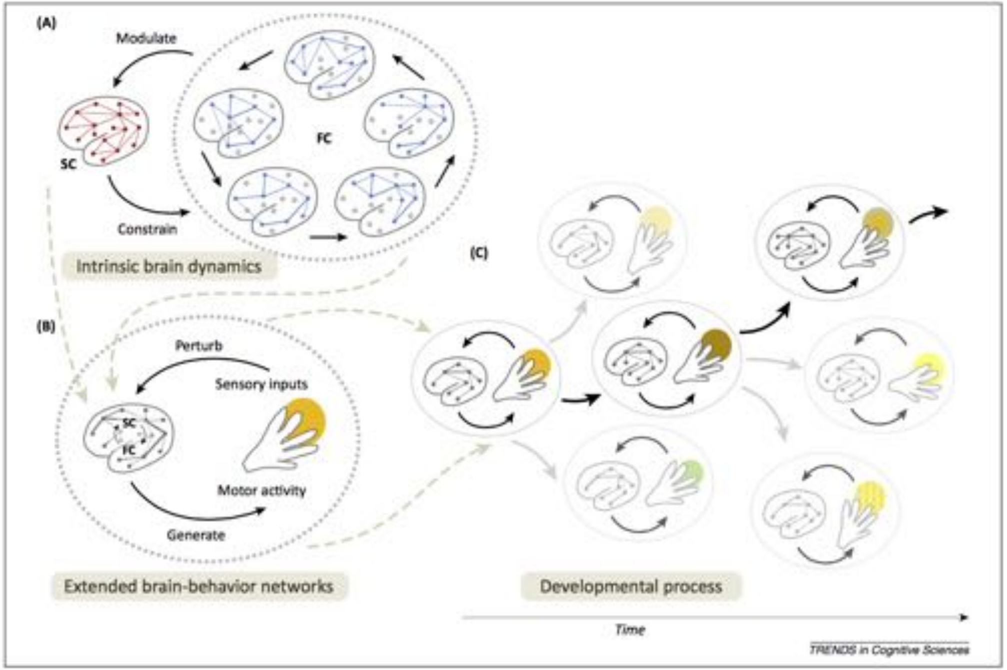
Developing visual environments

(aka *learning from the infant's point of view*)

Linda B. Smith
Indiana University
University East Anglia (part time)

The infant and the infant's visual environment co-develop



Byrge, L., Sporns, O. & Smith, L. B. (2014) Developmental process emerges from extended brain-body-behavior networks. *Trends in Cognitive Sciences*, 18(8), 395-403.

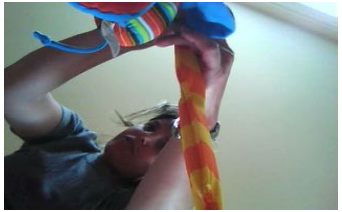

Multisensory

High resolution but naturalistic

Toyroom

Free-flowing, cluttered
In the lab

Homeview



Indiaview

At home, everyday experience

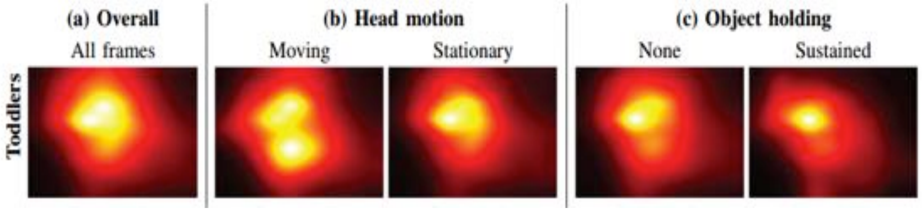
Infants: 1 month to 24 months of age

Yoshida & Smith (2008) *Infancy*
Yu et al (2009) *IEEE Transactions on Autonomous Mental Development*
Pereira, James, Jones & Smith (2010) *Journal of Vision*
Smith, Yu, Pereira, (2011) *Developmental Science*
Street, James, Jones & Smith (2011*) Child Development*
Yu & Smith (2012) *Cognition*
Yurovsky, Smith & Yu (2013) *Developmental Science*
Yu & Smith *(2013) PLoS One*
James, Swain, Jones & Smith (2014*) JCD*
Pereira, Smith & Yu (2014)*, Psychological Bulletin & Review*
James et al (2014) *Developmental Science*
Jayaraman, Fausey & Smith (2015) PLoS One
Fausey , Jayaraman & Smith, (2016) *Cognition*

In aggregate 1000 hours of head –camera video, 100s million images extracted

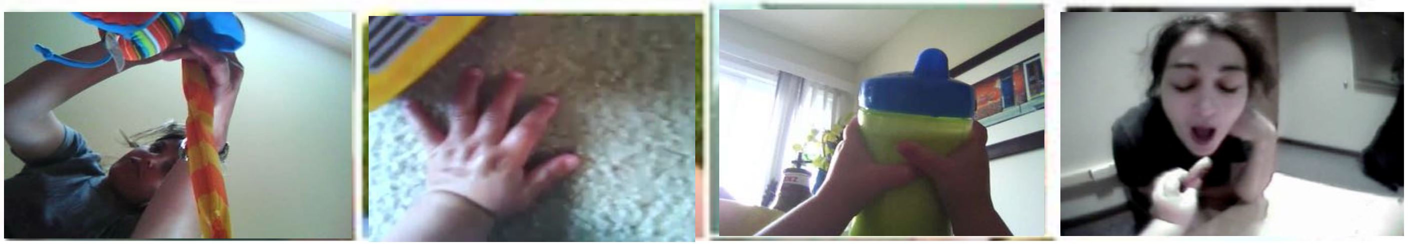
Yu & Smith (2016) *Current Biology*
Clerkin, Hart, Rehg, Yu & Smith (2017) *Royal Society B*
Jayaraman, Fausey & Smith (2017) *Developmental Psychology*
Suanda, Smith& Yu, (2017) *Developmental Neuropsychology*
Yu. & Smith (2017) *Child Development*
Smith, Jayaraman, Clerkin & Yu (2018) *Trends in Cognitive Science*
Jayraman & Smith (2018) *Vision Research*
Li et al (2017) ICML
Bambach,Yu,Smith & Crandall(2018) NeurIPS
Slone , Smith & Yu (2019*) Developmental Science*

Head cameras



Gaze distribution toddlers in various active contexts

Bambach et al (2016)
Borjon et al (2019)
DeSerio, Yu, Gold, Candy, Smith (in progress)



HIGHLIGHTS: what we have learned

Highlight 1: The data for visual learning changes markedly with every advance in motor abilities, **a curriculum for visual learning ordered and structured by development itself**

Smith, L. B., Jayaraman, S., Clerkin, E. & Yu, C. (2018) The Developing Infant Creates a Curriculum for Statistical Learning. **Trends in Cognitive Sciences**, 4, 325-336. PMID: 29519675 PMCID: PMC5866780

Smith, L. B. & Slone, L. K. (2017) A Developmental Approach to Machine Learning?. Fr**ontiers in Psychology**, 8:2124. PMCID: PMC5723343

Homeview Project

National Science Foundation

Indiana University (EAR): Learning: Brains, Machines & Children

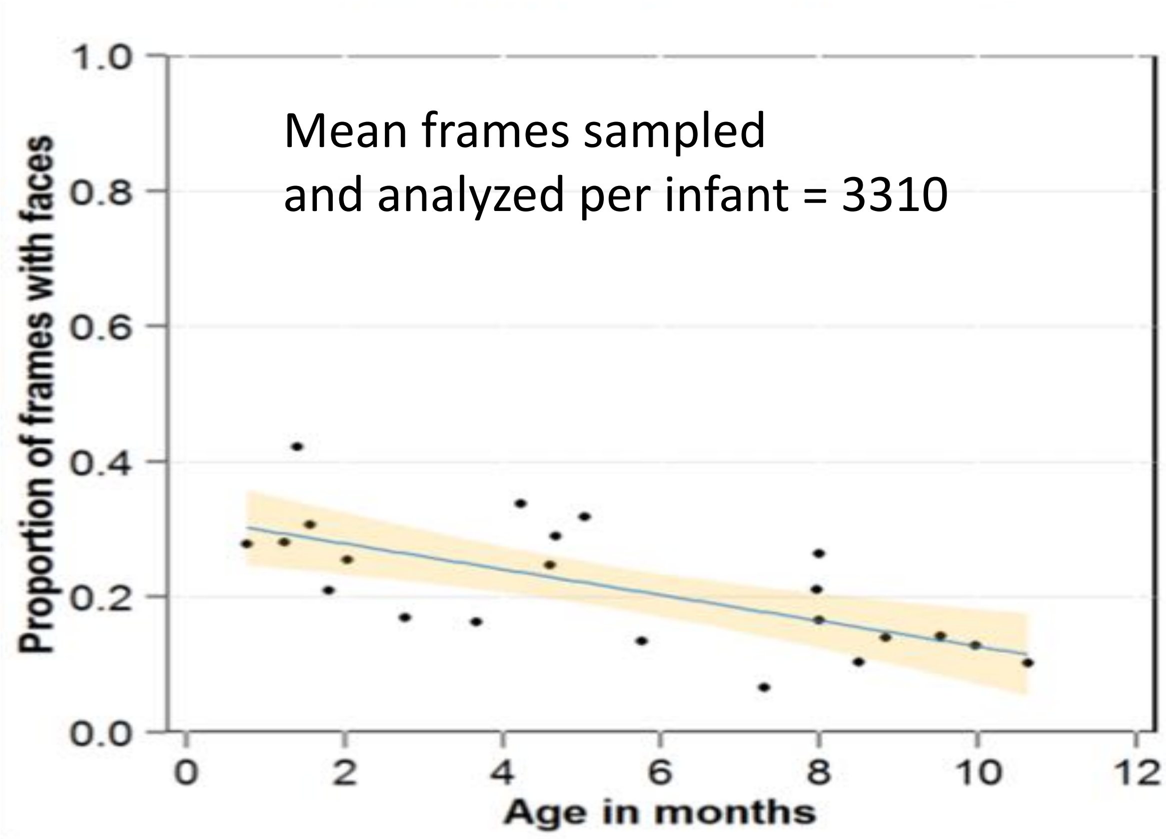
A corpus of developmentally indexed egocentric scenes

- 101 infants
- 3 weeks – 24 months
- 4 to 7 hours of head-camera video daily life
- (no experimenters present, camera on hats)
- records at 30Hz
- over 500 hours of video

