
151	95	122	131	87	100
160	156	159	197	178	172
136	219	188	218	204	202
184	190	235	215	198	186
175	163	223	218	199	203
221	210	224	167	170	134

• • •



Objects
Actions
Agents
Goals
Tools
Interactions

Examples: Learning Hands, Gaze

Hands

Gaze

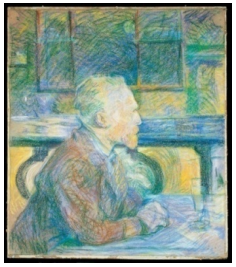
Difficult, appear early, important for subsequent learning of agents, goals, interactions,

Hands and body parts are important

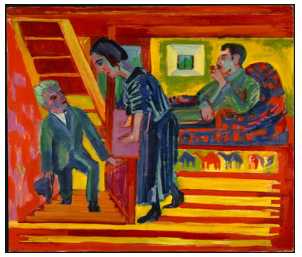
Action recognition
Gesture and communication
Agents interactions

Hands are difficult

Multiple appearances



Van Gogh



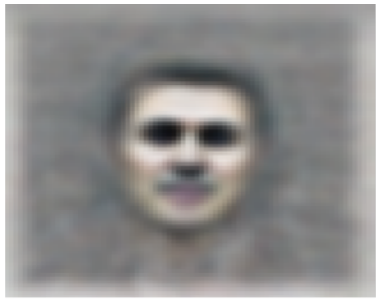
Kirchner

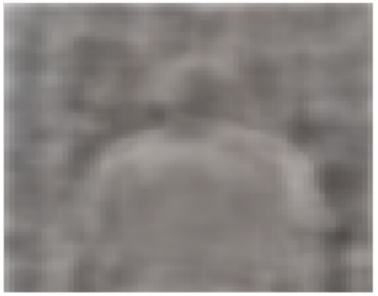
Small and inconspicuous

Models: Require extensive supervision

Unsupervised learning does not discover hands

‘Building High-level Features Using Large Scale Unsupervised Learning’
Ng et al Stanford and Google ICML 2012





1B connections, 10M YouTube images, 1000 machines, 16,000 cores, 3 days

Some statistically significant structures emerge with large data (unlike the original ‘deep nets’ goal)

In humans:
Selectivity to hands appear early in infancy

Using a Head Camera to Study Visual Experience.

'Overall…hand were in view and dynamically acting on an object in over 80% of the frames'.

Yoshida & Smith 2008

What makes hands learnable by humans?

Motion, Hand as ‘mover’

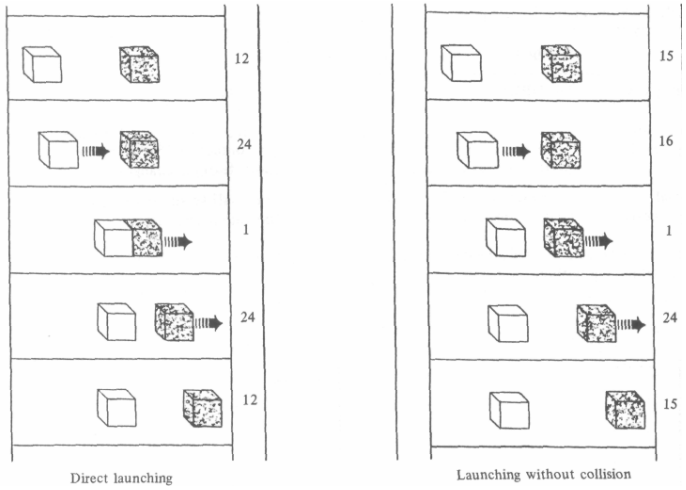
See: Saxe, Carey The perception of causality in infancy. *Acta Psychologica* 2006

Early sensitivity to special motion types

- High sensitivity to motion in general

  (detecting motion, motion segmentation, tracking)

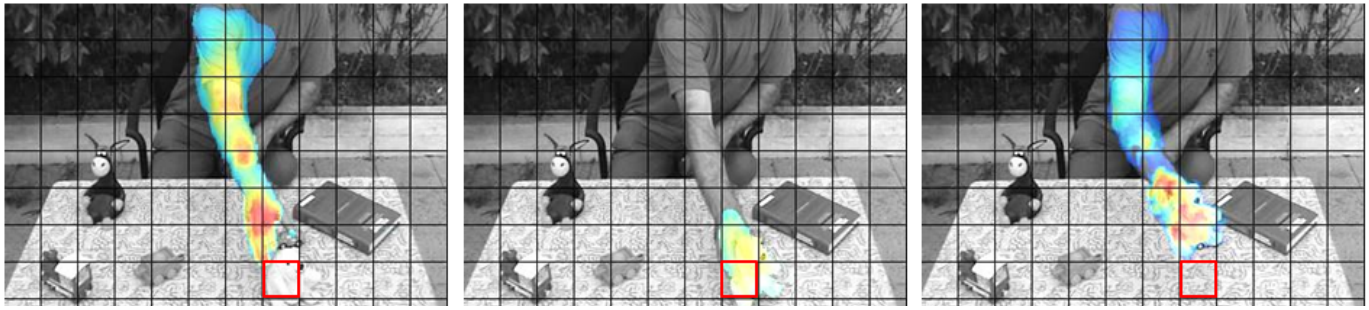
- Specific sub-classes of motion



A specific motion event is highly indicative of hands

‘Mover’ event

Detecting ‘Mover’ Events



A moving image region causing a stationary region to move or change after contact.

