9.85 Cognition in Infancy and Early Childhood

Statistical reasoning in infancy

Welcome back ...

- Questions?
- Poster session Wednesday
- Office hours by appointment this week

Two intuitions about number

- “The knowledge of mathematical things is almost innate in us … This is the easiest of sciences, a fact which is obvious in that no one’s brain rejects it;for layman and people who are utterly illiterate know how to count and reckon.” (Roger Bacon, 1219-1294)

Two intuitions about number

- "It must have required many ages to discover that a brace of pheasants and a couple of days were both instances of the number two."(Russell, 1872-1970)

However ...

- You do not need a concept of number to be sensitive to statistical information in the environment ...
- What is statistical learning? (Aslin)
  - acquisition of structured information by passive observation
  - no feedback or reinforcement

Why is statistical learning important?

- Because much learning (most notably, language learning) appears to occur without feedback, reinforcement or instruction.
- Provides possible insight into the type of input to which infants attend.

Word segmentation **

Figure removed due to copyright restrictions. Source: Figure 1. Saffran, J. R. "Statistical language learning: Mechanisms and constraints." *Current Directions in Psychological Science* 12 (2003): 110-4.

Word parsing

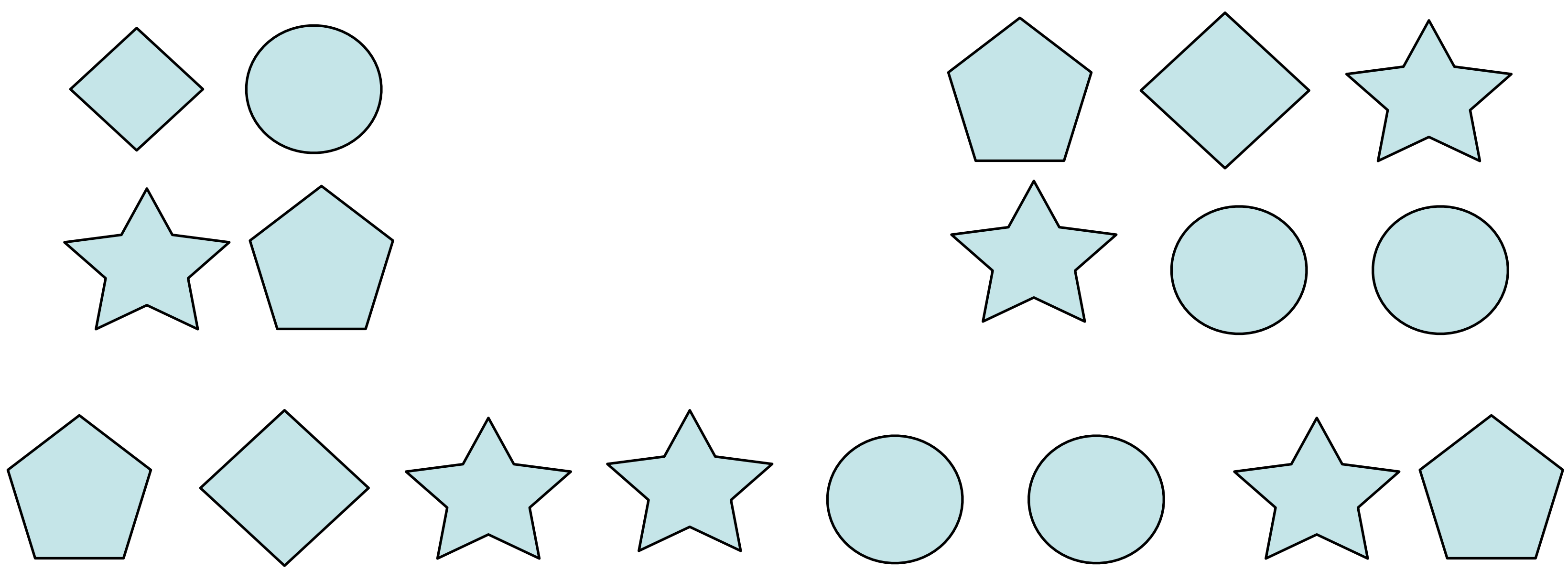
- theredonateakettleoftenchips
- The red on a tea kettle often chips
- There Don ate a kettle of ten chips

Word segmentation cues **

- Words in isolation
- Pauses/utterance boundaries
- Prosodic cues (e.g., word-initial stress in English)
- Correlations with objects/events in the environment
- Phonotactic/articulatory cues
- Statistical cues

Statistical learning simulation

- Make two 2-syllable words (each shape is a syllable).
- Now make two 3-syllable words.
- Write several sentences with your four words but put no spaces between the words.



Transitional probabilities **

PRETTY BABY

$$\frac{\text{(freq) } pretty}{\text{(freq) } pre} \quad .80$$

versus

$$\frac{\text{(freq) } tyba}{\text{(freq) } ty} \quad .0002$$

Statistical learning **

High likelihood High likelihood

PRE TTY BA BY

Low likelihood

Continuations *within* words are systematic
Continuations *between* words are arbitrary

Infants can use statistical cues to find word boundaries **

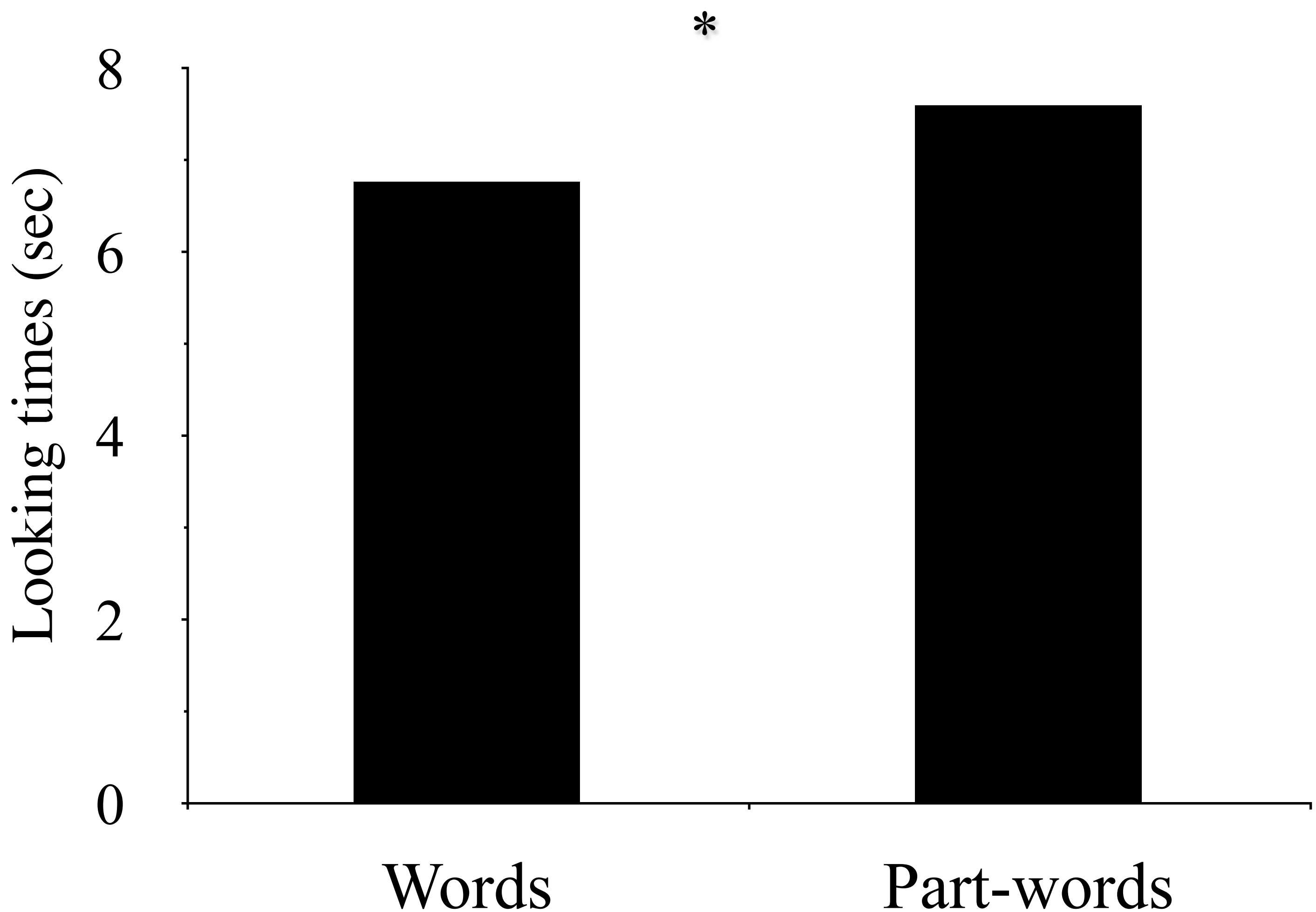
- Saffran, Aslin, & Newport (1996)
  - 2 minute exposure to a nonsense language (*tokibu*, *gopila*, *gikoba, tipolu)*
  - Only statistical cues to word boundaries
  - Tested on discrimination between words and part-words (sequences spanning word boundaries)