The visual data for learning change with development

Movie by Jason Gold

Proportion frames with faces

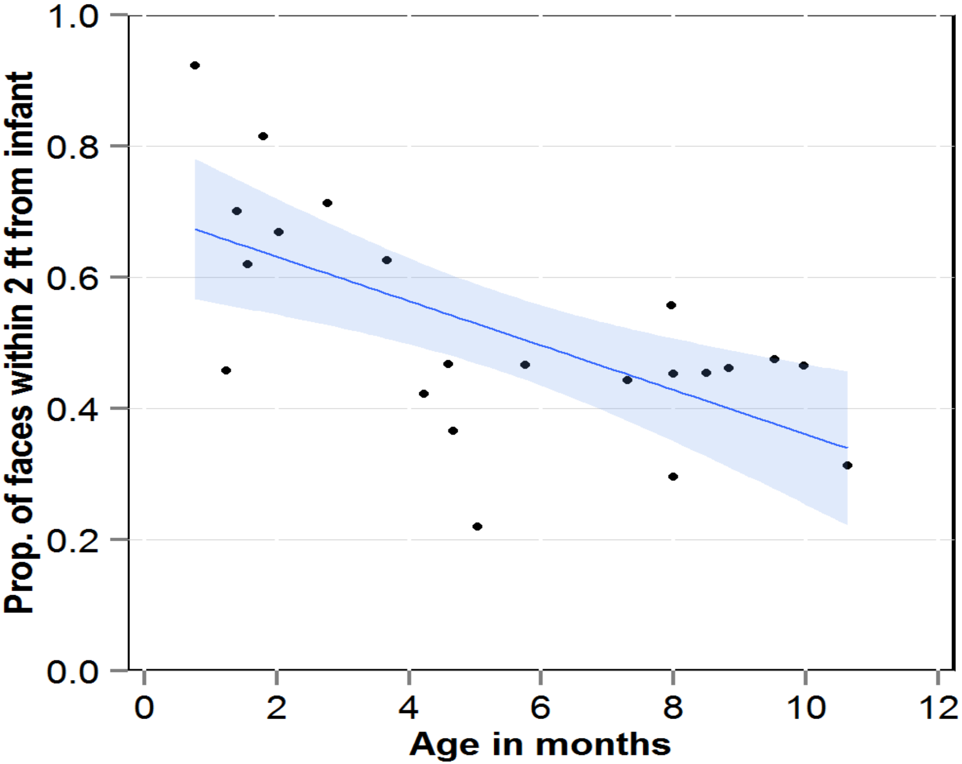


Mean frames sampled
and analyzed per infant = 3310

Jayaraman, Fausey & Smith (2015) *PLoS One*

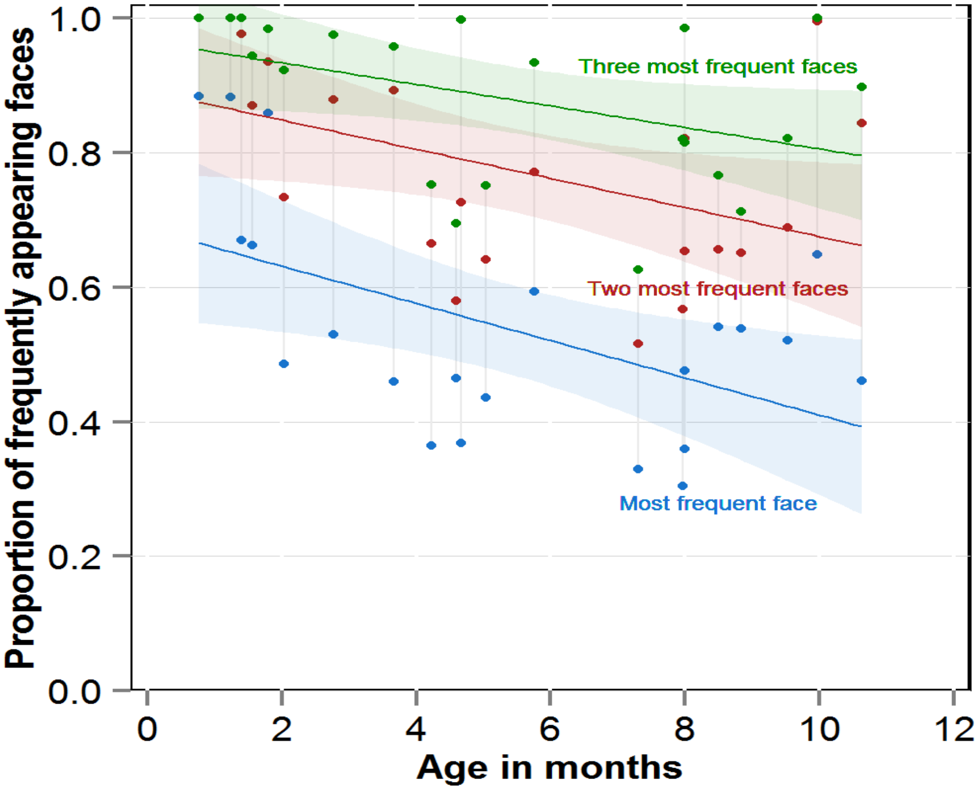
$R^2 = .42$, $F(1, 20) = 16.11$. $p < .001$

Proportion faces within 2 feet of head camera



$R^2 = .37, F(1, 20) = 13.61, p < .05$

Three individuals account for most of the faces

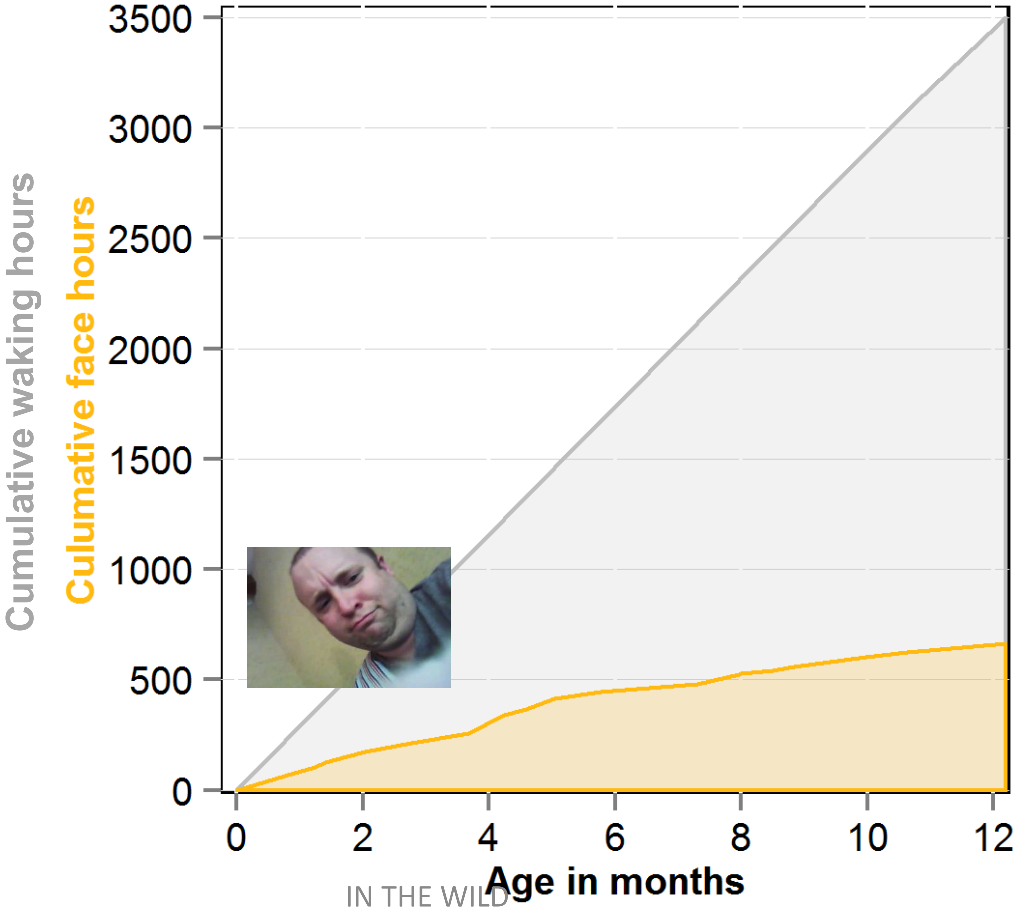


R^2 = .14, $F(1, 20) = 4.516$. $p < .05$.

R^2 = .16, $F(1, 20) = 5.24$. $p < .05$.

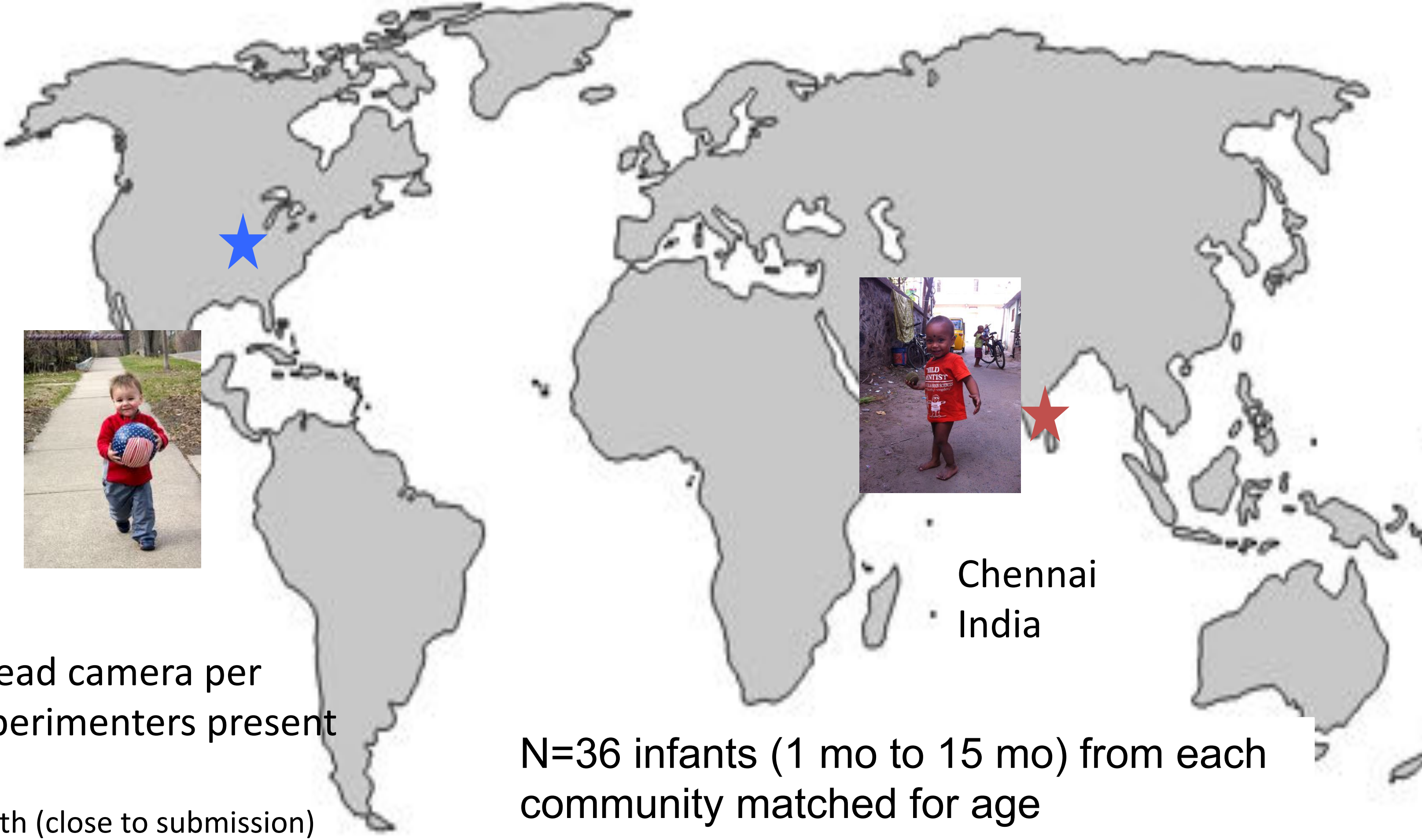
R^2 = .23, $F(1, 20) = 7.413$. $p < .05$

Cumulative waking hours and face hours



Highlight 2: Early environments may be tightly constrained.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325-336.