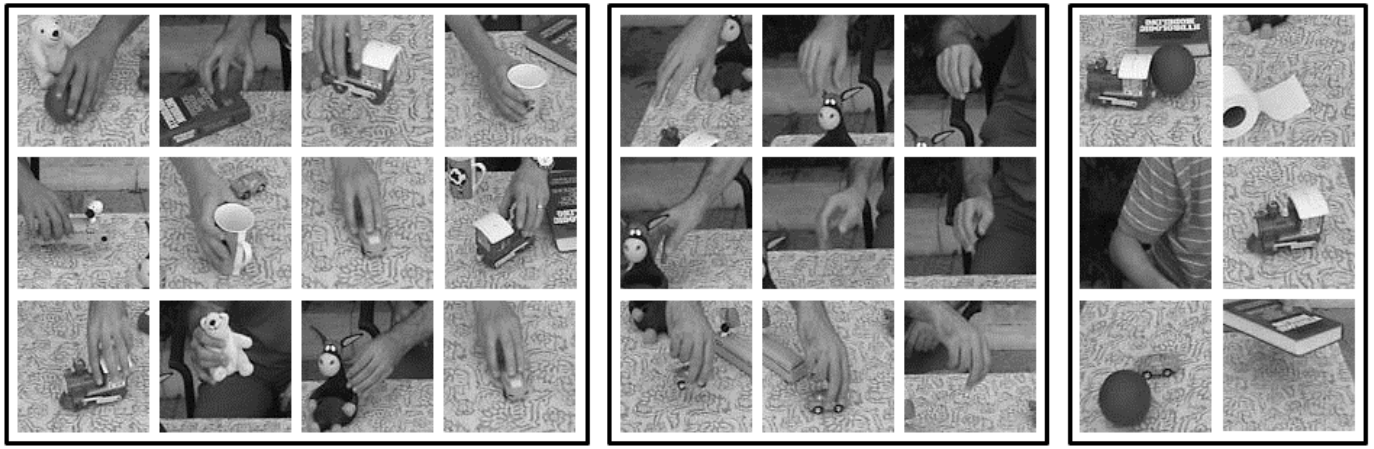
Simple and primitive, prior to objects or figure-ground segmentation

Movers detection

‘Mover’ as an innate teaching signal for hand

Motion alone is insufficient

‘Mover’ events extracted from videos



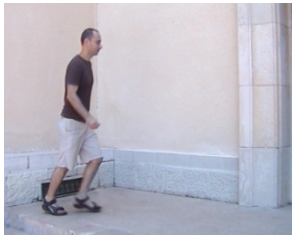
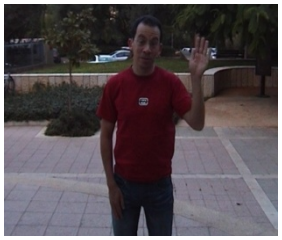
High fraction of Hand images
(90% recall 65% precision)

Internal supervision by movers and by tracking

Training Videos





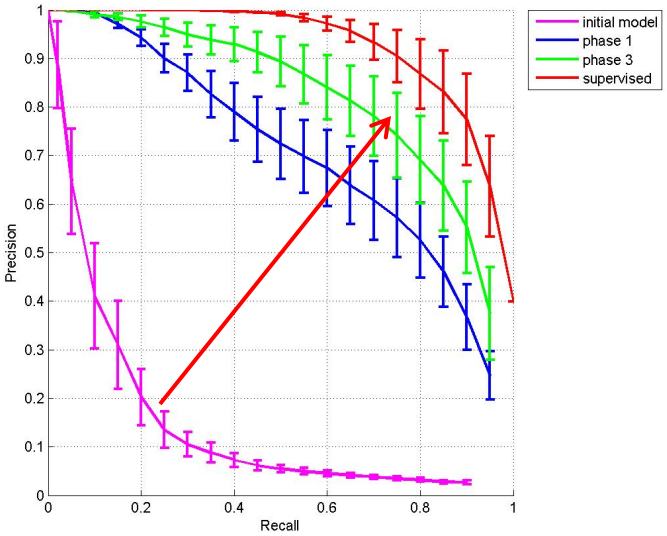

Movies of scenes, people moving, manipulating objects, moving hands.

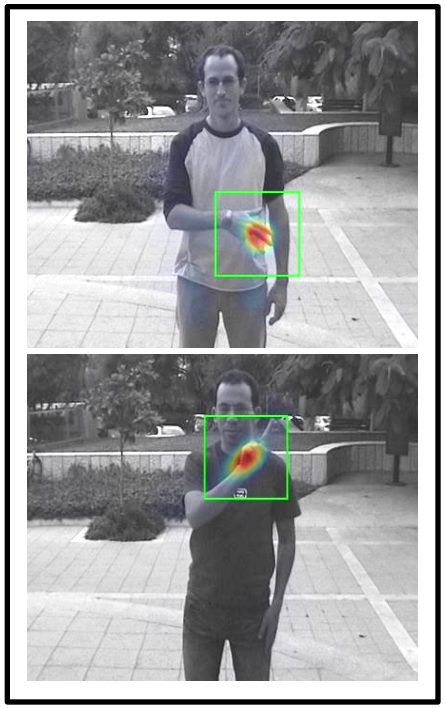
'Mover' events are detected in all movies and used for training

Hand detection in still images

Detection mainly of hands in object manipulation scenes

(a)



