24.949
Language Acquisition

Class 2
Words

“Words”

- Arbitrary combinations of form (e.g. sound in spoken language) and meaning, $\langle \phi, \mu \rangle$.

- not predictable, hence need to be learned and stored in memory (⇒ listemes)

- learning task: for any given listeme w_i ($\langle \phi_{wi}, \mu_{wi} \rangle$)

  - What are the properties of ϕ_{wi} (boundaries, length, phonemes, co-occurrence constraints, stress, ...)

  - What are the properties of μ_{wi} (semantic features/concept, conditions on use, ...).

Challenges of word learning

1. Finding words in the speech stream

2. Associating words with meaning

Challenges of word learning

1. Finding words in the speech stream

    - don’t have time to cover
    - relevant readings: Saffran et al. 1999, Gout et al. 2004, Shukla et al. 2011

2. Associating words with meaning

    - today’s topic

Better defining our problem space

- How children learn words that go with medium sized everyday objects and some other stuff

- Not much that we will talk about today will tell us how the child learns the meanings of words like “the” or “most” or “even”

Better defining our problem space

- How children learn prosodic words that correspond somewhat closely to “roots” (borrowing DM parlance) in isolating languages

A note on learning

- What does it mean to *learn* a word?

Learning as hypothesis testing

- Jerry Fodor's point
  - You are shown cards, some of which are *miv* and some *non-miv*.
  - You are rewarded when you correctly identify the *miv* cards and punished when you fail

Learning as hypothesis testing

- Jerry Fodor's point
  - What you do, according to classical learning theories: consider a hypothesis of the form "X is *miv* iff X is …"
  - The data is then used to assess the truth of the hypotheses with various values of "…".
    - So if *miv* means "red and round" then the data will tend to confirm "X is *miv* iff X is red and round" and disconfirm everything else.

Learning as hypothesis testing

- Given a hypothesis space, the data (environmental input) can be used to assign a number (a probability) of how well that hypothesis fits the data.

- But where do the hypotheses that are tested come from? i.e. where do the fillers of "…" come from?

- Inductive theories of learning **presuppose the hypothesis space**

The problem of identifying word meanings

Word meanings

- What kind of objects are word meanings?
- Let’s try: What is the meaning of *chair?*
  - *an item of furniture*

Word meanings

- What kind of objects are word meanings?
- Let’s try: What is the meaning of *chair?*
  - Collins Pocket English Dictionary: *A seat with a back and four legs, for one person to sit on*

Word meanings

- What kind of objects are word meanings?
- Let's try: What is the meaning of *chair?*
  - Collins Pocket English Dictionary: *A seat with a back and four legs, for one person to sit on*
  - The Oxford English Dictionary (2nd edition, 1989): *A seat for one person (always implying more or less comfort and ease); now the common name for the movable four-legged seat with a rest for the back, which constitutes, in many forms of rudeness or elegance, an ordinary article of household furniture, and is also used in gardens or wherever it is usual to sit*

Word meanings

- What kind of objects are word meanings?
- Let’s try: What is the meaning of *chair?*

  - an architect proposes a dining room in which the chairs are sculpted from the stone that forms the floor. are these *chairs*?
  - in society *x*, all chairs are systematically designed for two people and have 7 legs. are these *chairs*?

discussion adapted from Elbourne 2005

Word meanings

- What kind of objects are word meanings?
  - *extensions* and *intensions*
    - extension of chair: the set of all chairs
    - intension of chair: CHAIR
- Whatever the meaning of *chair*, it lets us talk not only about actual chairs we encounter in the world, but also about chairs that are merely possible

Other candidates

- Meanings as definitions (Jackendoff 1996)
- Prototype theory (Rosch 1975)
- System of association among (copies of) sensory representations (Hume 1739, Prinz 2002)

Word meanings

- How do we learn word meanings?

A simple story: associative learning

- How do we learn word meanings?

“It looks simple. A 14-month-old toddles after the family dog, smacking it whenever she gets close. The dog wearily moves under the table. ‘Dog,’ the child’s mother tells her. ‘You’re chasing the dog. That’s the dog.’ The child stops, points a pudgy hand the dog, and shrieks ‘Daw!’ The mother smiles: ‘Yes, dog.’”

Bloom 2000

Learning from observation

- Most people agree that at least some of the time, you learn word meanings by observing the world

- How does an infant mind do that? What ‘start-up package’ (biases, learning strategies) is it equipped with?

POS problem

- If word meanings are intensions/concepts, we have a POS problem

- How do you go from observing a perceptual slice of the world (e.g. a referent associated with the word) to something qualitatively different?

POS problem

- Now if word meanings are equivalence classes of entities grounded in perceptual similarity, the POS problem does not arise

- The task instead is to identify which perceptual features are **relevantly** similar (not all are) and where this knowledge comes from

The problem of referential uncertainty

- Quine (1960):

The problem of referential uncertainty

- Quine (1960): the problem of referential uncertainty
  - [[gavagai]] =
    - Rabbit?
    - Rabbit ear?
    - Dinner?
    - Something cute?
    - An animal?
    - A thing that hops?
    - Disconnected rabbit parts?
    - A rabbit but only to the year 2021, then carrots?

The problem of referential uncertainty

- The same problem that the child faces

- Even if something is explicitly labeled in their input (“Look! There’s a doggy!”), how does the child know what specifically that word refers to?