tokibugikobagopilatipolutokibu
gopilatipolutokibugikobagopila
gikobatokibugopilatipolugikoba
tipolugikobatipolugopilatipolu
tokibugopilatipolutokibugopila
tipolutokibugopilagikobatipolu
tokibugopilagikobatipolugikoba
tipolugikobatipolutokibugikoba
gopilatipolugikobatokibugopila **

tokibugikobagopilatipolutokibu
gopilatipolutokibugikobagopila
gikobatokibugopilatipolugikoba
tipolugikobatipolugopilatipolu
tokibugopilatipolutokibugopila
tipolutokibugopilagikobatipolu
tokibugopilagikobatipolugikoba
tipolugikobatipolutokibugikoba
gopilatipolugikobatokibugopila **

tokibugikobagopilatipolutokibu
gopilatipolutokibugikobagopila
gikobatokibugopilatipolugikoba
tipolugikobatipolugopilatipolu
tokibugopilatipolutokibugopila
tipolutokibugopilagikobatipolu
tokibugopilagikobatipolugikoba
tipolugikobatipolutokibugikoba
gopilatipolugikobatokibugopila **

Results **



So are babies tracking the transitional probabilities?

- What might be the problem with this interpretation?

tokibugikobagopilatipolutokibu
gopilatipolutokibugikobagopila
gikobatokibugopilatipolugikoba
tipolugikobatipolugopilatipolu
tokibugopilatipolutokibugopila
tipolutokibugopilagikobatipolu
tokibugopilagikobatipolugikoba
tipolugikobatipolutokibugikoba
gopilatipolugikobatokibugopila

toki**bugiko**bagopilatipolutokibu
gopilatipolutoki**bugiko**bagopila
gikobatokibugopilatipolugikoba
tipolugikobatipolugopilatipolu
tokibugopilatipolutokibugopila
tipolutokibugopilagikobatipolu
tokibugopilagikobatipolugikoba
tipolugikobatipolutoki**bugiko**ba
gopilatipolugikobatokibugopila

Controlling for frequency

- Words (tokibu, gikoba) occurred twice as often as part words (bugiko)

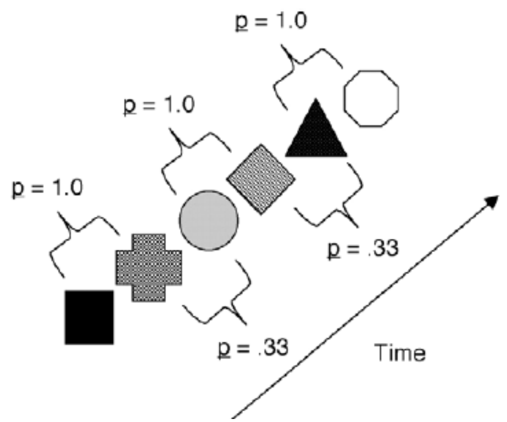
ABGHCDEFABGHABEFCDGH

part word BG as frequent as words CD and EF

- In other experiments (Aslin, Fiser), rare words were introduced to match the frequency of the part words.
- Thus only difference was in conditional probability (1.0 for words v. .5 for common part-words)
- 8-month-olds still distinguished part-words and words

Multimodal

- Is this kind of statistical learning domain general or specific to linguistic stimuli?
  - works for tones, not just syllables
  - same effects in non-human primates (although frequency controls have not been run)
  - same effects with visual stimuli



Also, what's happening here?

What do you see?

- “A baby playing.”
- “A baby’s stacking toy letters onto a bike, looking around and smiling.”

- But of course you didn’t “see” that. Even ignoring all the visual processing to get to objects and agents, you also saw
- A baby halfway through a reach.
- A baby turning.
- A letter in mid-air.

Meaningful units of action

- ‘If I am going for a walk to Hyde Park, there are any number of things that are happening in the course of my walk … So for example, I am also moving in the general direction of Patagonia, shaking the hair on my head up and down, wearing out my shoes and moving a lot of air molecules. However, none of these other descriptions seems to get at what is essential about this action, as the action it is.’ (Searle, Minds, Brains and Science)

Action parsing

- Level of representation -- meaningful intentional actions
- How do you get there?
- How do we “parse” action?

Action parsing

- Pause at the end of an “action”
- Pause in the middle of an action

Action parsing

- 10-month-olds dishabituated when the action was paused in the middle of a sequence but not when the action stopped at the end.

- Why?

Action parsing

- Top-down
- Use inferences about intentions to find meaningful units in action.
- Bottom-up
- Use low-level cues (changes in motion trajectories, eye gaze, **transitional probabilities**?) to parse action.

Statistical learning in action parsing

- **Habituate to: Stretch, shake, smell, knock, waggle, cap, head, stare**
- **Test to:**
- **shake, smell, cap, head**
- **Stretch, shake, smell, knock**

- **Looked longer at "part actions" than "whole actions"**

Okay but ...

- In all the examples so far, all the test stimuli were present during habituation.
- What about novel stimuli? Can infants extract general rules?

Rule-learning by 7-month-old infants (Marcus et al,. Science, 1999)

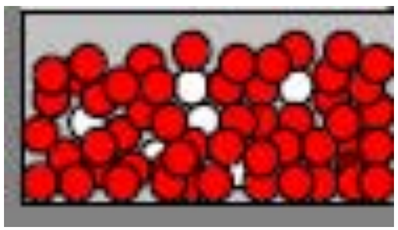
- ABA condition: ga ti ga, li na li, etc.
- ABB condition: ga ti ti, li na na

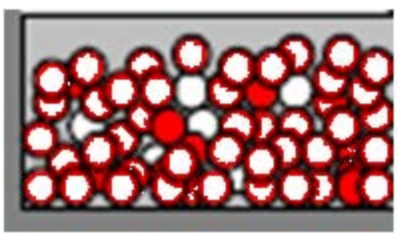
- Mean listening time was longer for novel inconsistent than consistent words (wo fe wo or wo fe fe)