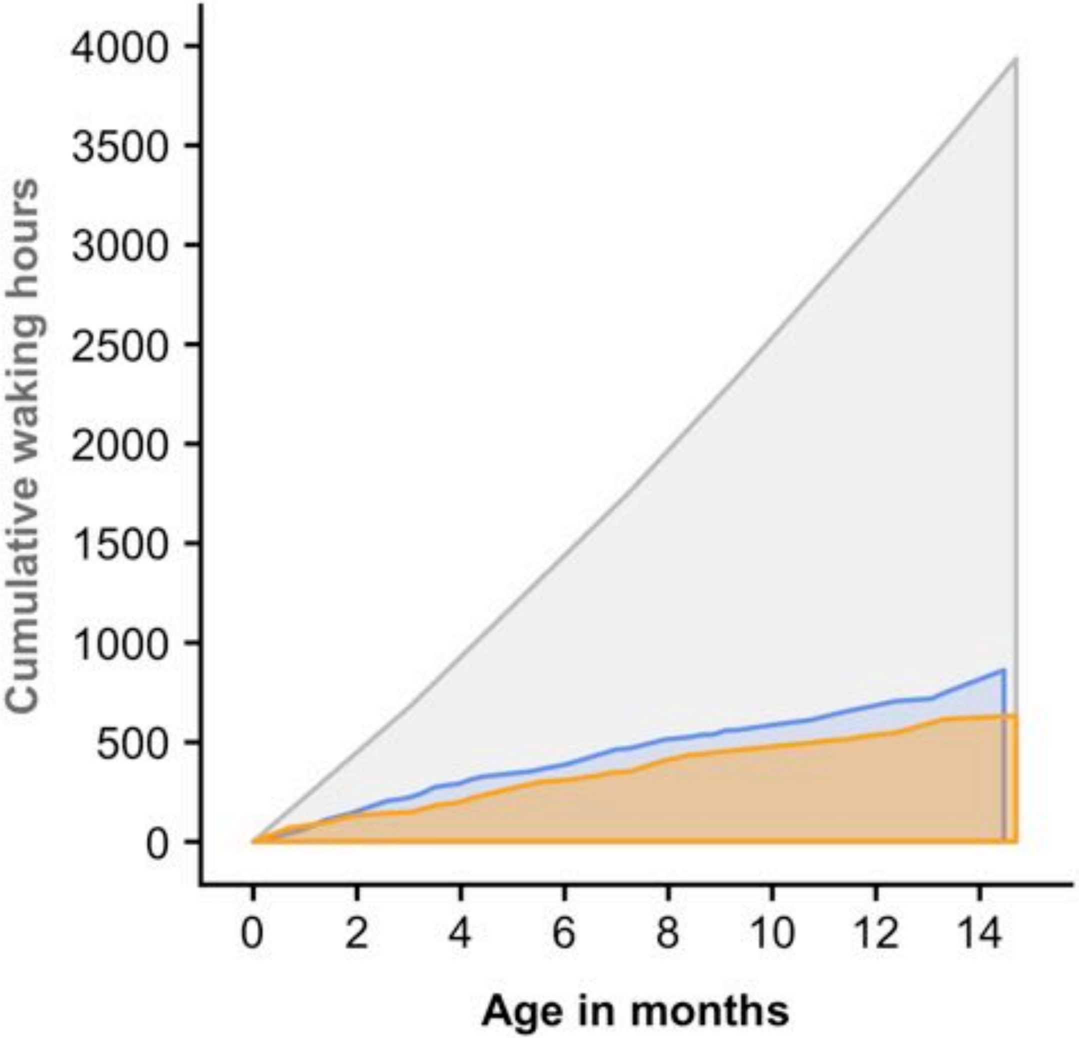
Bloomington
Indiana

Chennai
India

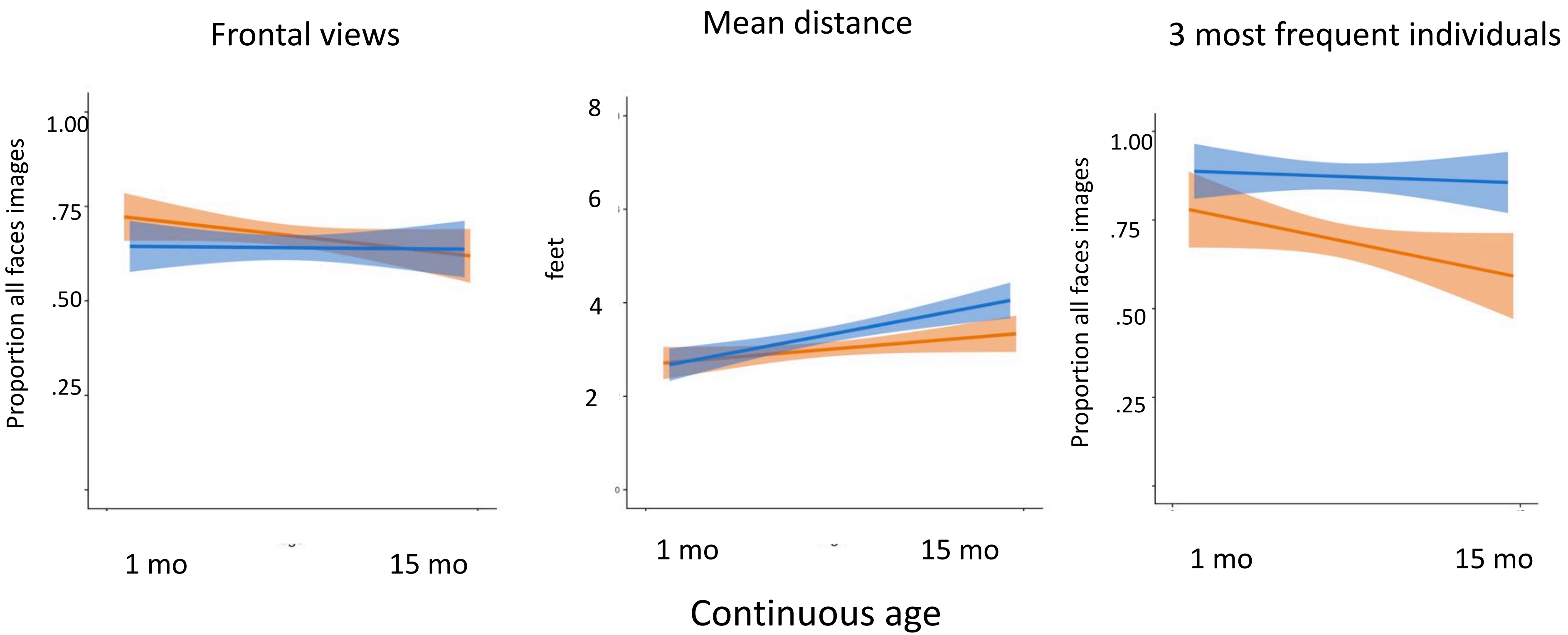
4 to 8 hours head camera per
Infant; No experimenters present

N=36 infants (1 mo to 15 mo) from each
community matched for age

Jayaraman & Smith (close to submission)

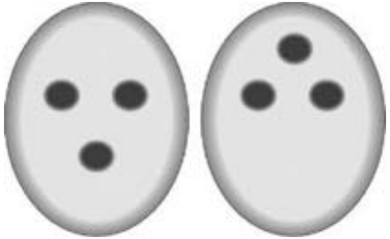


US blue
India orange



Signature properties

Close frontal views of faces



Fantz, 1963
Johnson & Morton, 1991

Developmental Science 10:1 (2007), pp 40–47 DOI: 10.1111/j.1467-7687.2007.00562.

Sleeper effects

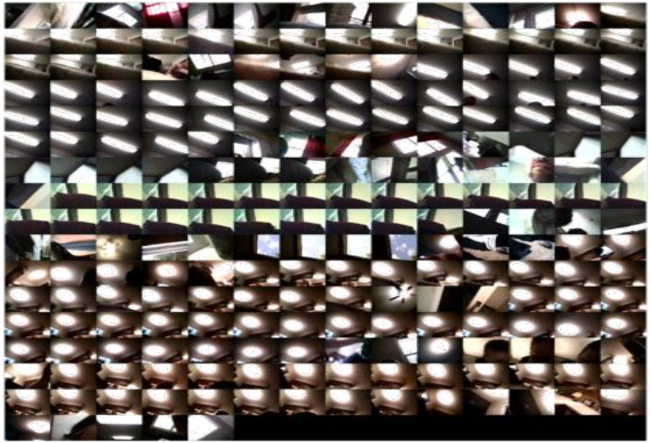
Daphne Maurer,[1,2] Catherine J. Mondloch[3] and Terri L. Lewis[1,2,4]

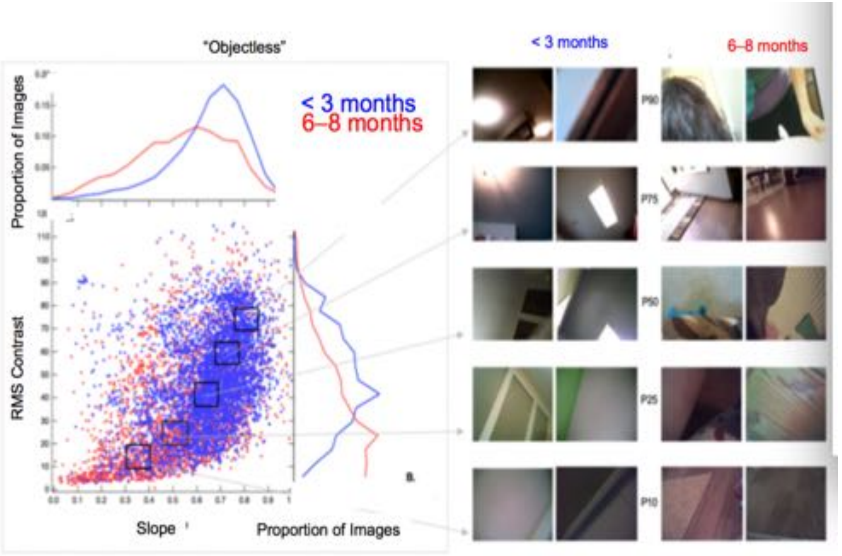
1. *Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Canada*
2. *Department of Ophthalmology, The Hospital for Sick Children, Toronto, Canada*
3. *Department of Psychology, Brock University, St. Catharines, Canada*
4. *Department of Ophthalmology & Vision Sciences, University of Toronto, Canada*

Under 3 months of age

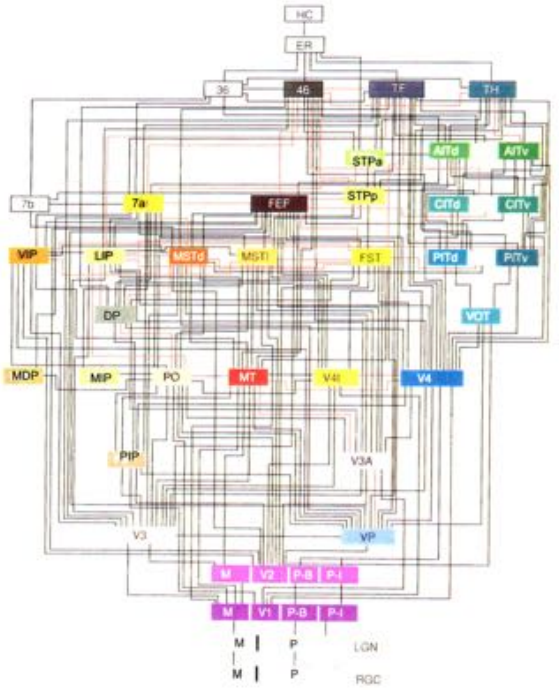
Nearly 10 minutes out of every hour, and enduring in time

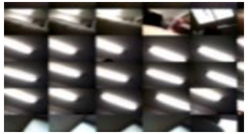
Candy, DeSerio, Gold & Smith
(close to submission…)

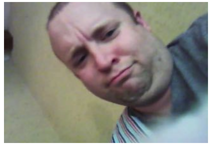




High contrast
Low spatial frequency
Simple structure (orientations)







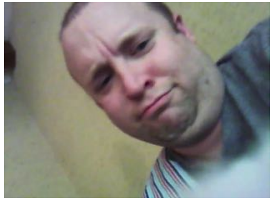
25 minutes out of every hour

High contrast, low spatial frequency
faces and simple edges

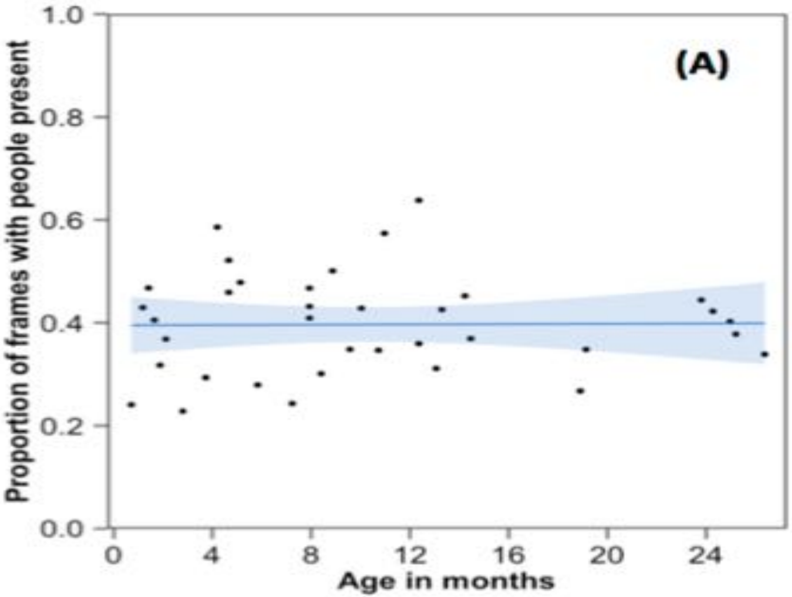
What is this teaching the visual system?