- An infinite number of hypotheses about word meaning is possible. **Input under-specifies the word’s meaning.**

- Yet children figure it out. They have mapped meanings onto some words as early as 6mos of age (Bergelson and Swingley 2012)

One solution

- “As for the Gavagai problem, it seems to me that a lot of it is pragmatic. [H]umans can have a notion of “informativeness”, or “utility” for a given word in context, such that, in a conversation, each participant has an idea of what is the right level of informativeness, even cross-linguistically. So, if a speaker of an unfamiliar language shows me a Gavagai, the likelihood that it will indeed denote an animal and not something extremely complex or specific, is relatively high. This is because my informant knows I am trying to translate words in his language, and wants to be cooperative… I think that humans can share a “scale of specificity” for words and their referents…"

One solution: learning biases

- Innate biases that help constrain hypothesis space
  - mutual exclusivity, newness, ...
  - basic level categories, whole objects as opposed to parts, properties or relations, ...
  - shape (for artifacts), color, texture and shape (for animals), ...

One solution: cross-situational learning

The basic idea:

- Quine’s problem is a problem only if learning must happen from one instance.

- If infants can accrue statistical evidence across multiple referential situations, they can home in on the right target

One solution: cross-situational learning

Utterance situation 1: “ball!”

One solution: cross-situational learning

Utterance situation 2: “ball!”

One solution: cross-situational learning

Utterance situation 3: “ball!”

One solution: cross-situational learning

Utterance situation 4: “ball!”

One solution: cross-situational learning

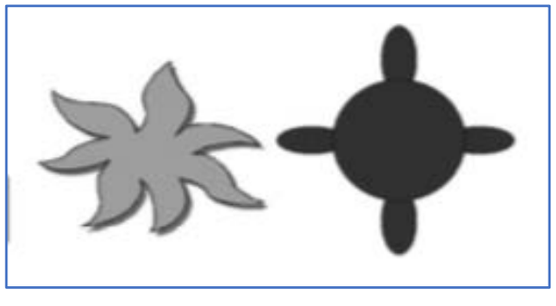
- Only one hypothesis — that "ball" refers to  — is consistent with all of these situations.
- Can babies reason like this?

Situation	Objects				
	b	***c***	***d***	***s***	***t***
1	1	1			
2	1		1	1	
3	1			1	
4	1				1

Smith & Yu 2007

- Preferential Looking Paradigm w/ 12-to-14-mo olds
- Training:
  - 6 novel words (*bosa, gasser, manu, colat, kaki, regli*) each associated with a distinct brightly colored shape
  - 30 trials, with 2 objects named with 2 words

manu
colat

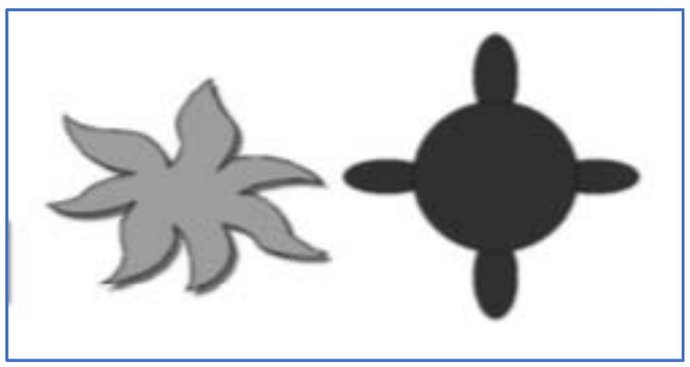


kaki
colat

Smith & Yu 2007

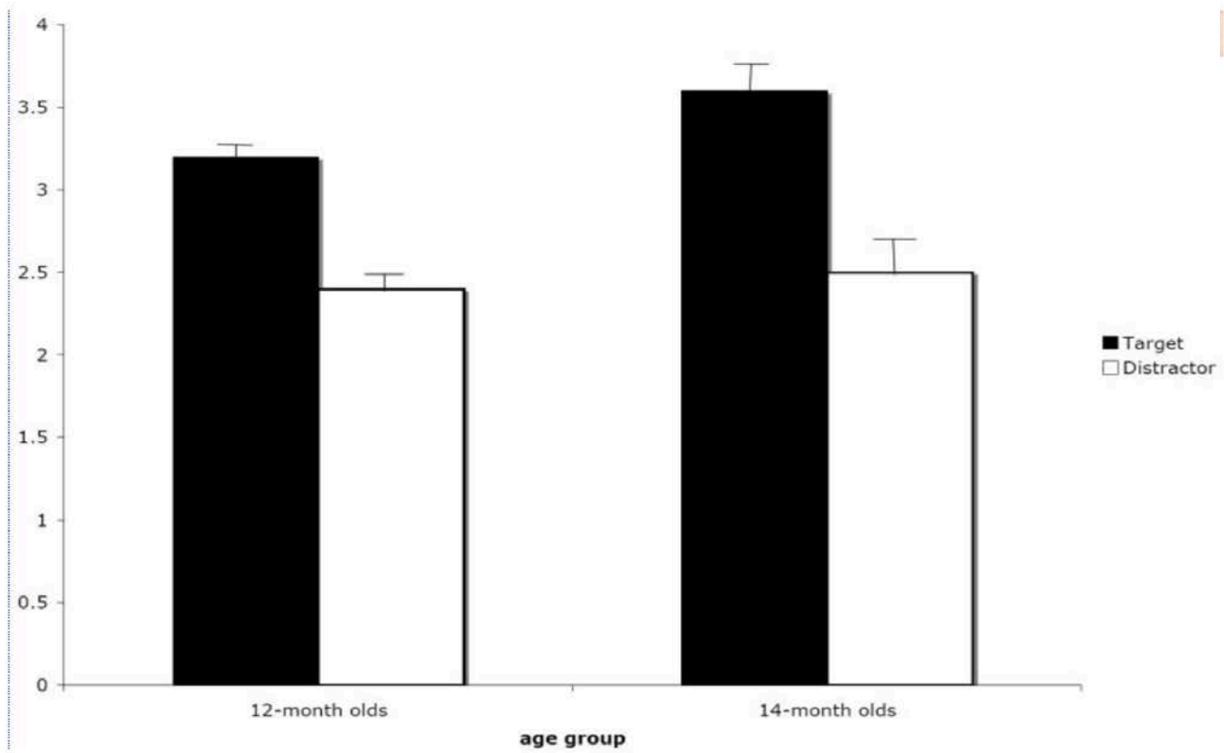
- Test:
  - Two objects; one word. Which object do they prefer to look at?
  - 12 trials (2x per target word)