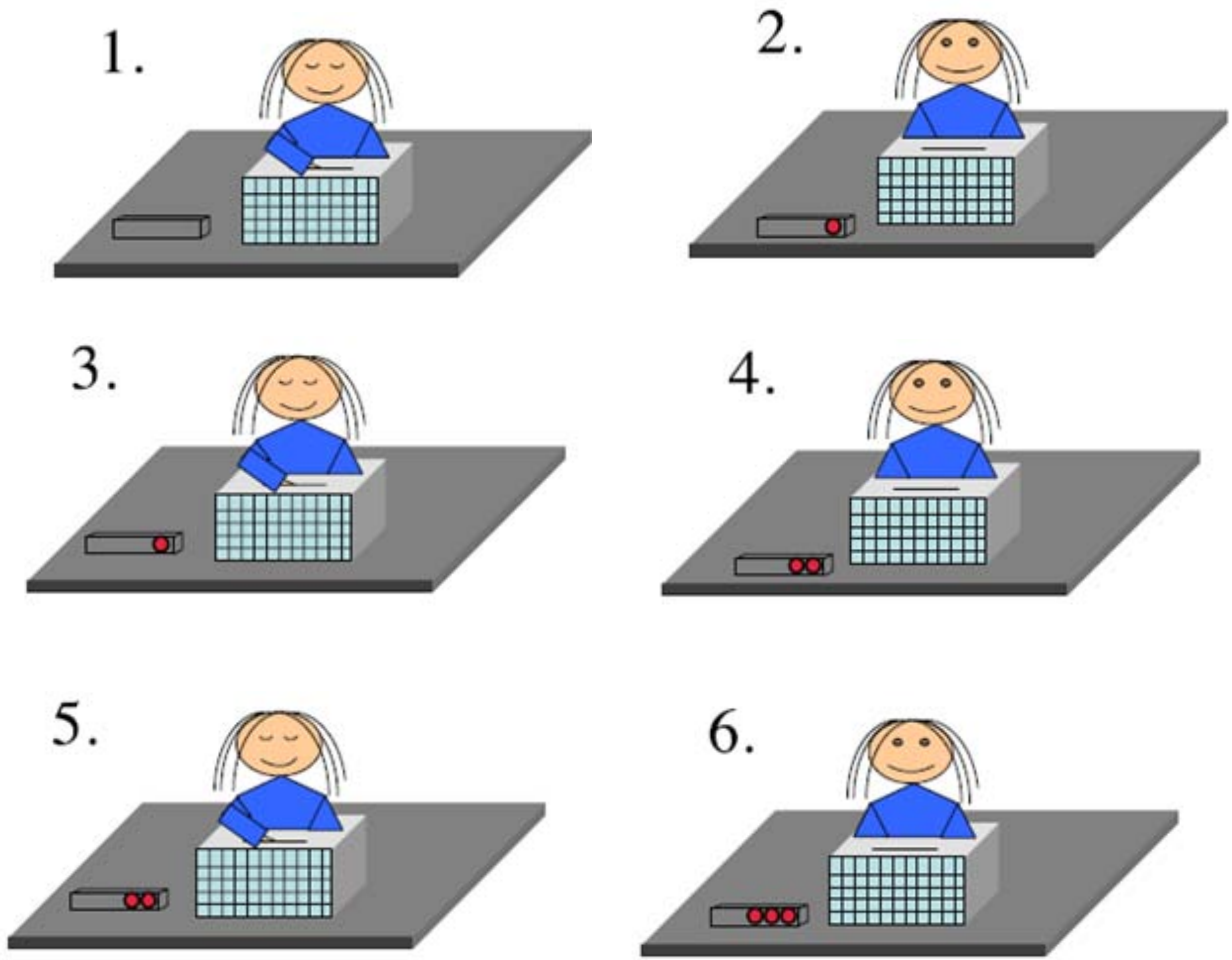
Statistical learning over categories more abstract than surface forms

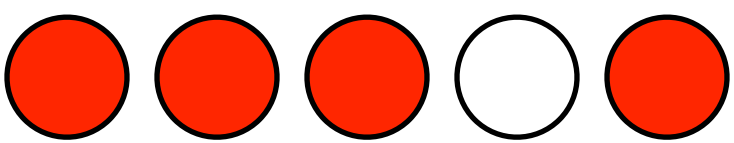
- Rule learning operates over categories more abstract than surface forms.
- And new research shows infants make even more sophisticated inferences ...

Xu & Garcia (2008) – 8 m.o.