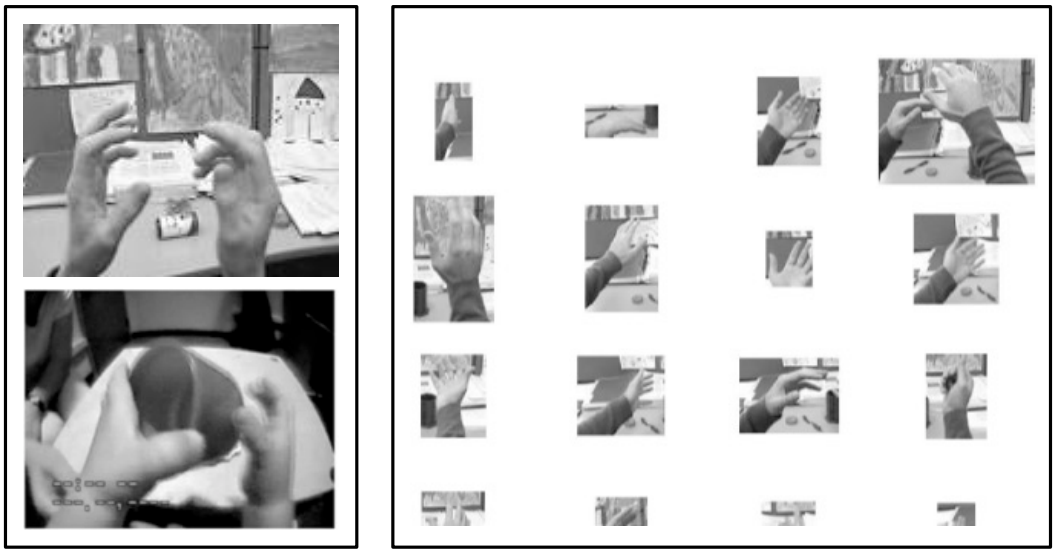


(c)

(d) Appearance

(e) Context

Own Hands



(A) (B)

Yoshida & Smith

A learned class, not the basis of hands in general

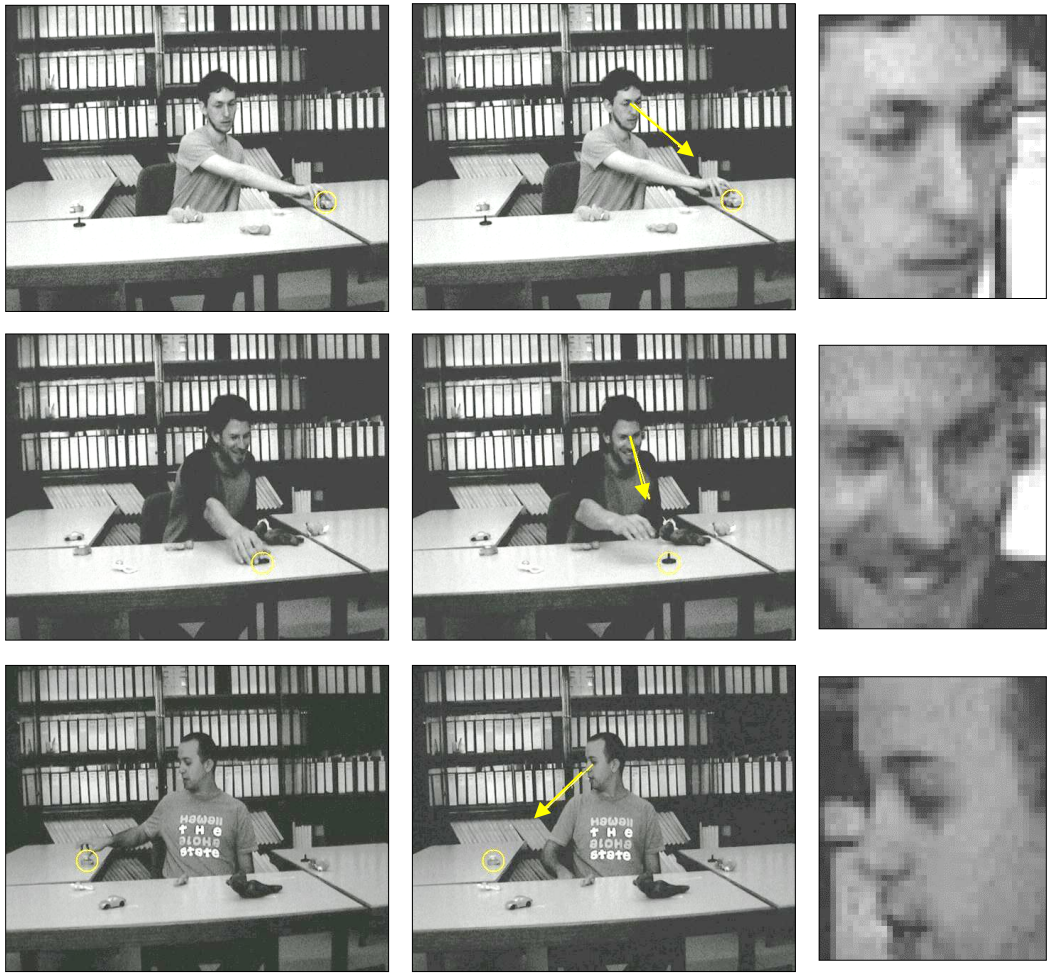
Gaze

Infants follow the gaze of others
Starting at 3-6 months and continues to develop
Head orientation first, eye cues later
Important in the development of communication and language
Modeling mainly head direction