colat



Smith & Yu 2007

- Results:
  - Infants preferentially looked at the target
  - Looking behavior varied by word: some word-referent associations better mapped than others



Mean looking time to target and distracter per 8 sec test trial (and standard error of the mean) for younger and older infants.

Smith & Yu 2007

"In sum, these results tell us that cross-situational statistical learning is in the repertoire of young word learners. Despite the ambiguity of word-referent mappings on any individual training trial, infants clearly **accumulate information across trials and use that information to determine the underlying mappings**."

Conceptual issue

- If we take seriously our conjecture that word meanings are not sets of referents, POS problem remains

- One possibility: cross-situational learning = accruing evidence for/against a hypothesis across situations (Tenenbaum & Xu 1999)

Empirical issue

- One-shot learning
  - Carey and Bartlett (1978): “fast-mapping”
  - 3-year-olds were presented with two objects, one blue, one olive-green, and asked: “bring me the chromium one, not the blue one”
  - All of the children retrieved the olive tray, correctly inferring that the experimenter intended chromium to refer to this new color.
  - When tested a week later on their comprehension of the word, over half of the children remembered something about its meaning, either that it named olive or that it named a color that resembled olive.

Empirical issue

- Landau & Gleitman 1985: language development in blind children proceed in a strikingly similar manner to sighted children
  - onset of speech below median but within normal limits
  - initial word combination below median, but within normal limits
  - by age 3, indistinguishable from sighted children in lexical and syntactic complexity

- earliest expressed meanings very similar to those of sighted children
  - “blind children talk about what most young children talk about: mommies, daddies, dolls, cookies, and toys" (p.30)

- Infants can seemingly tabulate word-referent co-occurrence statistics over multiple instances, but as a theory of word learning, there are some obvious empirical gaps

- But we might still suppose that those are the exceptional cases and cross-situational statistics is still a major tool recruited in the learning process

- How do you test this?

Starting point

- On the cross-situational learning model, learners learn by keeping track of multiple hypotheses about a word's meaning across successive learning instances, and gradually converge on the correct meaning via an intersective process.

- **A key premise:** Listeners who appreciate the indeterminacy of observation should not jump to a conclusion about word meaning on first observation, but rather, hold the choices in abeyance until evidence from further exposures has accumulated

Question

- Can learners hold onto the relevant situational information until evidence is strong enough to form a hypothesis about word meaning?

Question

- Can learners hold onto the relevant situational information until evidence is strong enough to form a hypothesis about word meaning?
  - Adults fail to track information across multiple instances (Trueswell et al. 2013)
  - Children fail to track information across multiple instances (Woodard et al. 2016, Aravind et al. 2017)