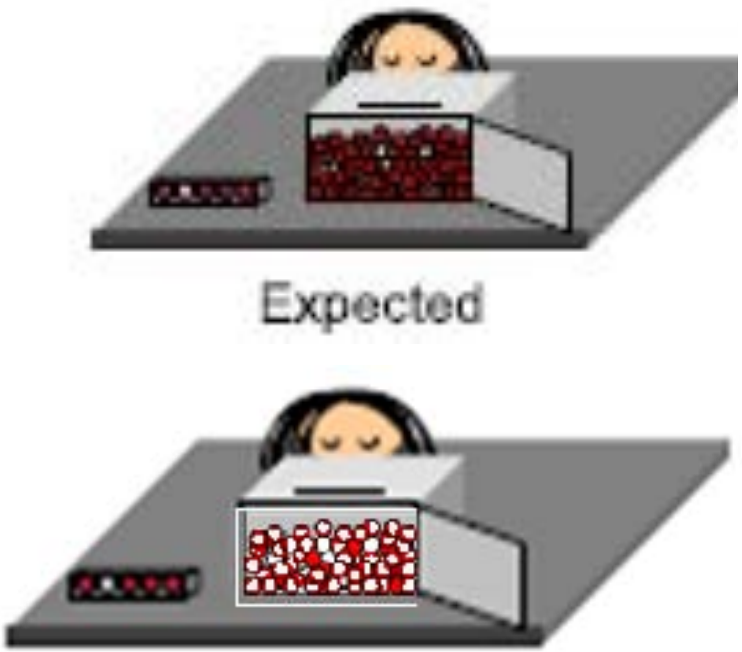








Test outcomes



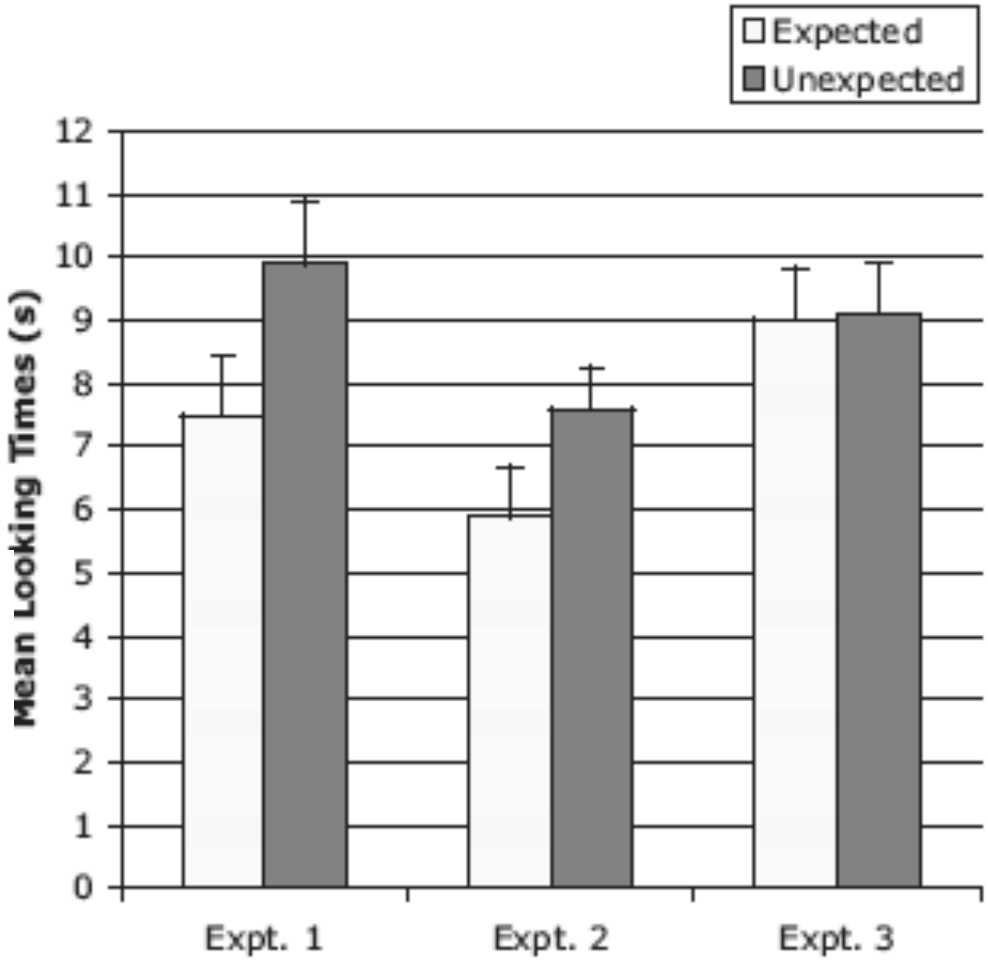
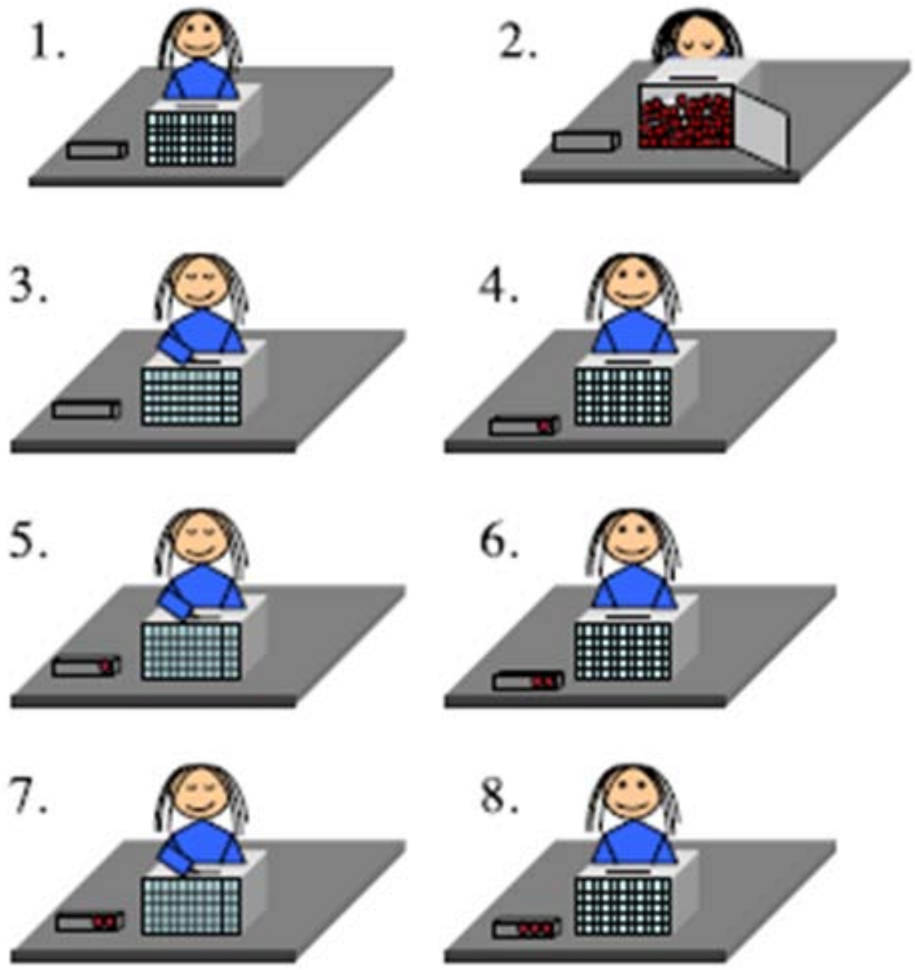
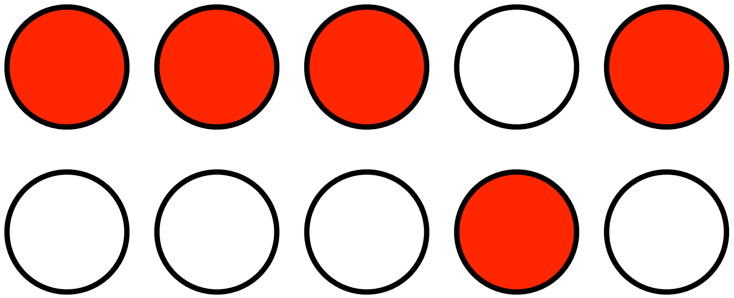


Fig. 2. Mean looking times for Exps. 1–3 with standard error.

Xu & Garcia (2008)





Xu & Garcia (2008)

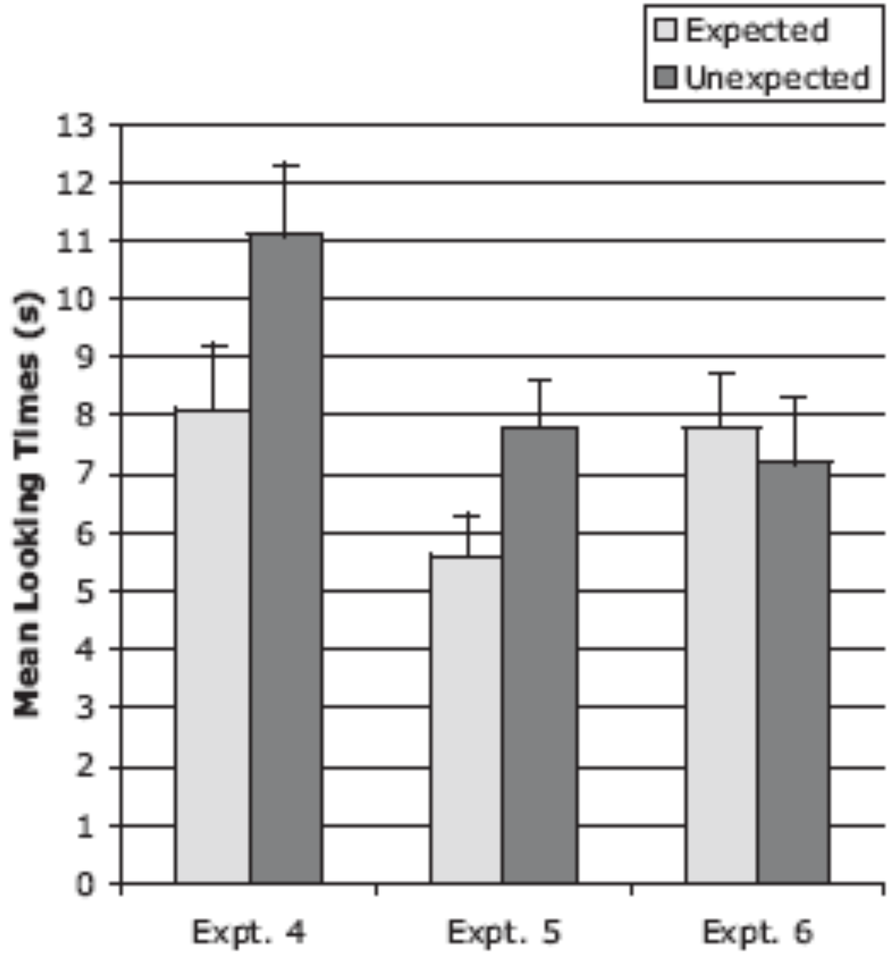


Fig. 4. Mean looking times for Exps. 4–6 with standard error.