Highlight 3: Sensory motor development creates **developmental niches** for solving different visual learning problems

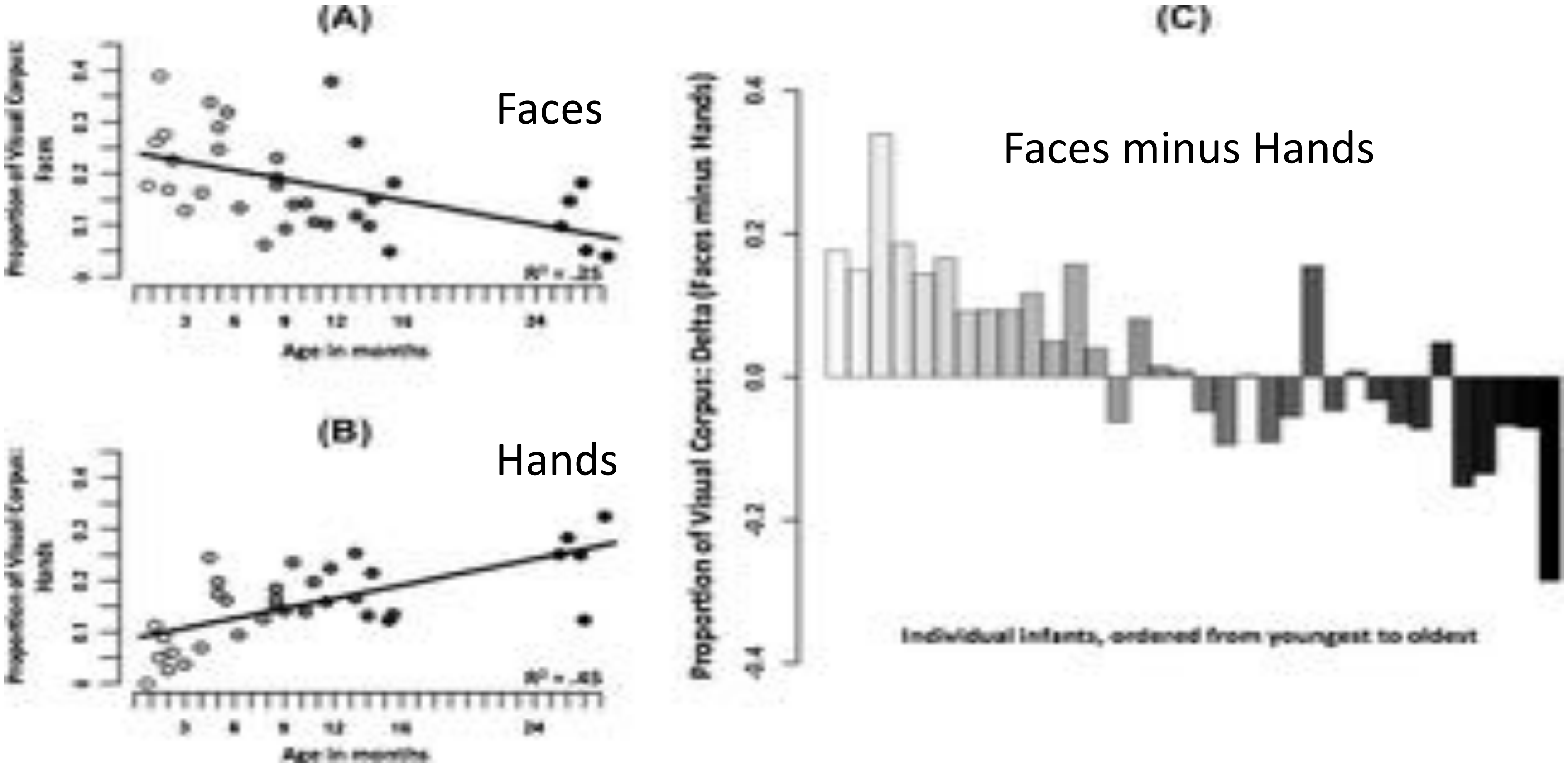
Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325-336.



It’s faces that decline with age, not people in view

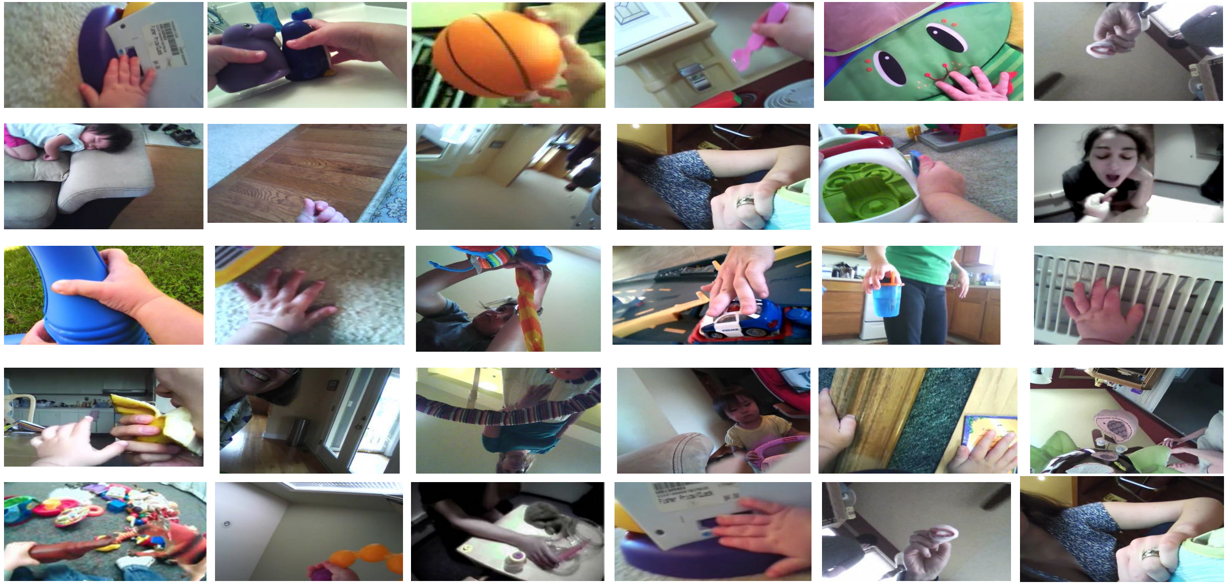


Jayaraman, Fausey & Smith (2017) *Developmental Psychology*

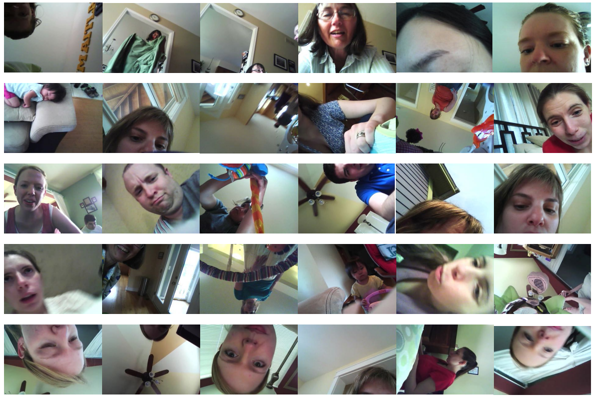


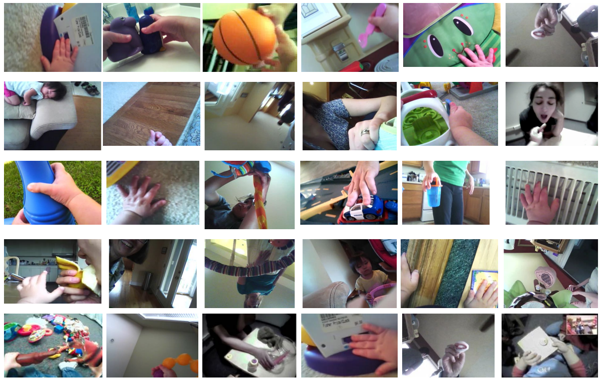
Fausey ,Jayaraman & Smith, (2016) *Cognition*

At all ages more than 70% of the hands in view are touching an object

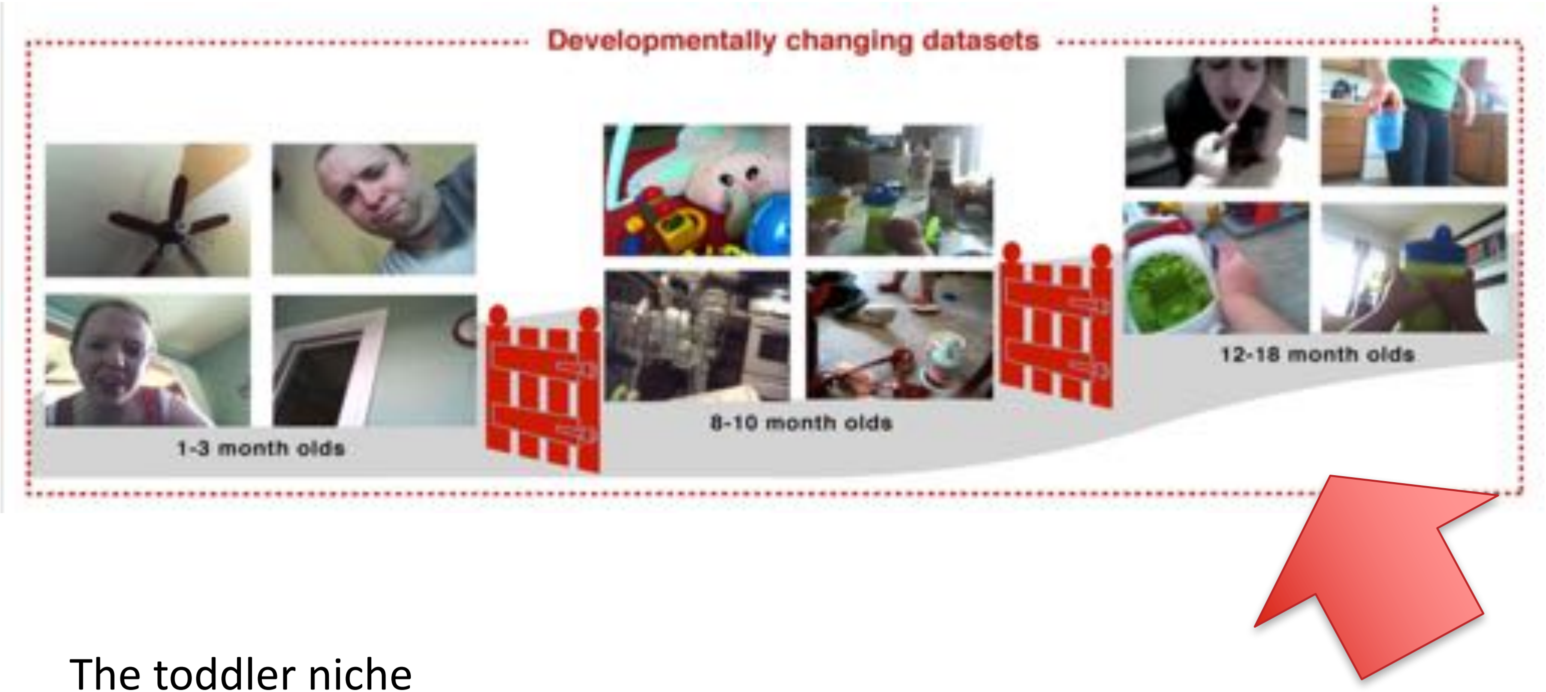


The contents of visual experience change: from dense with *faces to dense with hands acting on objects*





Highlight 4. Hands are as important as eyes for the development of visual object recognition



The toddler niche

Multi-sensory project
(NSF, NICHD, Eye-Institute, AFOSR)

Dual head-cameras, or head-mounted eye trackers

Motion sensors (hands, heads, eyes)

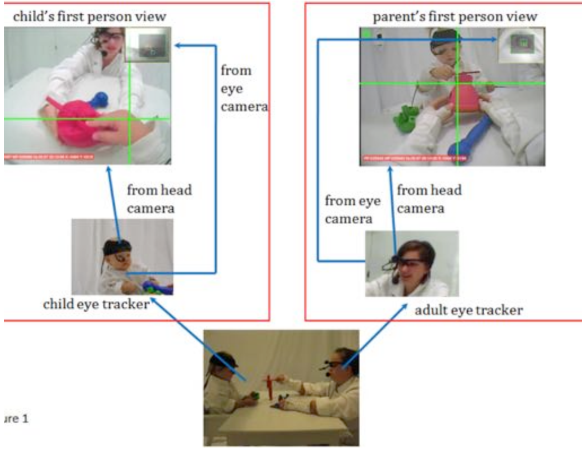
Audio

Multiple room cameras

Parent-infant play with multiple toys

Nearly 200 infants (longitudinal & cross-sectional)

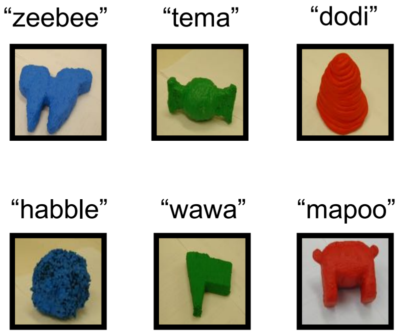
9 months to 36 months



into the wild -- CBCD

Optimal moments for **learning object names and visual object categories**?