Trueswell et al. 2013

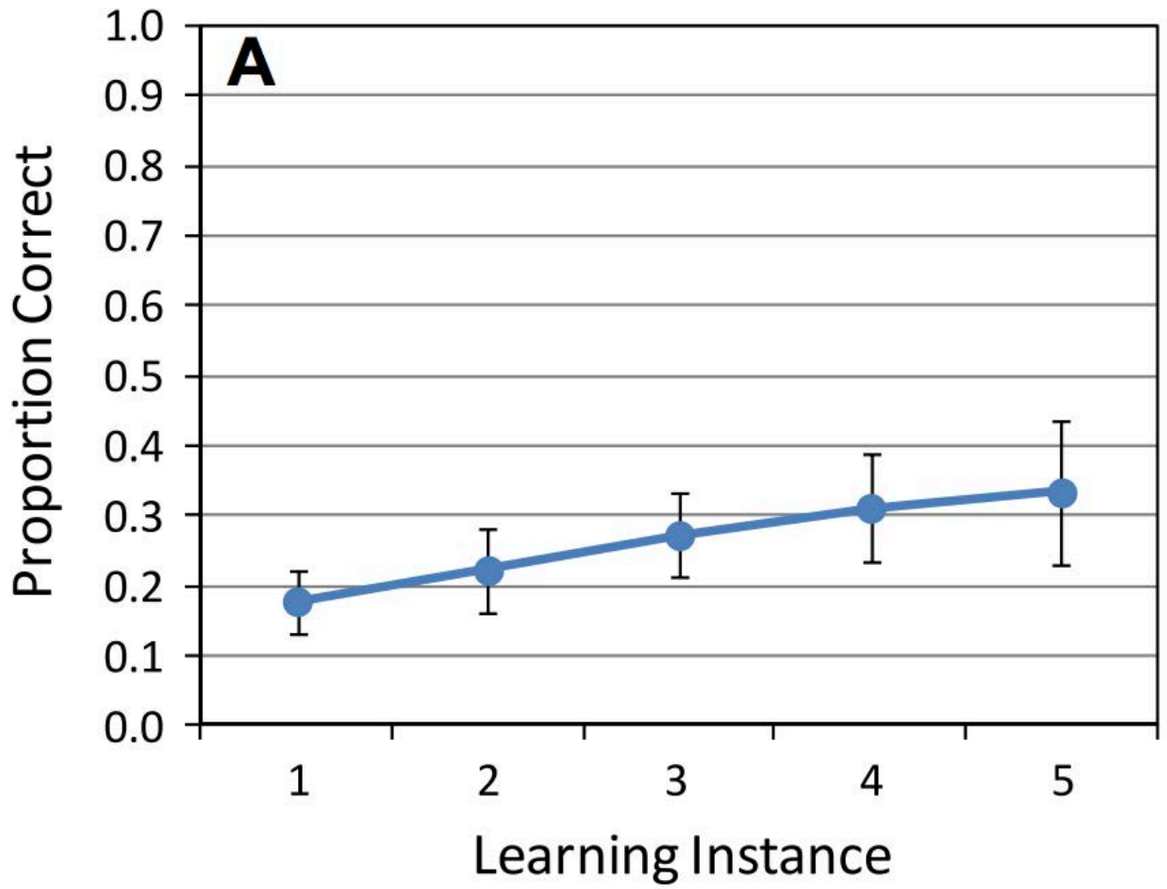
- Experiment 1
  - Presented adults with novel words used as names for familiar objects
  - Participants click on a hypothesized referent object
  - Each subsequent set replaces 4 of the 5 possible referents, in a way s.t. co-occurrence frequency between target and word (100%) is higher than any alternatives (max 40%)
  - 5 “learning instances” per word

Trueswell et al. 2013

- Measures
  - explicit: choice of referent
  - implicit: eye movements
- Question: does the rate of target-selection improve stepwise over time/across the 5 learning instances?

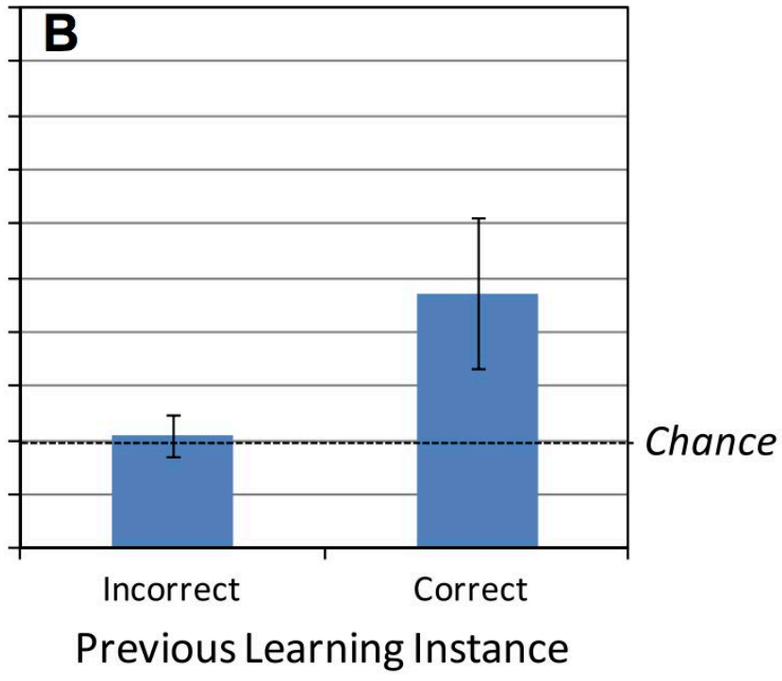
Trueswell et al. 2013

- Results
  - slow, but steady learning over time



Trueswell et al. 2013

- Results
  - accuracy-contingent learning
    - success on any given trial modulated by success on preceding one
  - confirmed by eye-movement patterns