Xu & Garcia (2008)

Babies as Intuitive Statisticians

- Babies have an intuition about the relationship between a **sample** and its **population**

- They expect a sample randomly drawn from the whole box to be representative of the population (and vice versa)

- Why should infants care about whether or not evidence is representative of a population?

Generalizing from samples

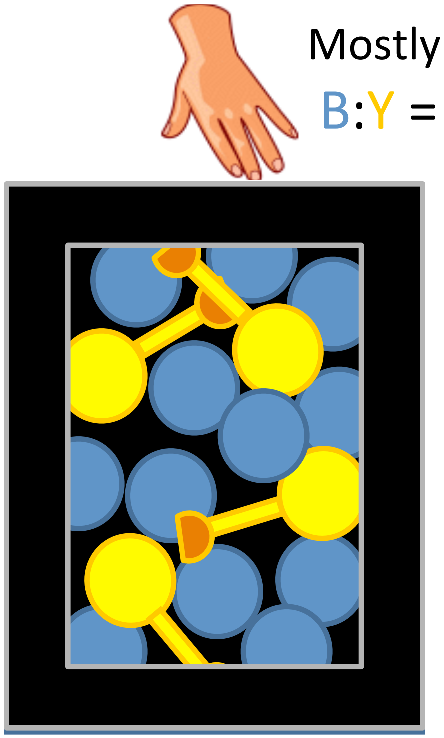
- Science requires generalizing properties from a small sample to a population.
- Can use feature similarity and category membership to infer that things that look alike or belong to the same category will share properties.
- If you know that this sample of Martian rocks has a high concentration of silica, may infer that other Martian rocks have a high concentration of silica.
- If you know that this sample of needles from a Pacific silver fir lie flat on the branch, may infer other Pacific silver fir needles lie flat on the branch.

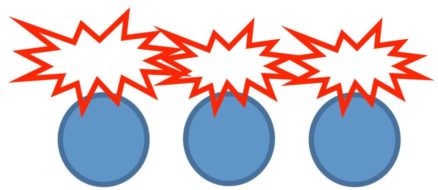
Generalizing from samples

- But as scientists we may know something about the sampling process that affects our inferences.
  - Do all Martian rocks have high concentrations of silica or only dusty rocks on the surface?
  - Do all Pacific silver fir needles lie flat or just those low on the canopy?
- How far we extend our generalizations depends on whether we think the sampling process was random or selective.
- Do infants' generalizations also take the sampling process into account?

Mostly Blue

B:Y = 3:1





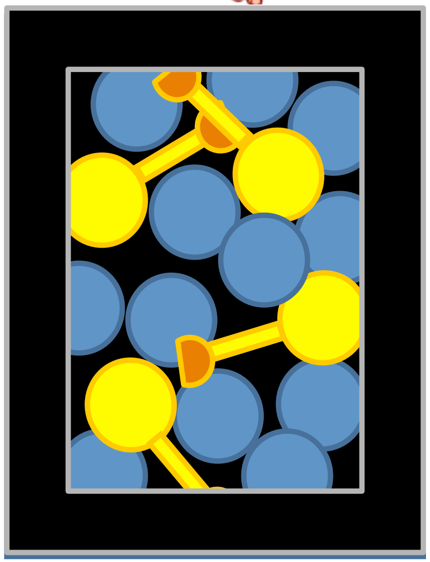
Consistent with sampling from the whole box

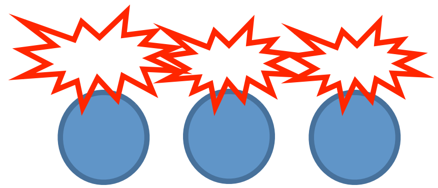
Looks a lot like others – should try squeaking it!

Gweon, H., Tenenbaum, J. B., et al. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." *Proceedings of the National Academy of Sciences* 107, no. 20 (2010): 9066-71.

Mostly Blue

B:Y = 3:1





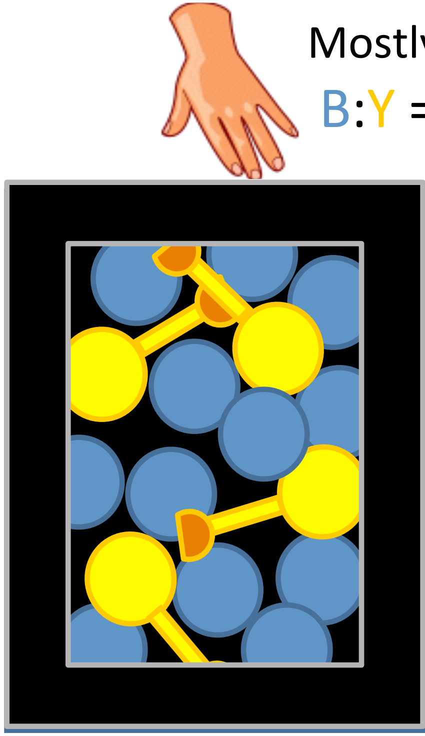
Consistent with sampling from the whole box

Prediction:
(1) many children should try squeezing
(2) and should squeeze often

Gweon, H., Tenenbaum, J. B., et al. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." Proceedings of the National Academy of Sciences 107, no. 20 (2010): 9066-71.

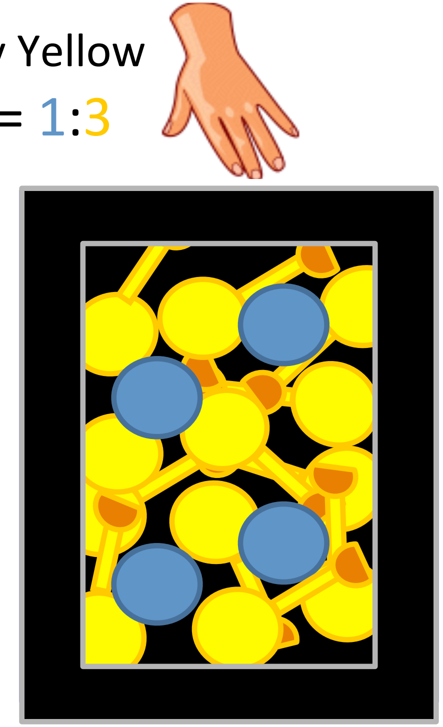
Mostly Blue

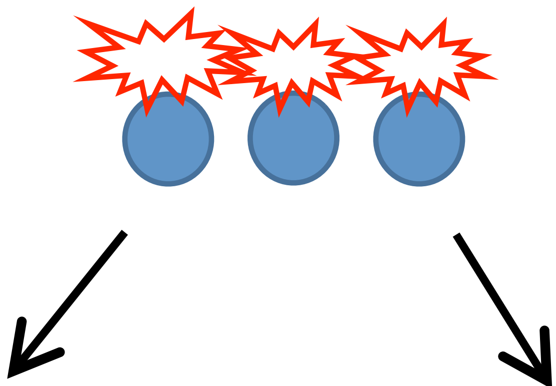
B:Y = 3:1



Mostly Yellow

B:Y = 1:3





Consistent with sampling from the whole box

Unlikely to have been sampled from the whole box

more likely to have been sampled ***selectively***

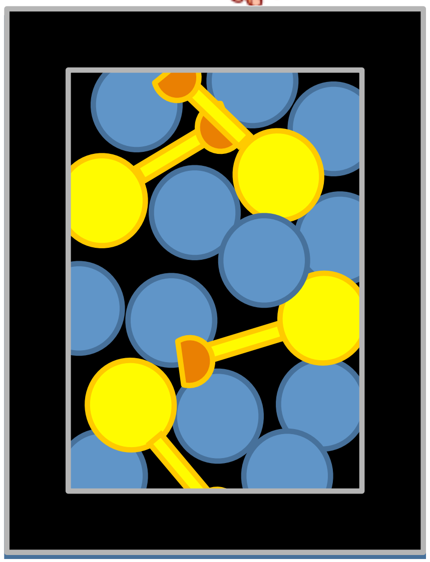
Prediction:
(1) many children should try squeezing
(2) and should squeeze often

The yellow one probably doesn't squeak

Gweon, H., Tenenbaum, J. B., et al. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." *Proceedings of the National Academy of Sciences* 107, no. 20 (2010): 9066-71.