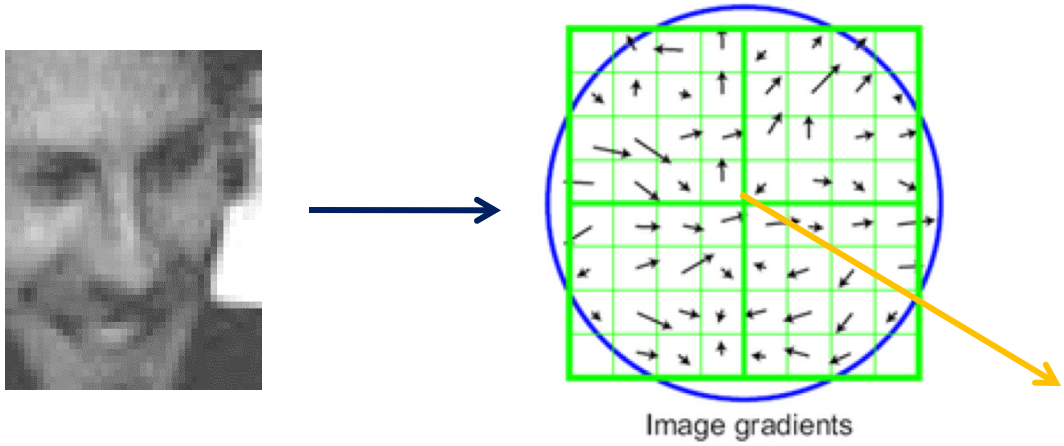
Gaze cues are subtle and inconspicuous

Mover supplies the teaching signal

Using hand ‘mover’ events to learn gaze direction

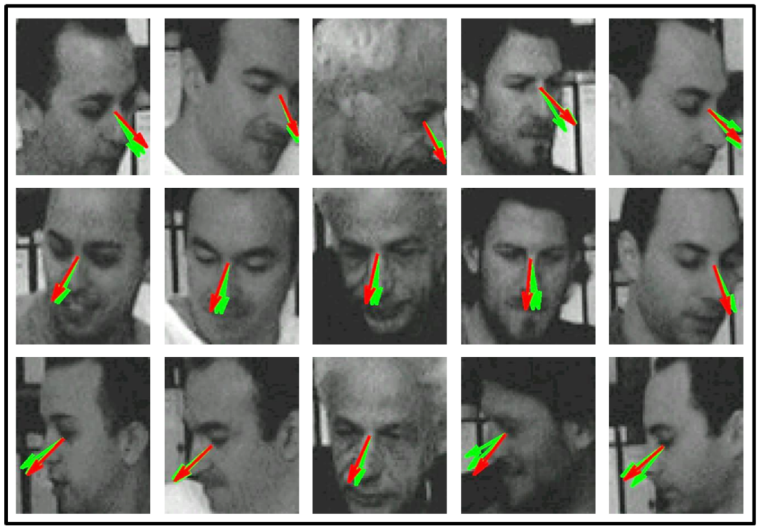


Face description



Gaze extraction 2D

(D)



(E)

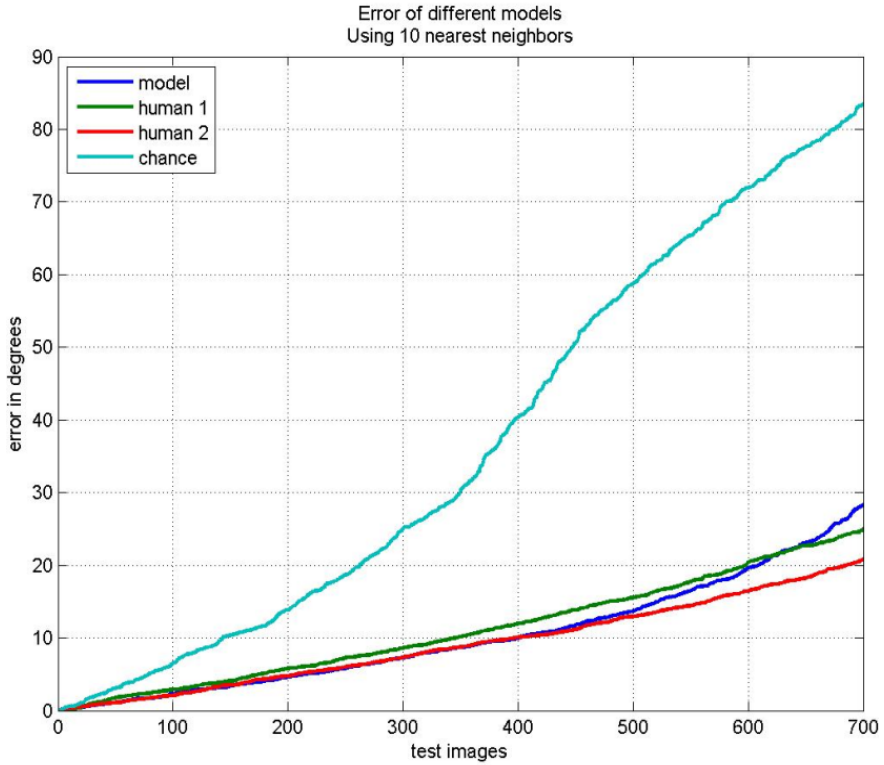
Training

Testing

Model

Humans

Gaze results, 700 test images
8 people, leave-one-out

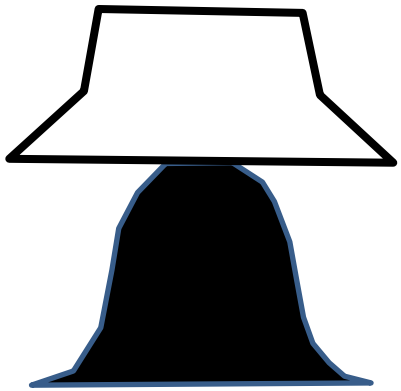


Emerging Interpretation

Both agents are manipulating objects;
The one on the left is interested in the other's object