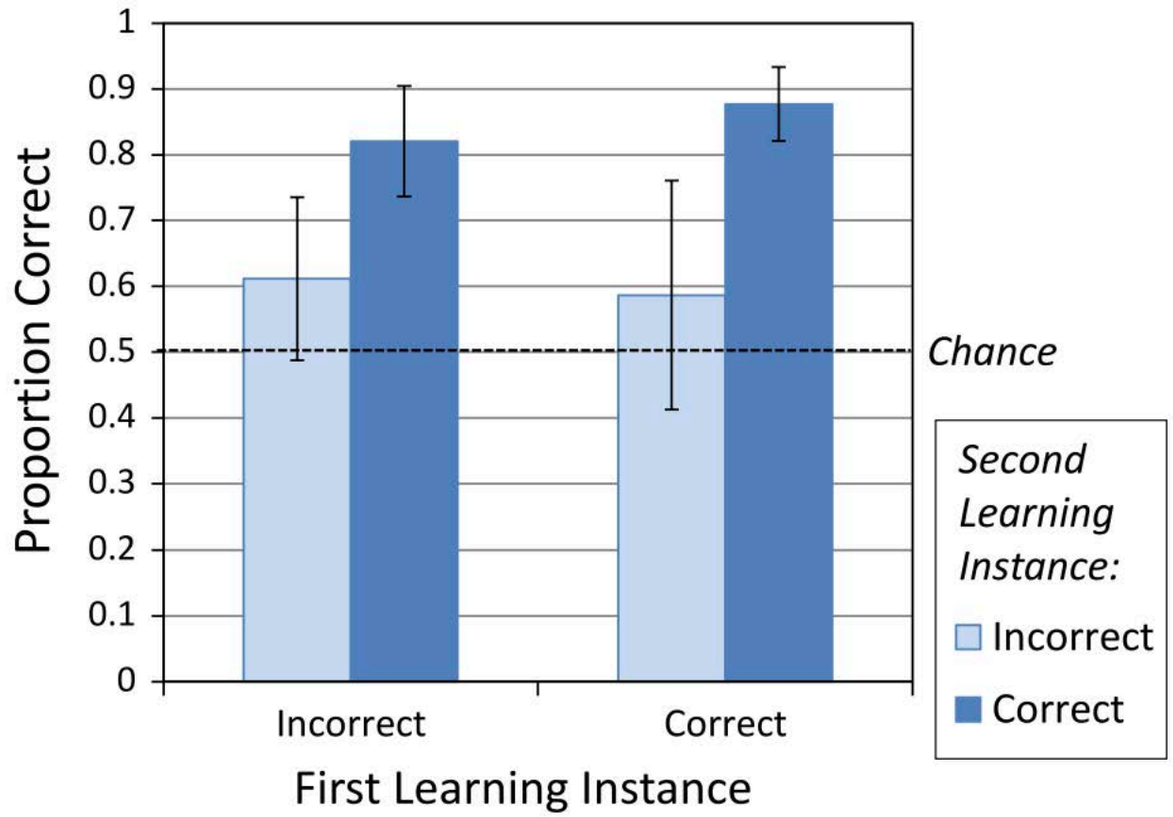


Trueswell et al. 2013

- Experiments 2/3: replication with “high" informativity learning instances

Trueswell et al. 2013

- Results
  - Highly local process
  - Success on trial n is determined only by success on trial n-1; success on n-2 doesn't matter



Trueswell et al. 2013

- **Upshot:** learners are failing to hold in mind possible candidate referents for a given word across learning instances

- How then can they carry out cross-situational learning?

- If not cross-situational learning, then what?

Propose but verify

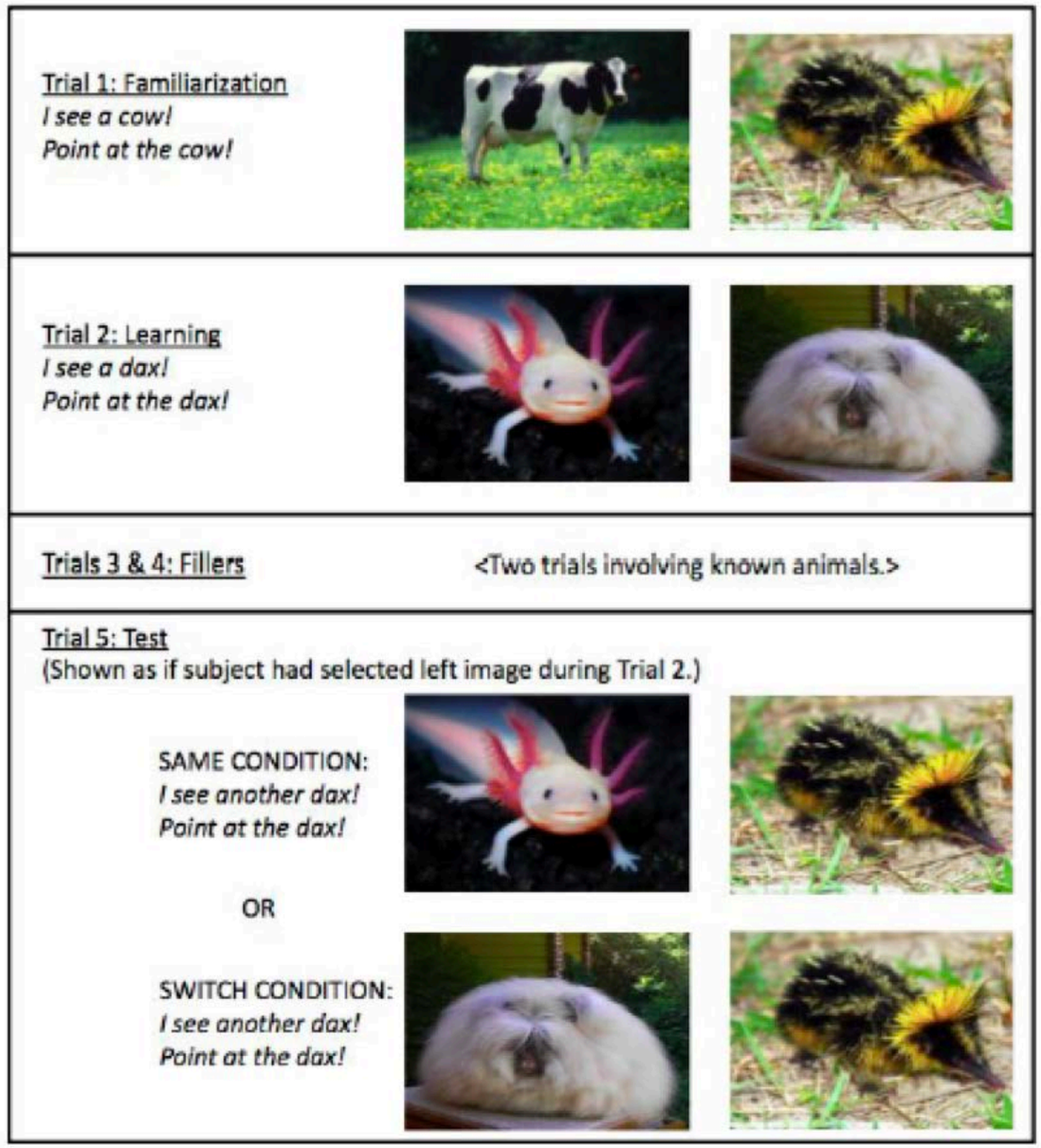
- The learner makes a single conjecture upon hearing the word and carry that conjecture forward to be evaluated for consistency with the next observed context.