Novel Words & Objects



Parent-Toddler Free play

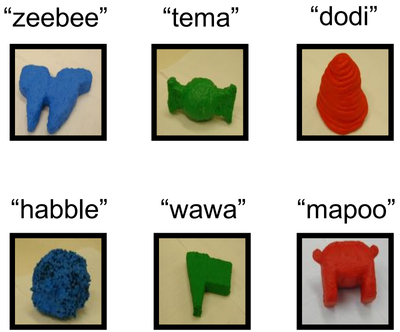
Word Learning Test

Pereira, A., Smith, L. B. & Yu, C. (2014) A Bottom-up View of Toddler Word Learning. *Psychonomic Bulletin & Review*, 21, 178-185.

Yu, C. & Smith, L. B. (2012) Embodied Attention and Word Learning by Toddlers. *Cognition*, 125, 244-262

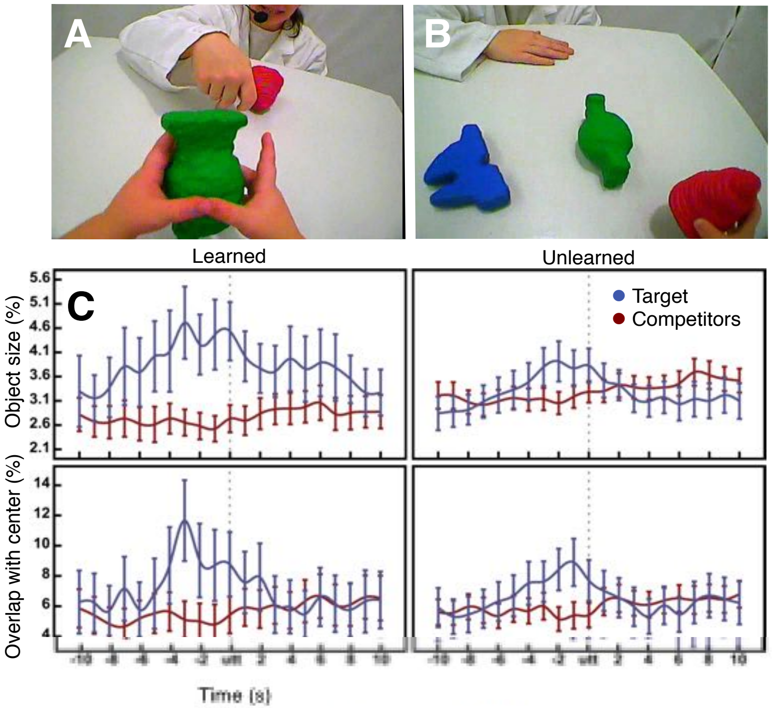
How do the naming moments associated with learned names differ from the naming moments associated not learned names?

Novel Words & Objects



Parent-Toddler Free play

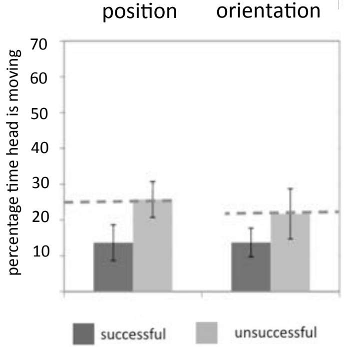
Word Learning Test



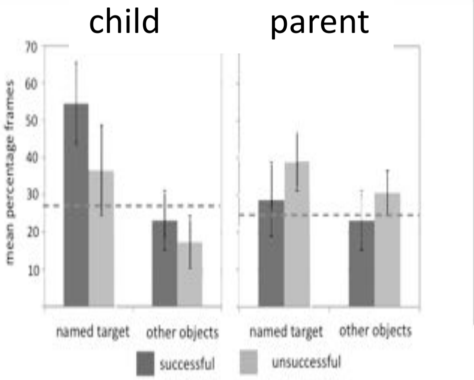
a direct consequence of the **toddler's body** of **how their sensory-motor system works**, and how they interact with the world

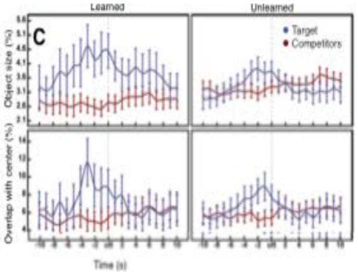
The body of a toddler old creates unique visual data for learning

Head stability



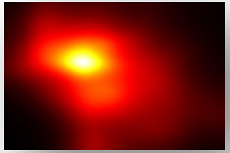
Holding behavior



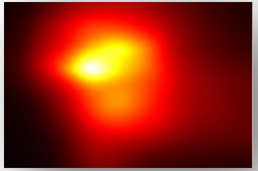


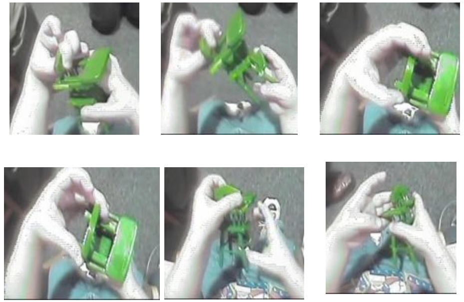
Head and eye alignment

Holding Toy



Empty Hands





Self-generated views

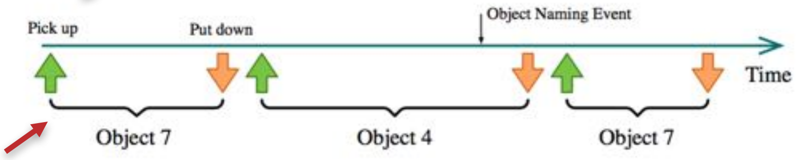
Pereira, A., Smith, L. B. & Yu, C. (2014) A Bottom-up View of Toddler Word Learning. *Psychonomic Bulletin & Review*, 21, 178-185.

Yu, C. & Smith, L. B. (2012) Embodied Attention and Word Learning by Toddlers. *Cognition*, 125, 244-262

Bembach, S., Smith, L.B., Crandall, D. & Yu C. Objects in the center: Heads, hands, eyes and sustained attention in toddlers (in preparation)

Slone, L. K., Smith, L. B. & Yu, C. (2019) Self-generated variability in object images predicts vocabulary growth. *Developmental Science.*

Many views of a single object



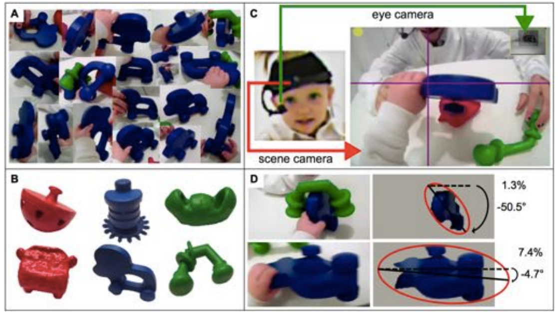
A teaching signal from holding– all these views are the same thing

Self-generated input related to the current state of knowledge

High image variability of a single object

Clustered repetitions in time

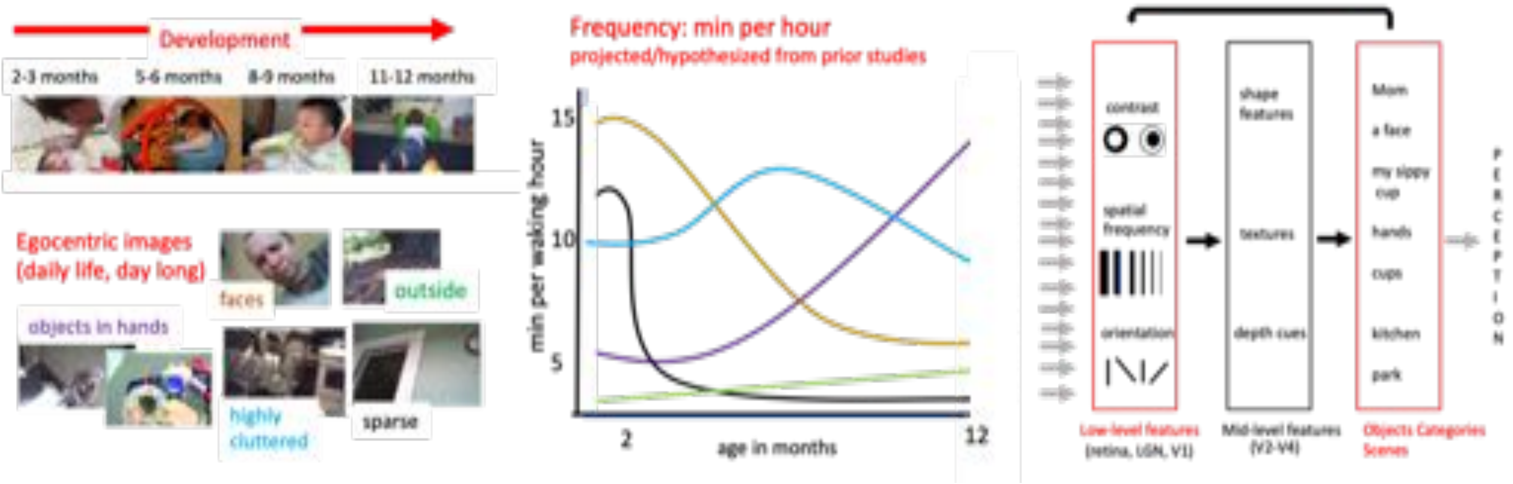
High image level variability of *individual* objects



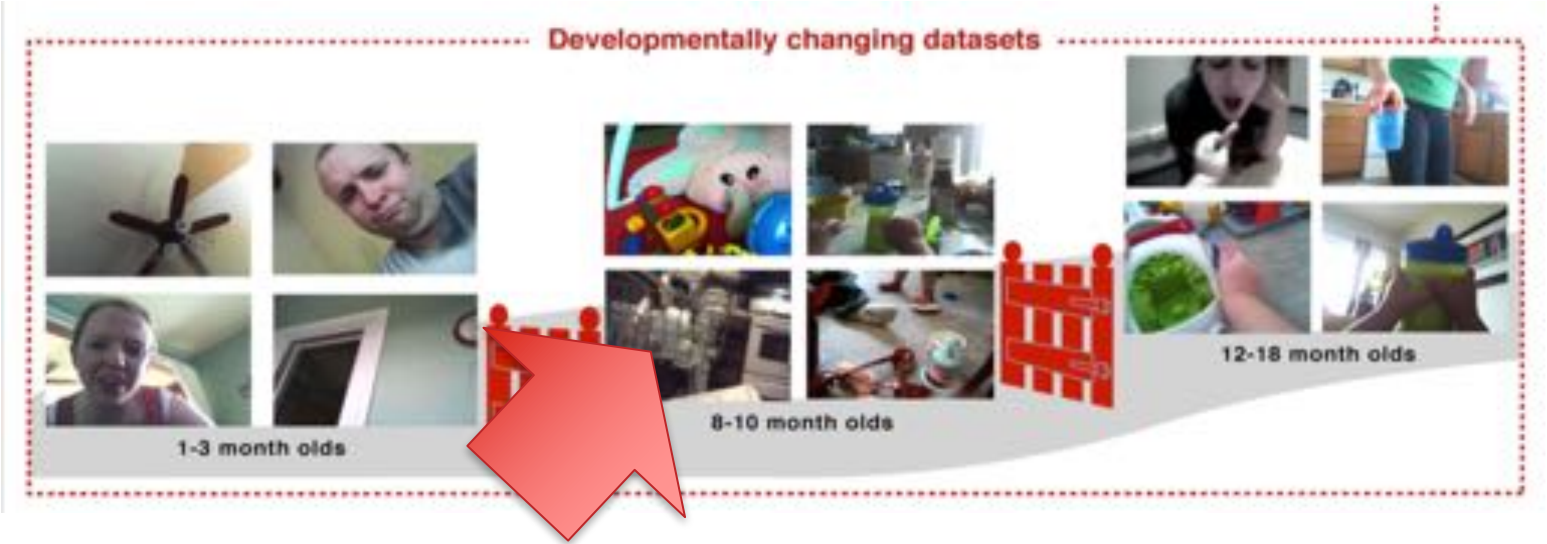
A developmental niche for generating optimal data for learning to visual recognize (and categorize) objects?