Learning to Perceive Coherent Objects

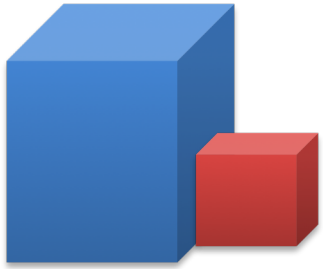
Learning to Perceive Coherent Objects



Dorfman, Harari, Ullman CogSci 2013

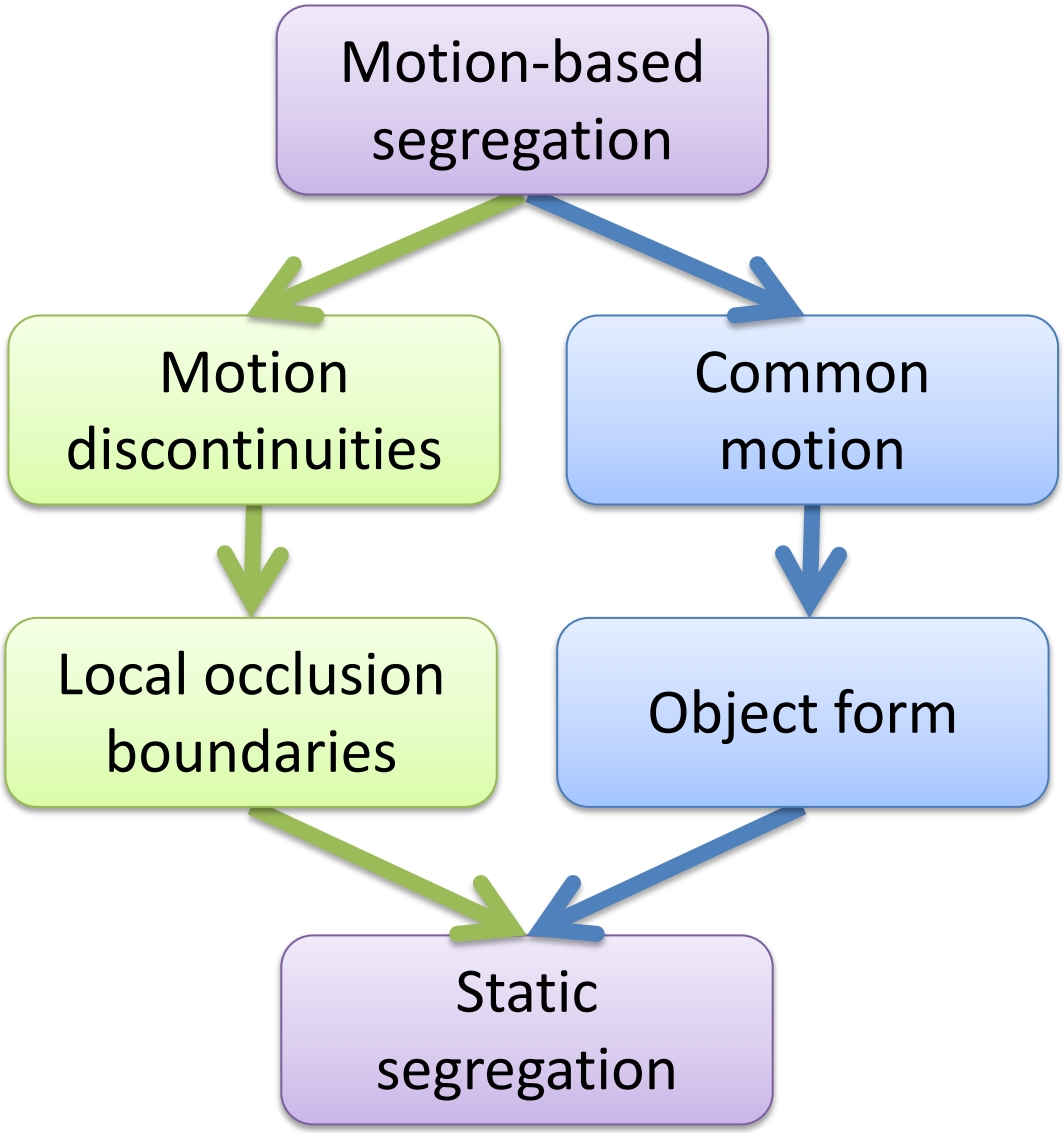
Motion provides segregation information

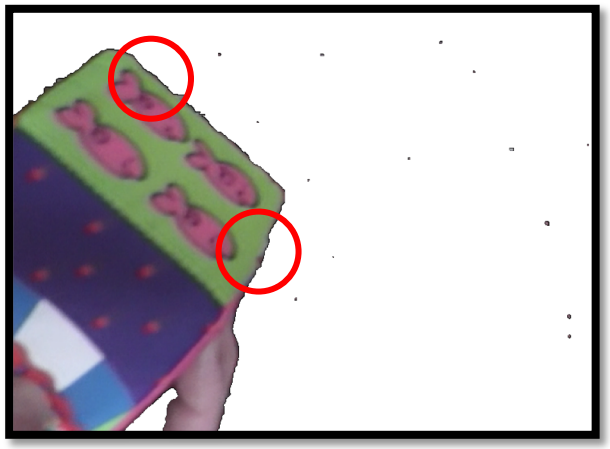
Grouping by *common motion*
[Spelke 1990]