- If the guess is “confirmed” in the next instance, the learner will further solidify the word meaning in memory.

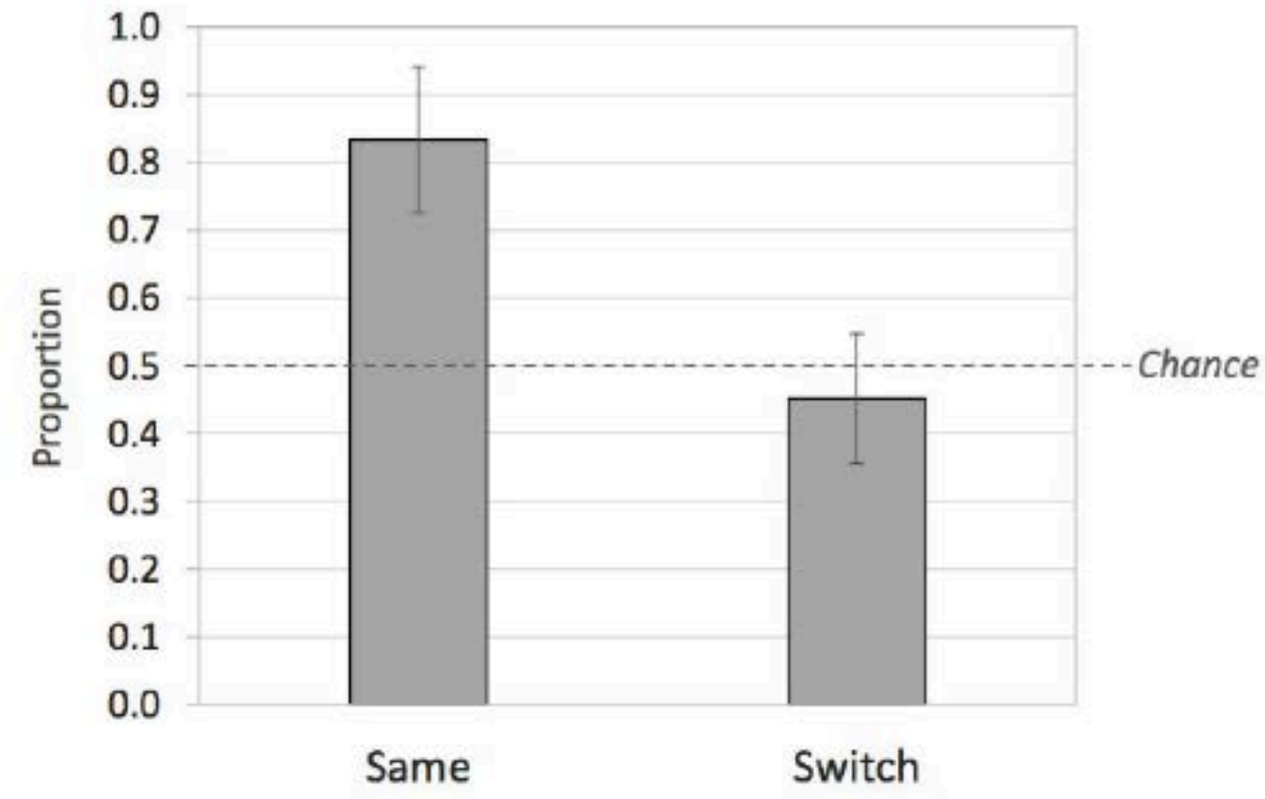
- If the guess is inconsistent with the succeeding observation, the learning machinery will abandon this interpretation and postulate a new one – which can be carried forward, in its turn, for subsequent confirmation or rejection.