Stojanova et al (2019) Incremental object learning from contiguous views **CVPR**
Bambach, S., Crandall, D. J., Smith, L. B. & Yu, C. (2018) Toddler-Inspired Visual Object Learning. Advances **in Neural Information Processing Systems** 31.
Slone, L., Smith, L.B., Yu, C (2019). Self-generated variability in object images predicts vocabulary growth. **Developmental Science**

Niches: How does this ordered visual development matter to the visual system?



The system is being changed at low, middle, and high-levels of features analysis.
What are the correlations – in the data?



Highlight 4: The frequency and temporal properties of real world environments have unchartered statistical regularities likely essential to understanding how infants learn so much in such relatively brief periods.

Clerkin, Hart, Rehg, Yu & Smith (2016) *Royal Society*

8 infants 8 to 10 months of age

147 **events Involving any kind of food or dishes (~9K images)**

Images sampled at 1/5 Hz

Scene to text coding

About 500 naïve adults (Amazon Mechanical Turk)

Task: name **up to five** objects,

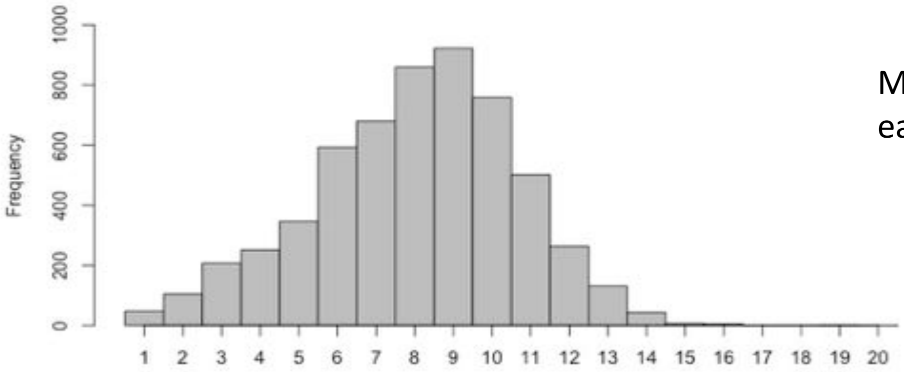
most obvious objects in the scene,
using **basic level nouns**
(no people, no body parts)

Coders saw images in **sequential** sets of 20

Sampled (within a "mealtime") at 1/5 Hz

Each scene was coded by 4 individuals

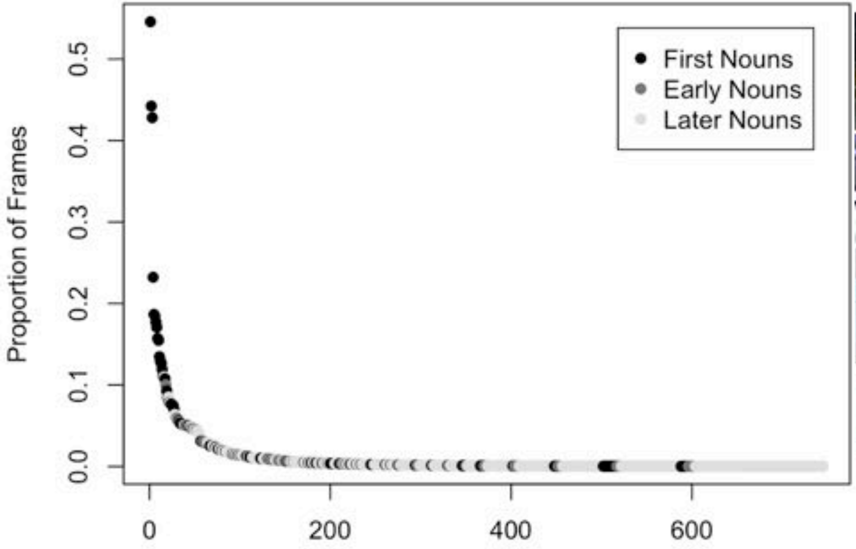
8 to 10 month olds



Number of unique object categories

Many different object categories in each scenes

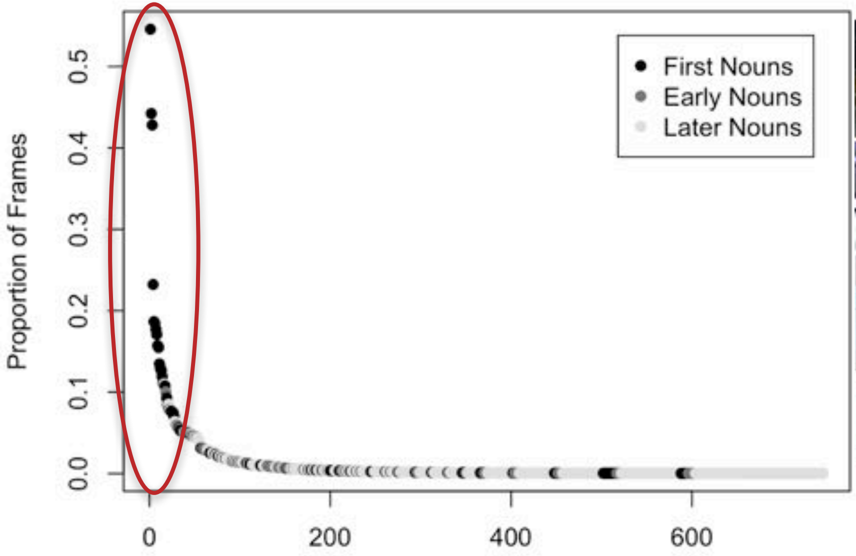
(a) Frequency of objects in head camera images



Unique object categories ordered by frequency

(b) Example images

(a) Frequency of objects in head camera images

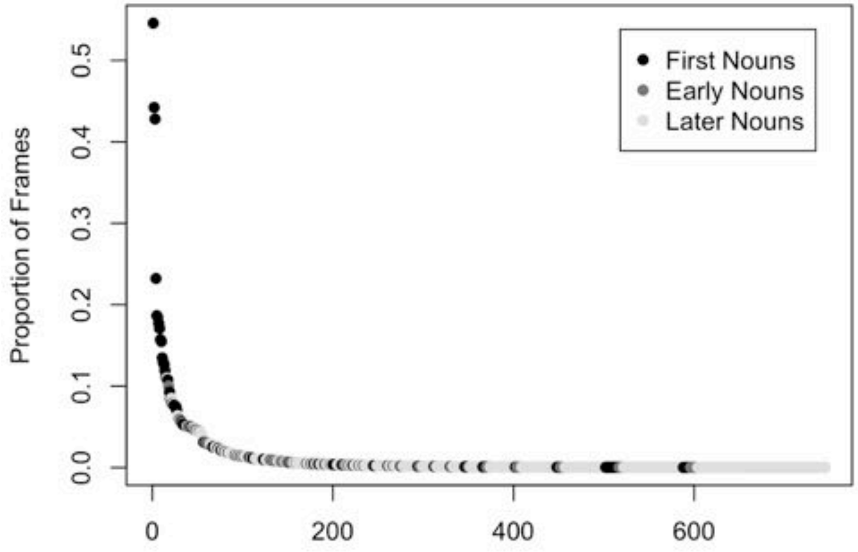


Unique object categories ordered by frequency

(b) Example images

Visual pervasiveness from many different viewpoints
may be critical to *visual* learning: **segmentation**, **finding things in clutter**, **building strong visual memories**

(a) Frequency of objects in head camera images



Unique object categories ordered by frequency

(b) Example images

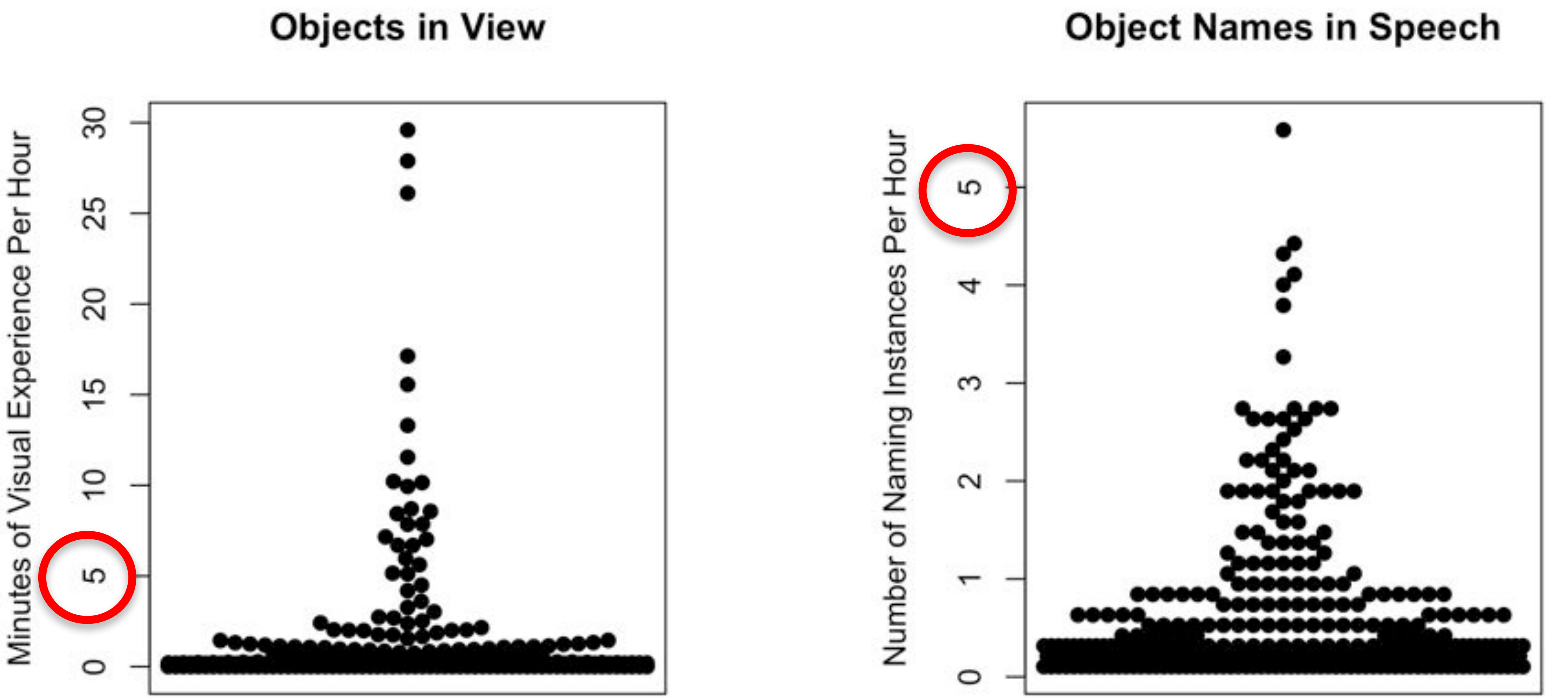
Visual pervasiveness likely also means the pervasiveness of their spoken names, right?