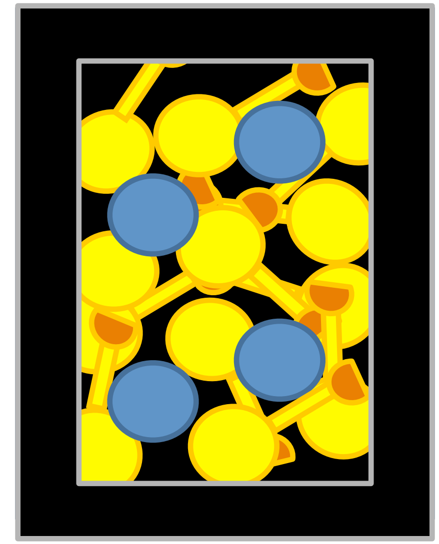
Mostly Blue

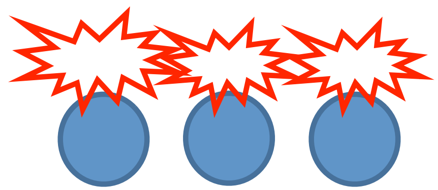
B:Y = 3:1



Mostly Yellow

B:Y = 1:3





Consistent with sampling from the whole box

Prediction:
(1) many children should try squeezing
(2) and should squeeze often

Unlikely to have been sampled from the whole box

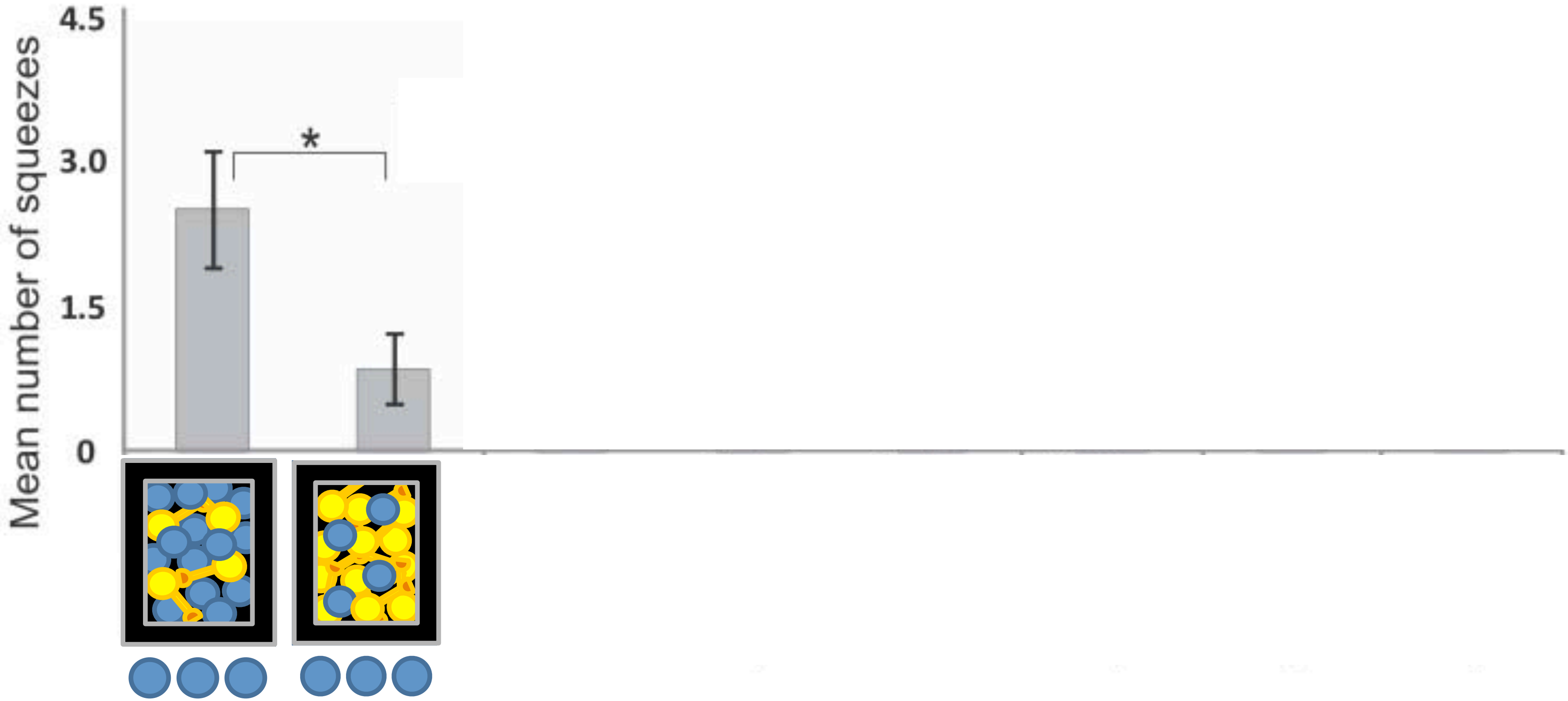
more likely to have been sampled ***selectively***

Prediction:
(1) fewer children try squeezing
(2) squeeze less often

Gweon, H., Tenenbaum, J. B., et al. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." *Proceedings of the National Academy of Sciences* 107, no. 20 (2010): 9066-71.

Results

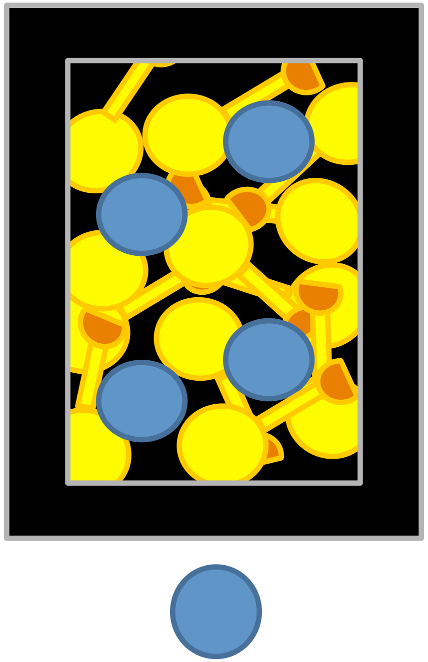
n = 15/condition, mean: 15 months, 15 days, range 13-18 months



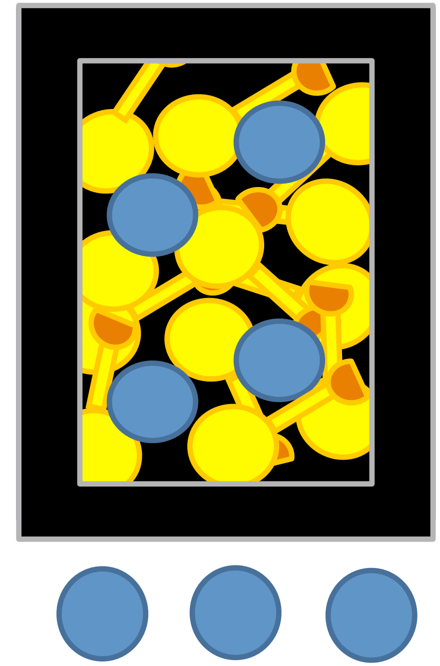
*$p < 0.05$

Gweon, Tenenbaum, & Schulz, 2010 *PNAS*

Gweon, H., Tenenbaum, J. B., et al. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." Proceedings of the National Academy of Sciences 107, no. 20 (2010): 9066-71.