Woodard et al. 2016

- Verifying predictions of PbV in 2-3-year-olds (N=32)
- In learning trials, presented with two objects and a label
- Manipulated the test trials following a choice to contain either:
  - the referent that was guessed (**same** condition), OR
  - the referent not guessed (**switch** condition)



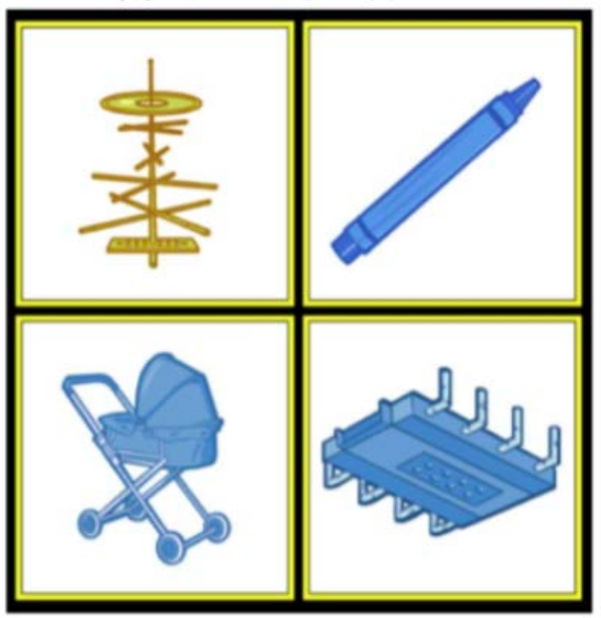
Woodard et al. 2016



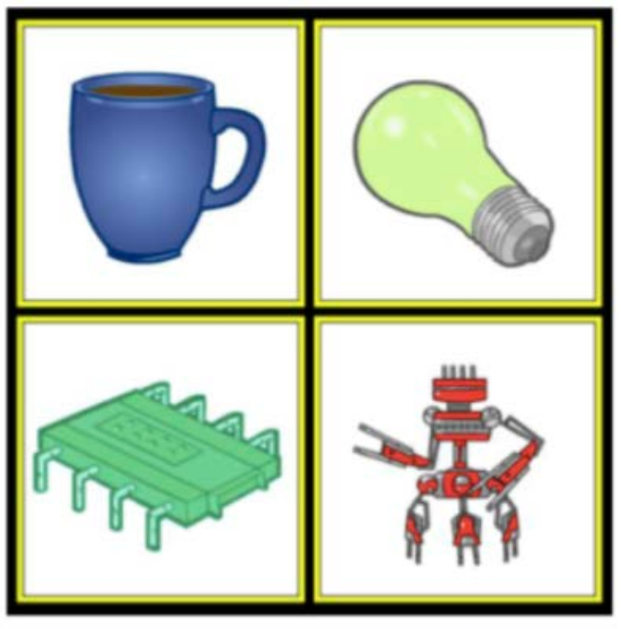
- Same: well above chance (mean = .83, SD = .22, p<.001)
- Switch: not significantly different from chance (mean = .45, SD = .19, p = .33)

Aravind et al. 2017

- 3-5-year-olds (N=674)
- Sufficient cues to make a good first guess
  - the idea: there is a "correct" answer
- Trials 2 immediately follows Trial 1

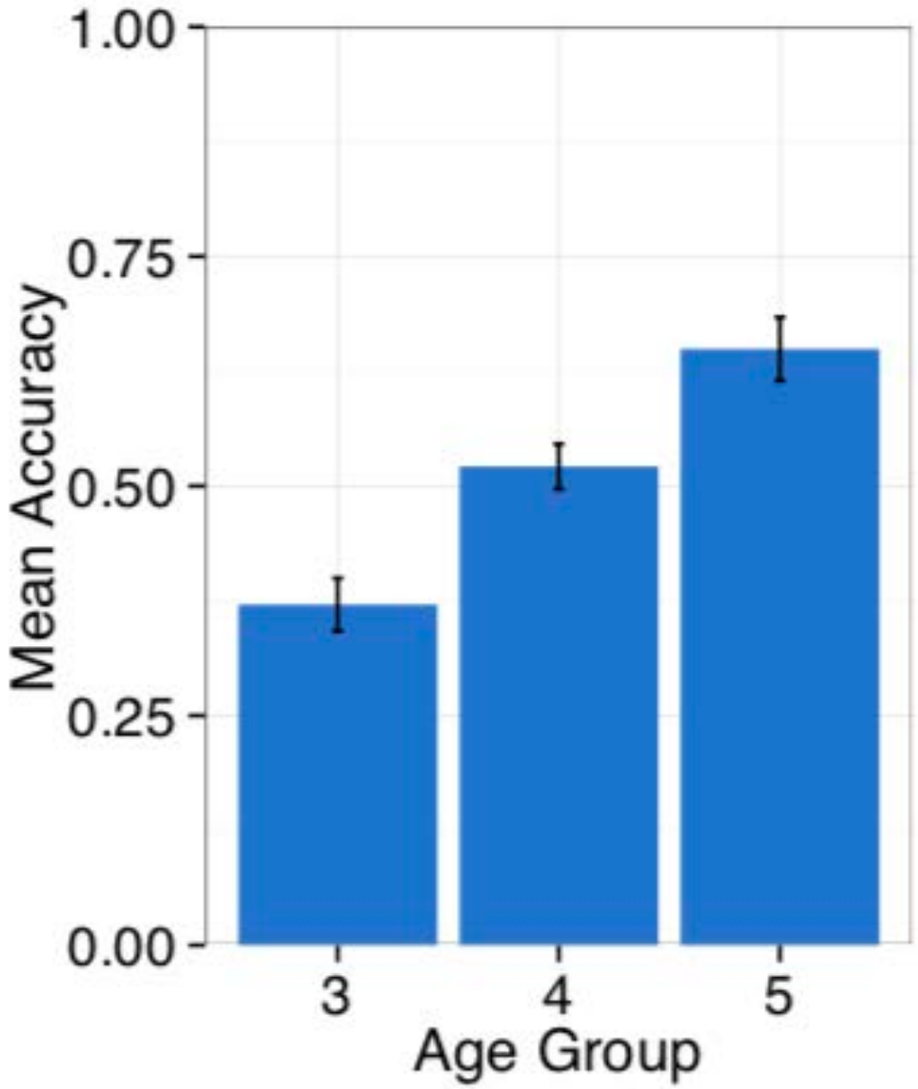


Trial 1:
"The fep is blue. Find the fep!"