Motion discontinuities: Early cue for occlusion Boundaries and '*ownership*' [Granrud et al. 1984]

model

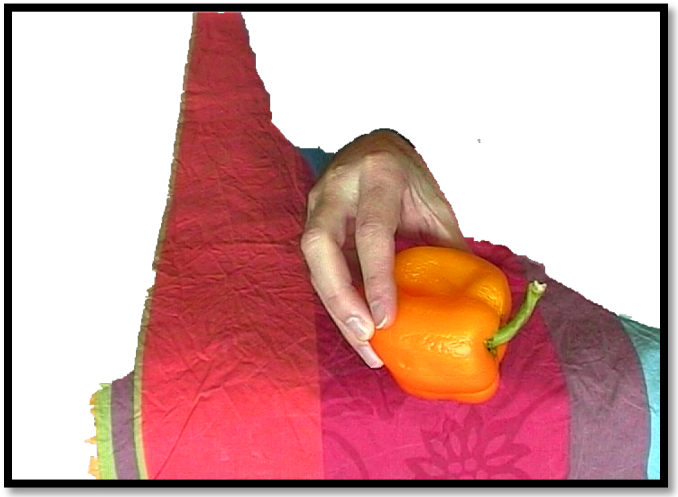

Motion-based segregation
Motion discontinuities
Common motion
Local occlusion boundaries
Object form
Static segregation



Motion segmentation can guide static segmentation of objects and their boundaries

Dorfman, Harari, Ullman CogSci 2013

Segmenting with ‘GrabCut’ algo.