Squeeze
once
squeeze x



..not an improbable sample. Coul hav been generated by samplin randomly from the whole box.

Unlikely to have been sampled from the whole box
more likely to have been sampled ***selectively***

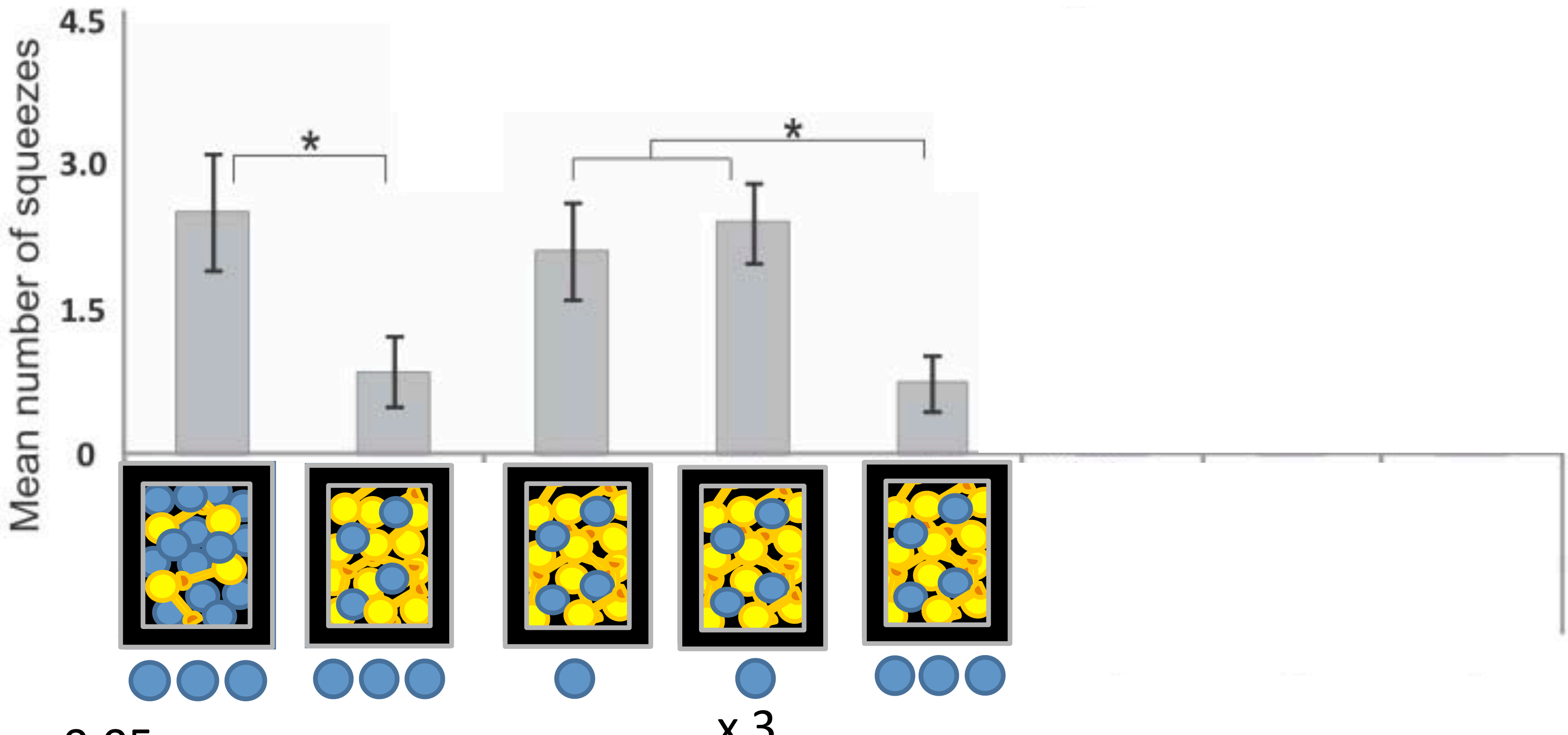
Prediction:
(1) many children should try squeezing
(2) and should squeeze often

Prediction:
(1) few children try squeezing
(2) squeeze less often

Gweon, H., Tenenbaum, J. B., et al. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." Proceedings of the National Academy of Sciences 107, no. 20 (2010): 9066-71.

Results

n = 16/condition, mean: 15 months, 15 days, range 13-18 months



*$p < 0.05$

Gweon, Tenenbaum, & Schulz, 2010 *PNAS*

Gweon, H., Tenenbaum, J. B., et al. "Infants Consider Both the Sample and the Sampling Process in Inductive Generalization." *Proceedings of the National Academy of Sciences* 107, no. 20 (2010): 9066-71.

Generalizing from samples

- 15-month-olds' generalizations take into account more than category membership and the perceptual similarity of objects.

- Infants make graded inferences that are sensitive to both the amount of evidence they observe and the process by which the evidence is sampled.

- And by preschool, inferences based information about the relationship between samples and populations support inferences about agent mental states, like preferences ...

Inferring Preferences from Sampling Information

Kushnir, Xu, & Wellman (2008)