New larger data set
Clerkin & Smith (in preparation)

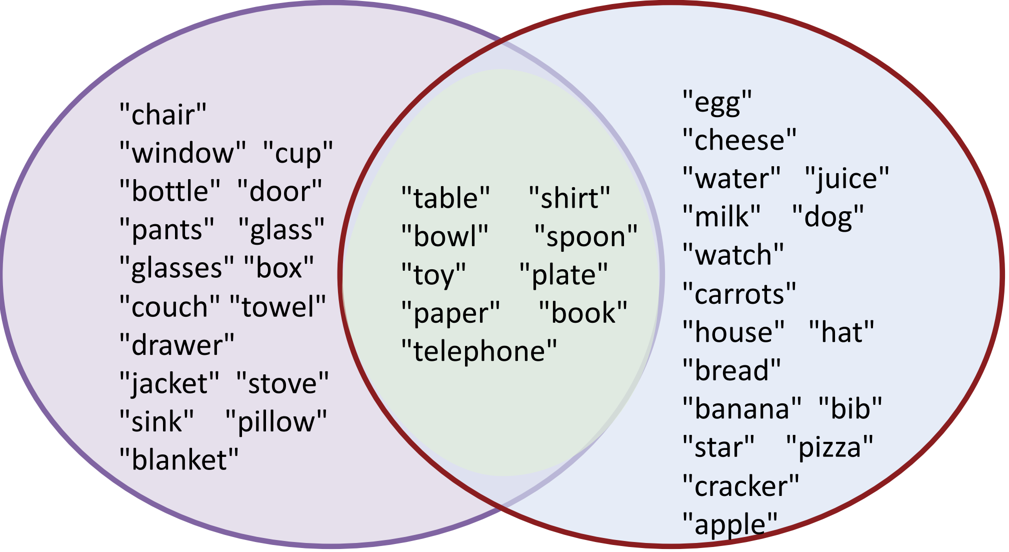
- 14 subjects (8 female)
- 7-11 months old

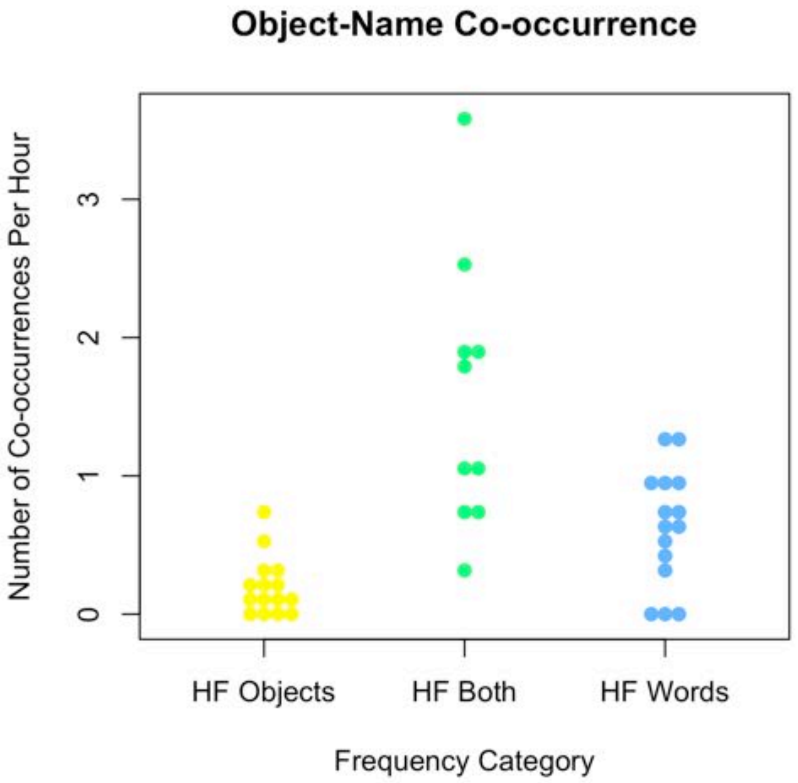
- 18.08 hours of clips involving food or dishes of any kind
- **458 individual meal times**
- sampled 1 image every 5 seconds
    - 13,000 coded scenes
  - All parent speech was also transcribed in 5 second increments
    - **47.87%** of the increments contained any speech
    - Parents are not talking all the time

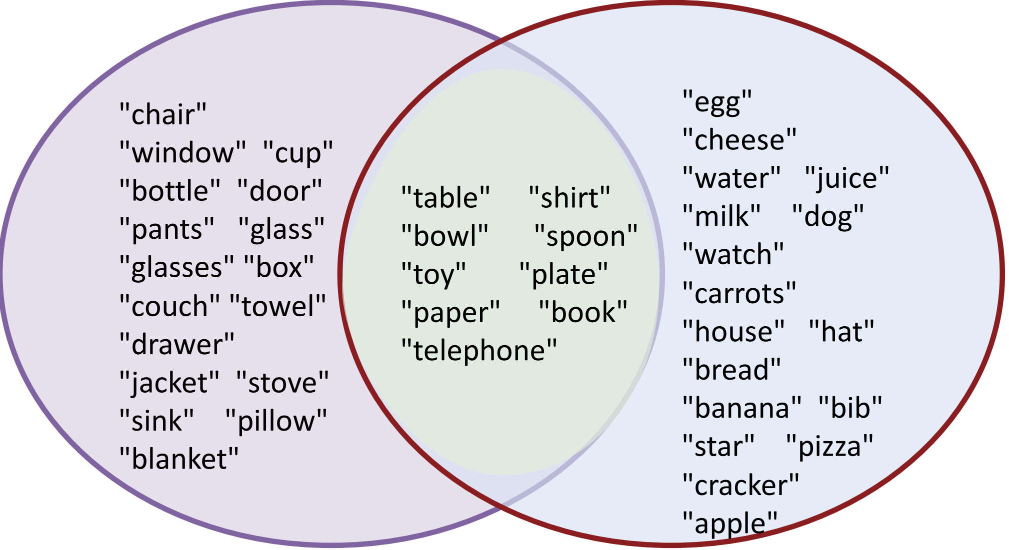
Frequencies of individual objects and names*

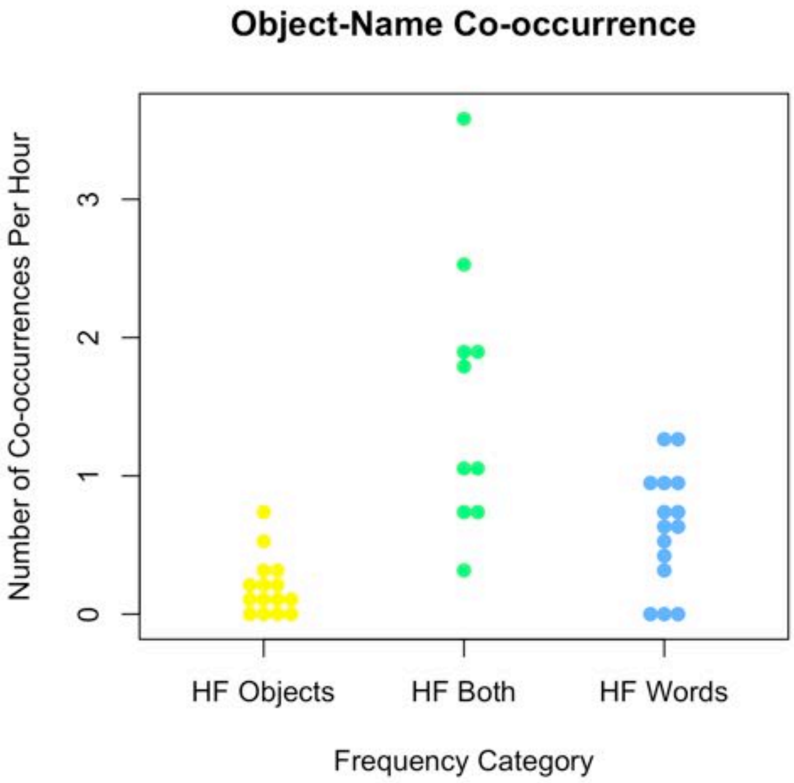


*Each dot represents a visual object/an object name that is normatively learned before 18 months and that occurred at least once visually OR was named OR both


"chair"
"window" "cup"
"bottle" "door"
"pants" "glass"
"glasses" "box"
"couch" "towel"
"drawer"
"jacket" "stove"
"sink" "pillow"
"blanket"
"table" "shirt"
"bowl" "spoon"
"toy" "plate"
"paper" "book"
"telephone"
"egg"
"cheese"
"water" "juice"
"milk" "dog"
"watch"
"carrots"
"house" "hat"
"bread"
"banana" "bib"
"star" "pizza"
"cracker"
"apple"

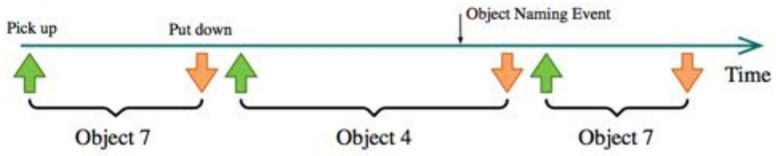

Object-Name Co-occurrence
Number of Co-occurrences Per Hour
0
1
2
3
HF Objects
HF Both
HF Words
Frequency Category





But is this the right scale to
to think about early object name learning?

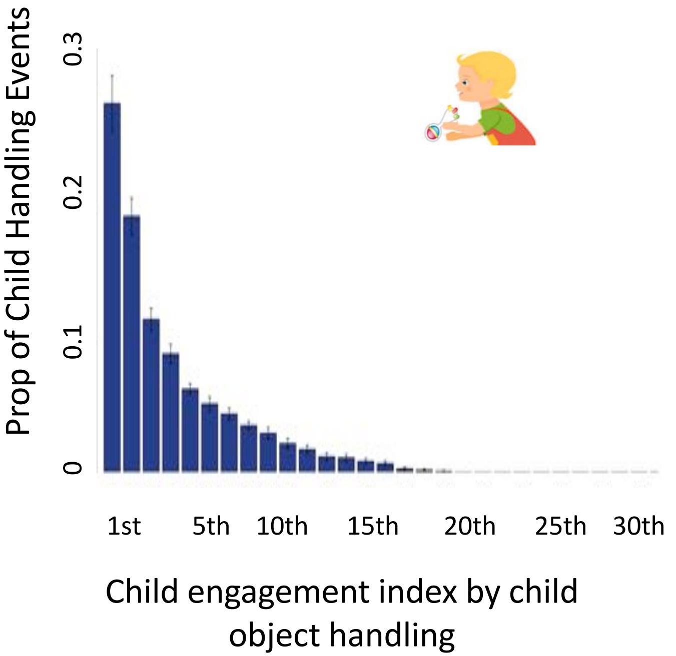
Are word-object co-occurrences really the right metric?



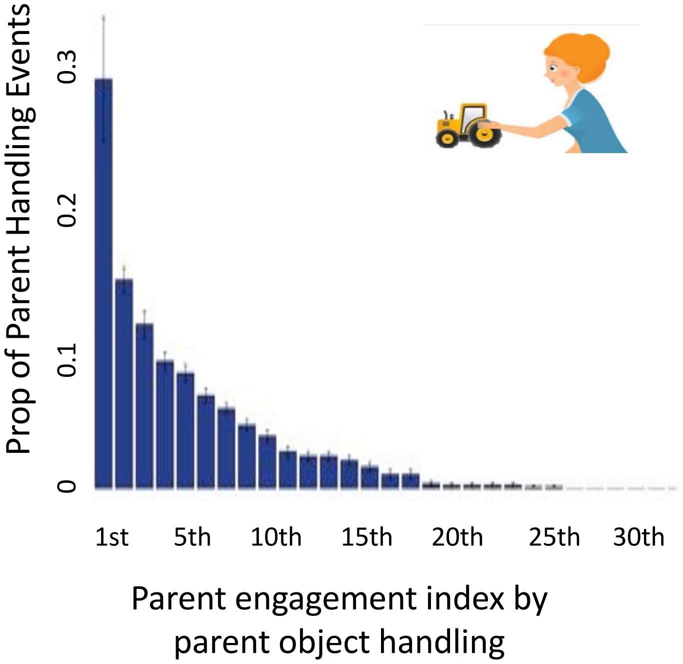
Raz & Smith, New!!!!