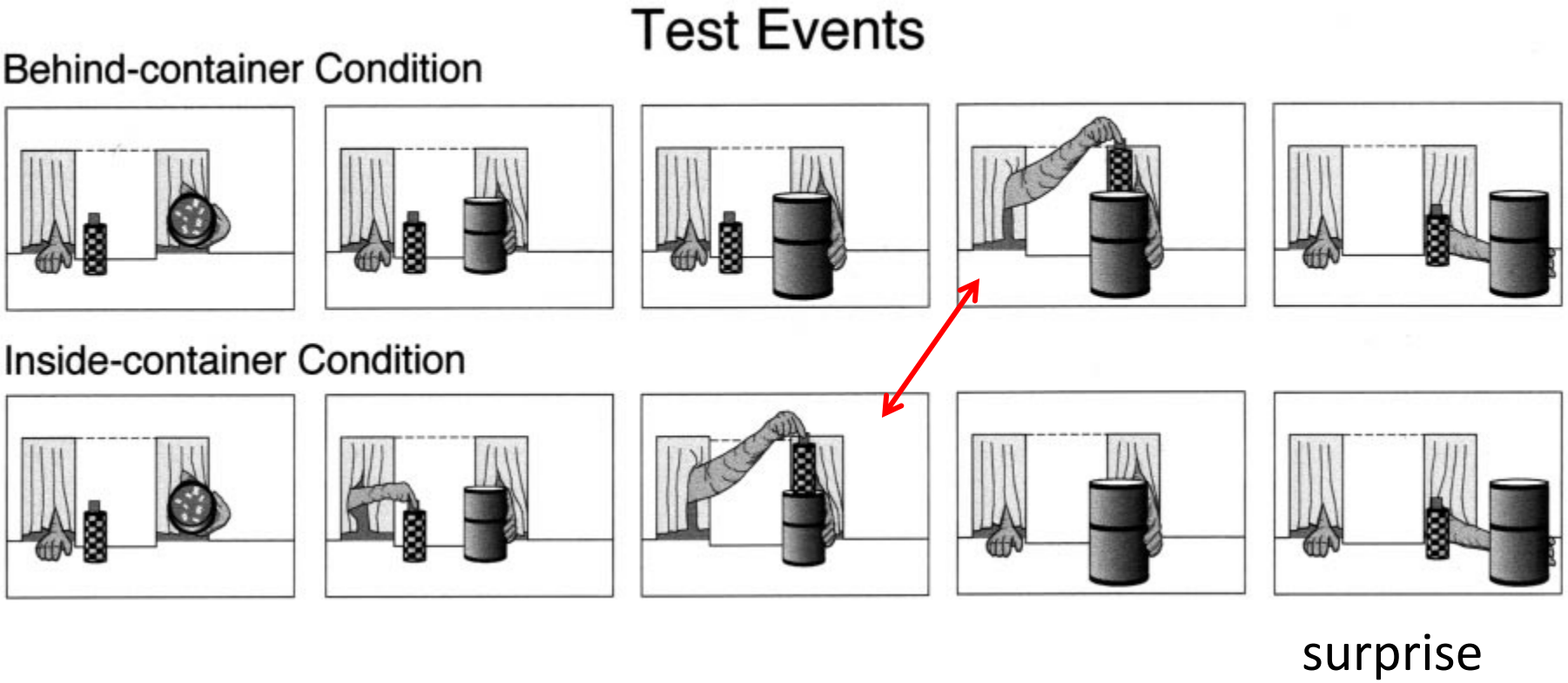


Default

With segregation cue

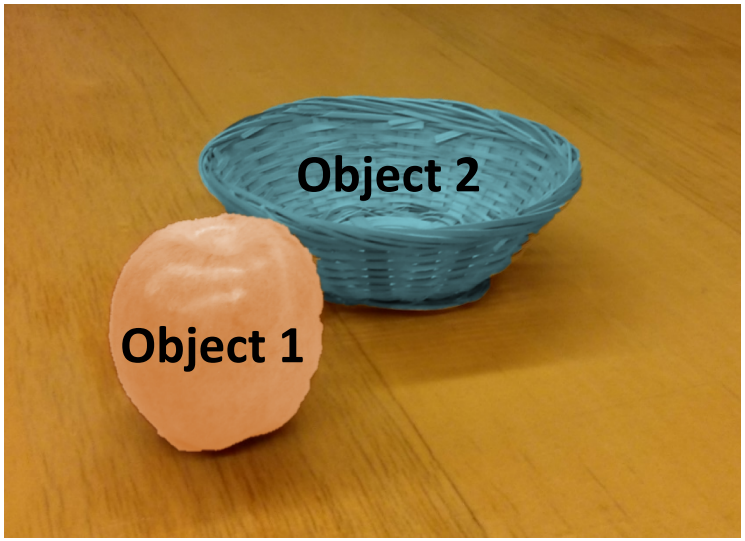
Spatial Relations

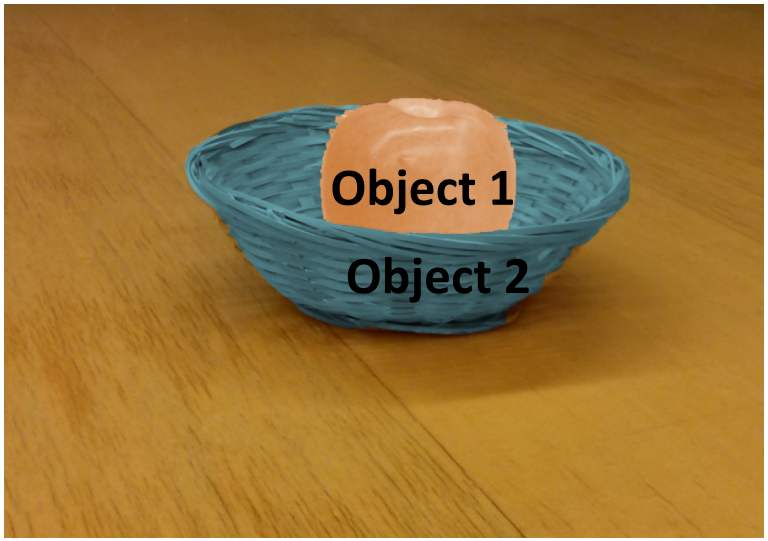
Expectations: container vs behind (transport)



[Hespos & Baillargeon 2001]