- Preschoolers (3 - 4yr olds)
- 3 conditions

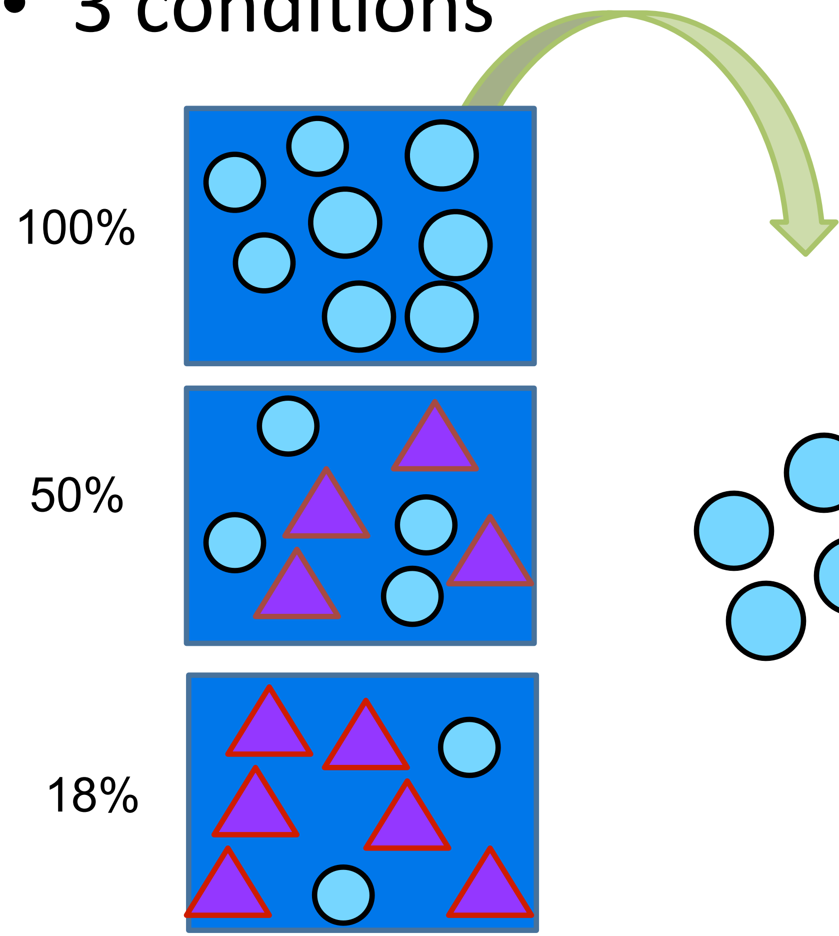
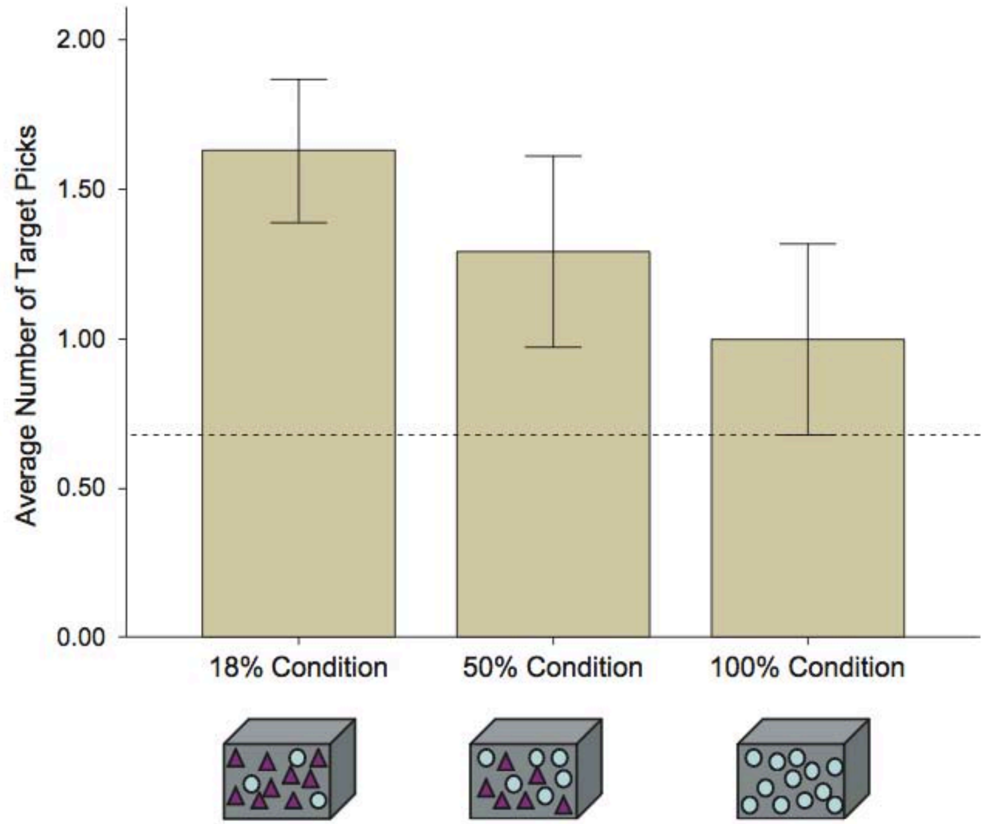


Image: OpenClipArt. Public Domain.

Results

- Violation of random sampling (selective sampling) as evidence for preference
- Make inferences about social/psychological phenomena from statistical information.



Statistical learning over single events

Objects are spinning around in the “lottery machine”. The machine is occluded and a single object drops out. Which shape will it be?

Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference

Ernő Téglás, Edward Vul, Vittorio Girotto, Michel Gonzalez, Joshua B. Tenenbaum, and Luca L. Bonatti

Length of time the lottery machine is occluded

Figure 3E. Téglás, Ernő, Edward Vul, et al. "Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference." *Science* 332, no. 6033 (2011): 1054-9.

B) immediately after occlusion, infants are surprised if objects far from the exit out, not if near objects fall.

C) 1 s after occlusion, infants consider both how far the object is from the exit and how many of them there are

D) 2 s after occlusion, infants disregard how far the object is from the exit and consider only how many of them there are

Figure 2. Téglás, Ernő, Edward Vul, et al. "Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference." *Science* 332, no. 6033 (2011): 1054-9.

Single event probabilities plus folk physics

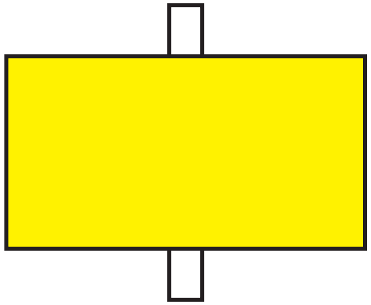
- Astonishing correlation between infants' looking time and the predictions of the computational model. (Previously we only ever tried to predict "longer looking" not graded inferences.)

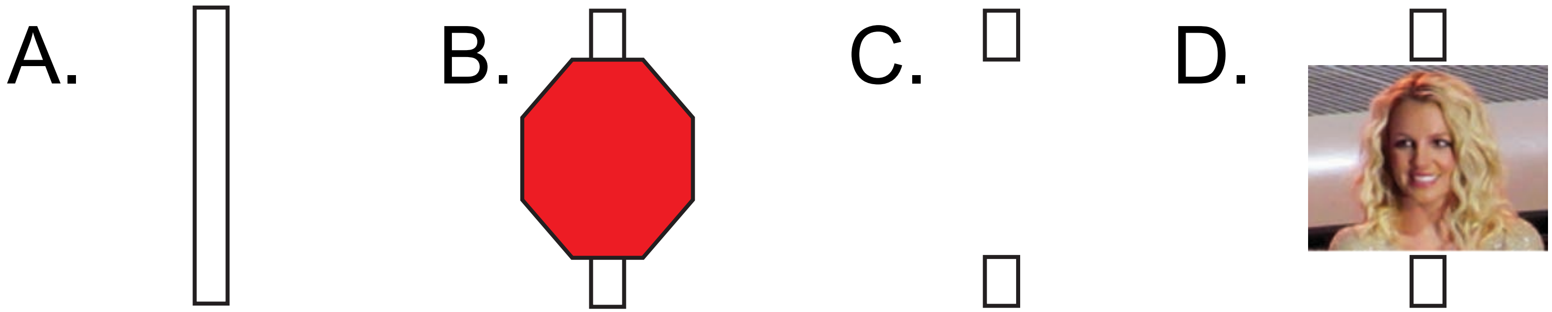
Big riddle of induction

- We need constraints to learn from data (because infinitely many hypotheses are consistent with sparse data)

What gap? Try this multiple choice quiz

- What's behind the rectangle?





Big riddle of induction

- We need constraints to learn from data (because infinitely many hypotheses are consistent with sparse data)
- But where do the constraints come from?
- Some are plausibly innate (i.e., object knowledge) but what about everything else?

Chicken and egg problem

- How do you use data to draw abstract inferences but have abstract inferences that constrain your interpretation of the data?
- Answer: the intuition that children get a lot of specific concrete examples and gradually abstract from them may be wrong.
- Instead, a few examples may suffice for children to make a very abstract inference, that constrains further induction.
- In some cases, more abstract inferences may be easier than more specific ones.

Figure removed due to copyright restrictions. Fig. 1. Schematic representation of the test events in the experimental and control conditions of Experiment 1. Dewar, Kathryn M. and Fei Xu. "Induction, Overhypothesis, and the Origin of Abstract Knowledge Evidence From 9-Month-Old Infants." *Psychological Science* 21, no. 12 (2010): 1871-7.

Powerful statistical learning

- Drawing rich, abstract inferences from sparse data provides a mechanism that can get rapid learning off the ground.
- Not all constraints on induction need to be innate ... some can be learned.
- The “blessing of abstraction”