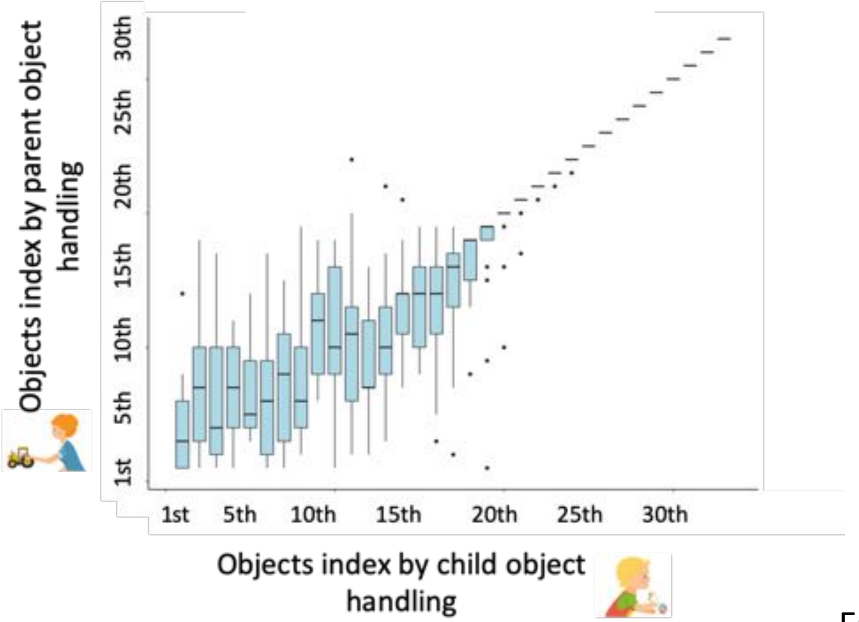
Parent and their 12 month old
10 minutes of play
33 objects on the floor

Prop of Child Manipulation
time per Object



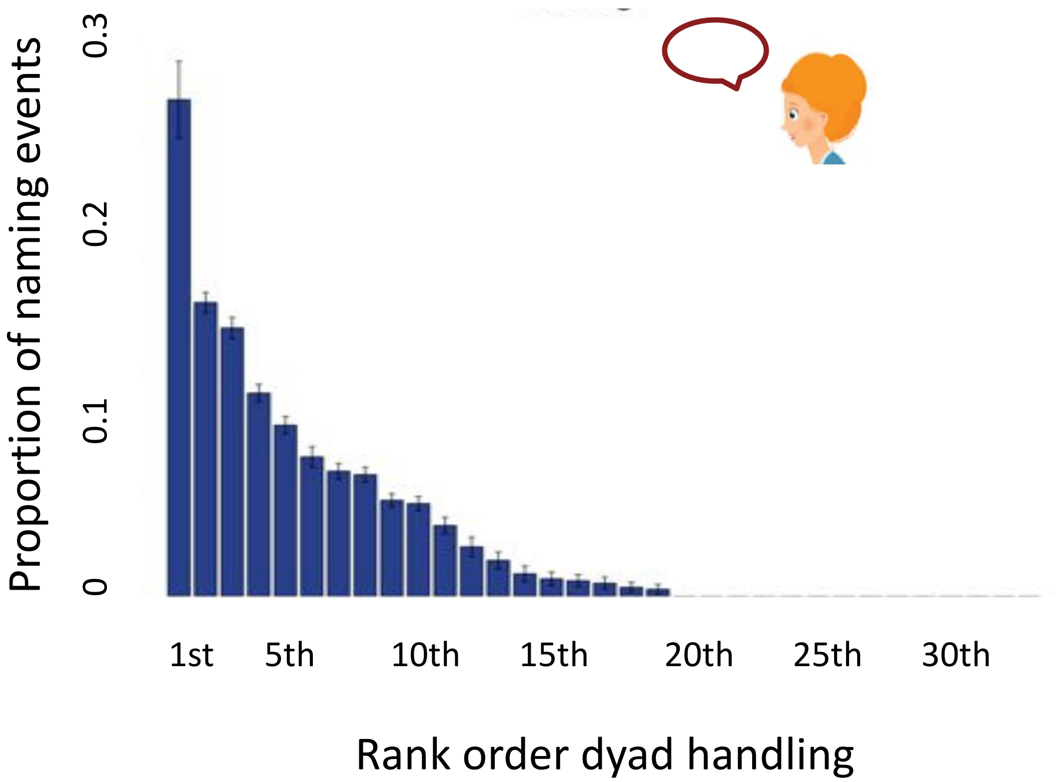
Prop of Parent Manipulation
times per Object





Each partner, mean
ranking of parent as function of
ranking by child in time
manipulating

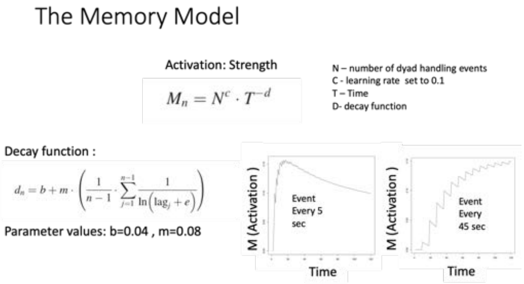
Naming –rare, about 5 -7 naming events in the whole 10 minutes of play



Why does this matter?

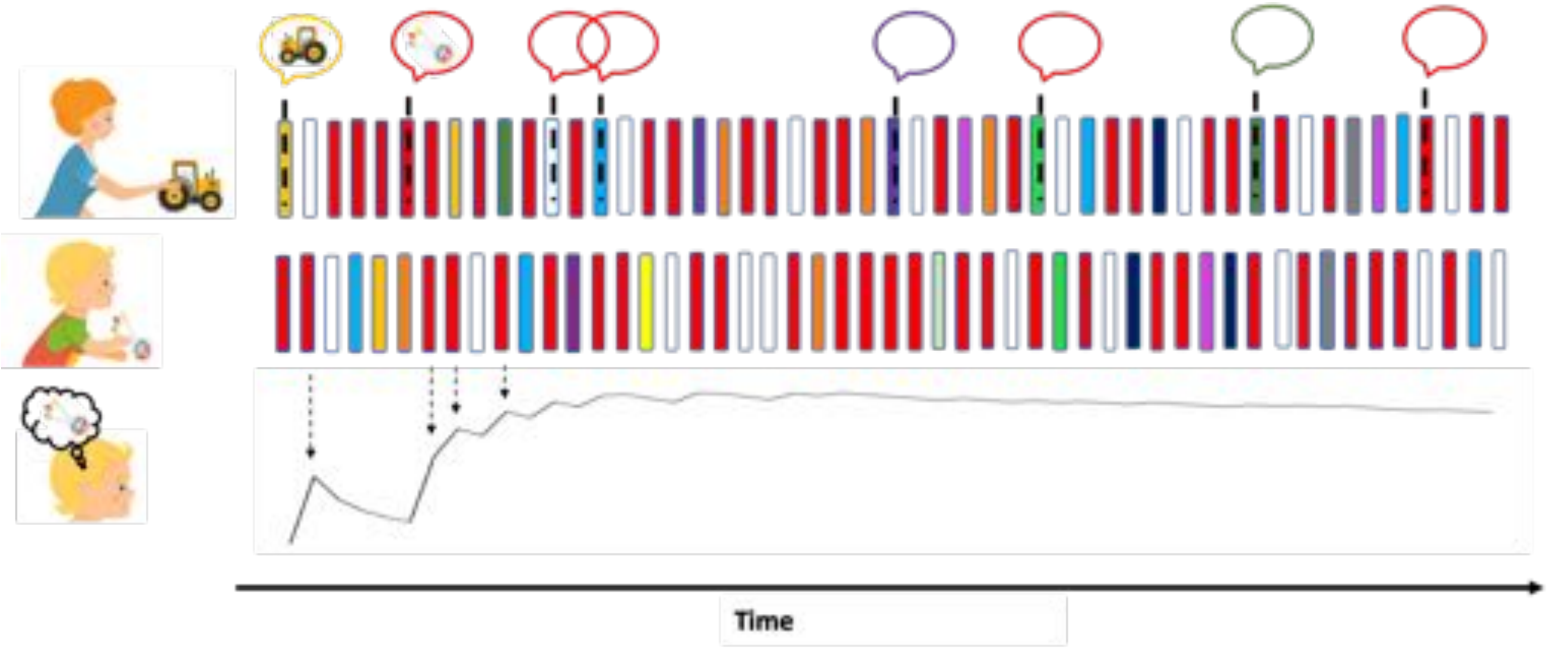
Ebbinghaus and modern versions:

Memory strength increases repetition and with longer spacing between repeating events.



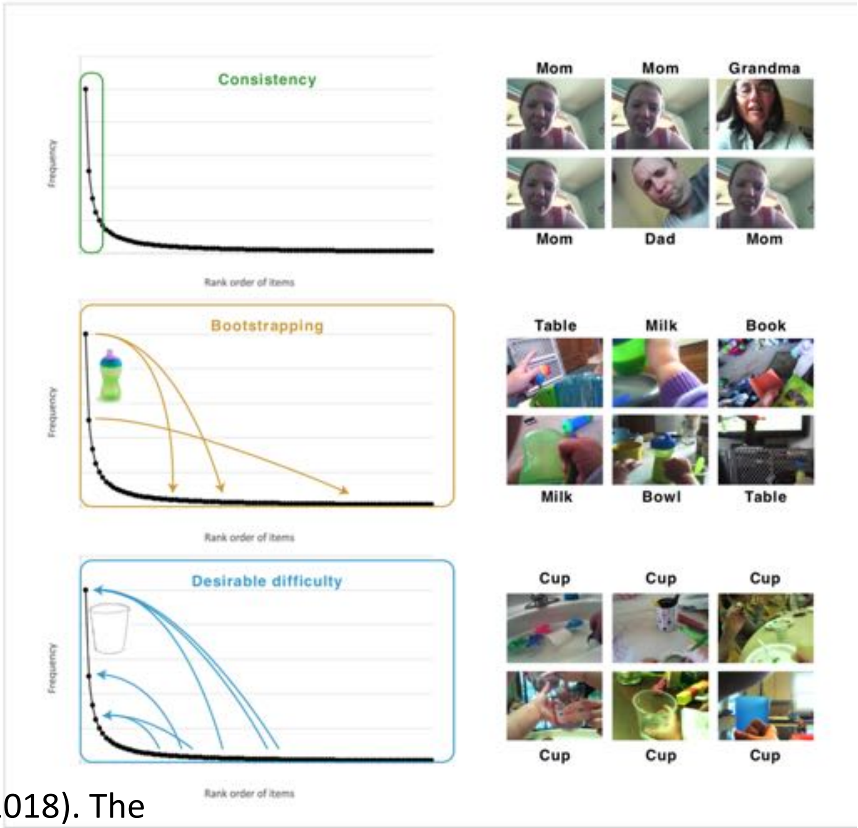
Walsh 2016

Beyond co-occurrence



Highlight 5: It's a right-skewed environment – for all contents, all time scales

Individual faces
Categories
Objects

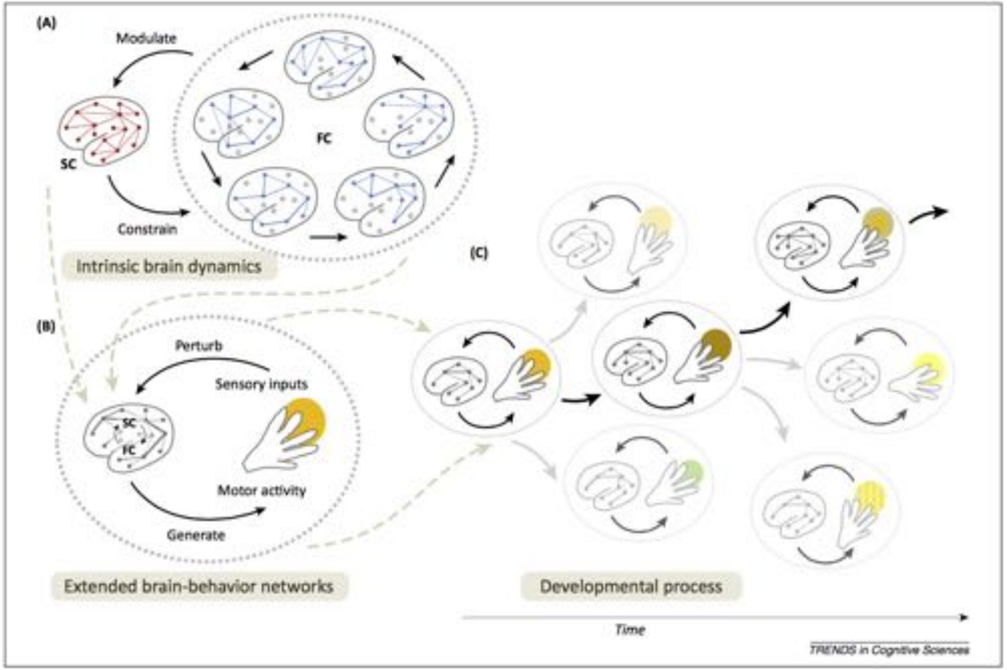


Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325-336.

Why the environment from the learner’s point of view







Byrge, L., Sporns, O. & Smith, L. B. (2014) Developmental process emerges from extended brain-body-behavior networks. *Trends in Cognitive Sciences*, 18(8), 395-403.

Collaborators

Chen Yu, David Crandall
Karin James, Jason Gold
Rowan Candy, Jim Regh (Georgia Tech)
Maithilee Kunda (Vanderbilt)

Former students and post docs

Alfredo Pereira (U. of Minho)
Caitlin Fausey (U. of Oregon)
Swapnaa Jayaraman (Western University)
Sumarga Suanda (U of Connecticut)
Daniel Yurovsky (Carnegie Mellon U)

Current students and post docs

Christian DeSerio, Elizabeth Clerkin
Hadar Raz, Lauren Slone
Catalina Suarez, Sven Bambach
Drew Abney, Violet Xiang

Lab Manager:
Charlotte Wozniak

And more!

Financial Support:
NSF, NIH, ASOFR
Indiana University (EAR):
Learning: Brains, machines & children