Paradoxical Occlusion

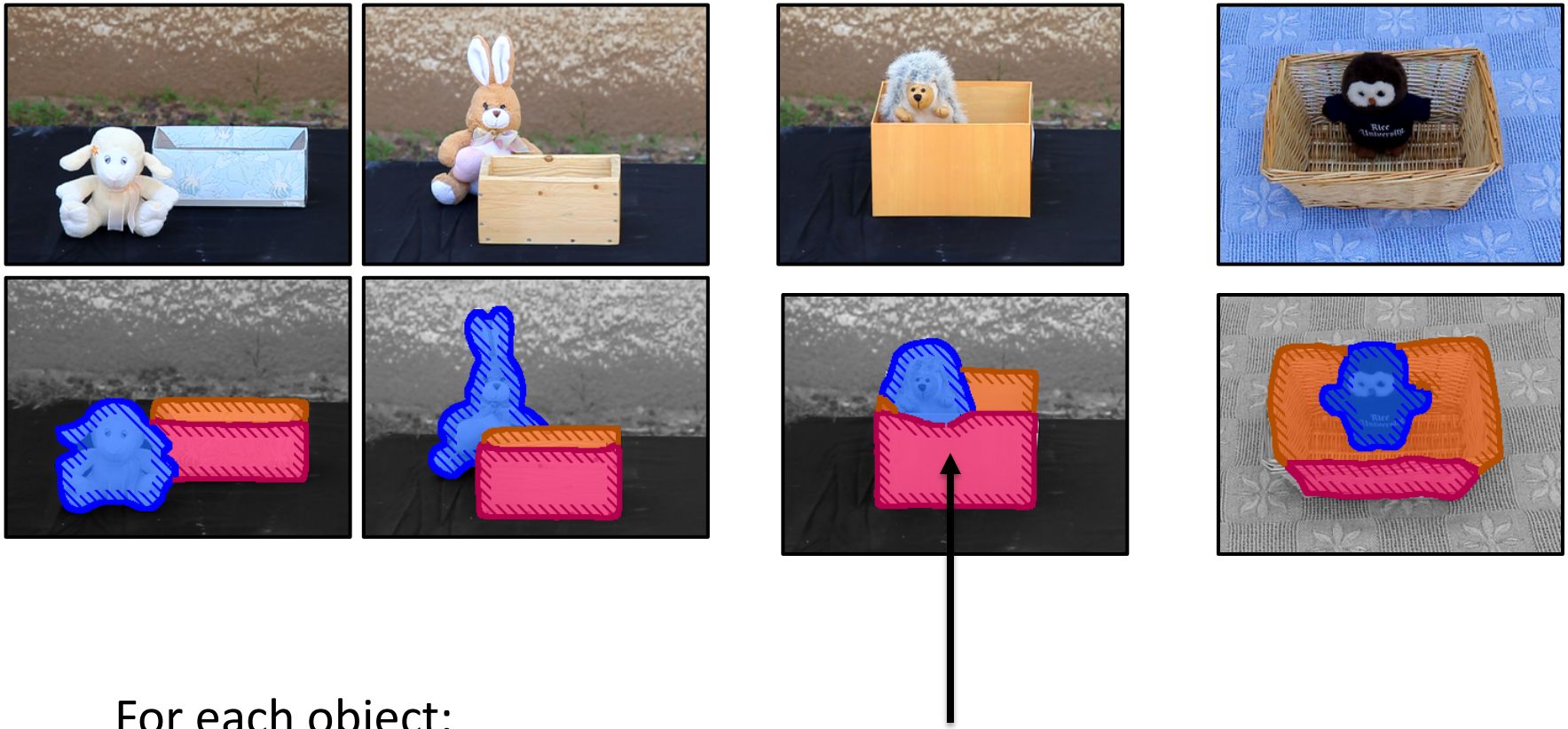




Object-2 > Object-1 > Object-2

Coincides with extensive learning of segregation and depth ordering
Violates normal expectation on depth ordering
Creates a special event, which guides the learning of containment

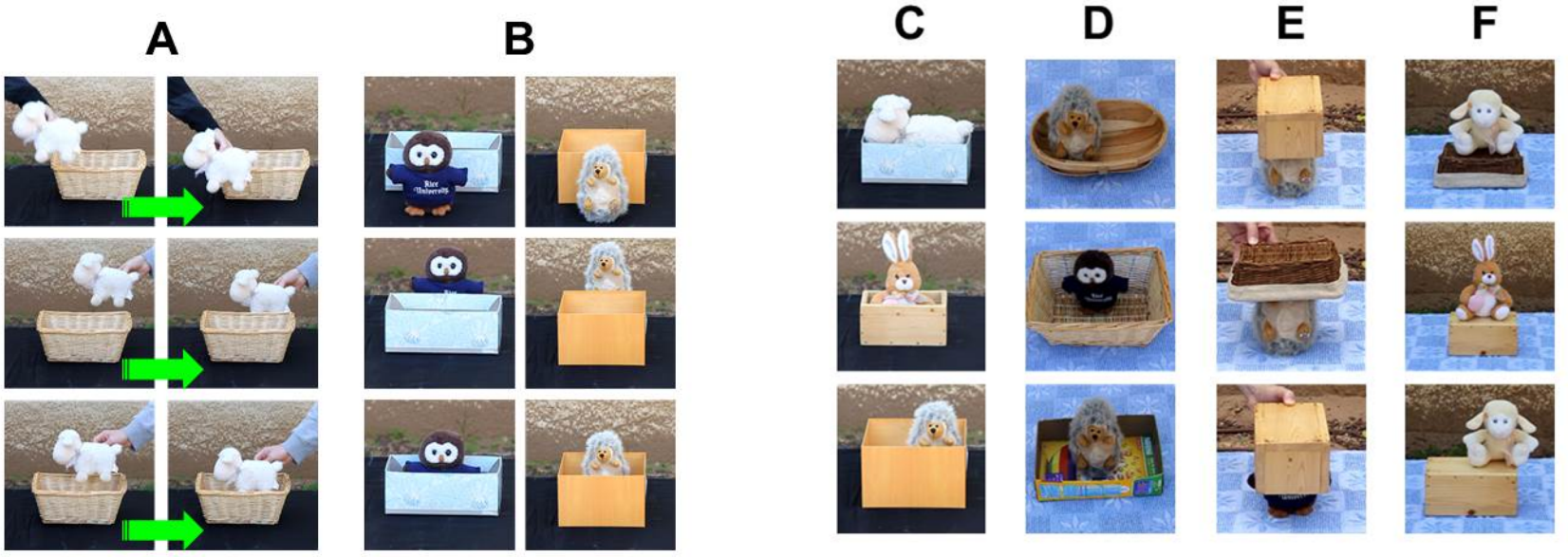
Model training on dynamic events
Learns dynamic and static containment



For each object:
Identify all the object boundaries
External and internal
With boundary ‘ownership’

Example Images handled by the model

Learning trajectory



- Innate structures and programs play a major role in intelligent behavior
- Innate structures are combined with learning

- For example: DNN with SGD
  - Initial structure with learning weights
- Humans have much more

- How do we find the useful innate structures?

- Studying humans
- Computational search (AutoML as a limited version)

Learning and innate structures

- Complex concept neither learned on its own nor innate.
- Domain-specific innate structures
- Not full solutions, but proto-concepts and strategies
- Not hands, but movers etc.
- Guide the system to develop meaningful representations
- Provide internal supervision
- 'Learning trajectories': mover – hand – reaching-for-object
  - Paradoxical occlusion – containers – hidden object location
- Can extract meaningful concepts event when they are non-salient in the input
- From cognition to AI: incorporate similar structures in computational systems