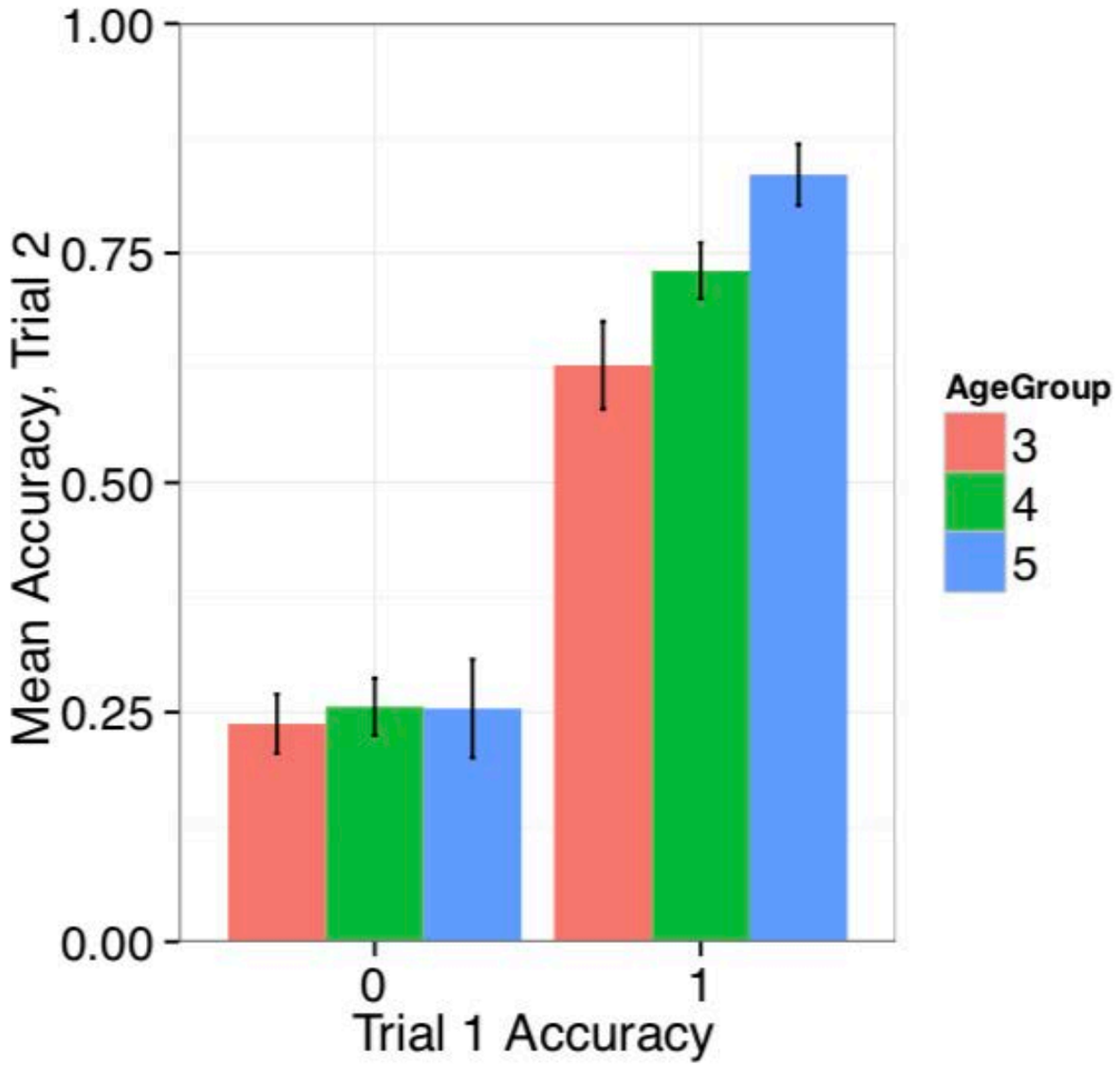


Trial 2:
"Find another fep!"

Aravind et al. 2017



Accuracy on Trial 1 by Age



Accuracy on Trial 2 by Trial 1 Success * Age

CSM v. PbV

"The advantage of PBV seems to be that the child needs to remember less. However, if the child's hypothesis is falsified, the process of establishing the meaning of a word has to start all over again because the child does not remember anything else about the prior contexts in which it was used. CSM on the other hand involves more cognitive load, but is more "fail-safe," in that if the initial hypothesis is incorrect, the child has prior information to form a different hypothesis. PBV becomes more advantageous **if the likelihood of forming false hypotheses is quite small**…"

Guessing well

- Much rides on making the right first guess; otherwise learning would be quite laborious

- What kind of evidence is needed to ensure that the initial guess isn’t radically off? What kind of evidence is utilized?

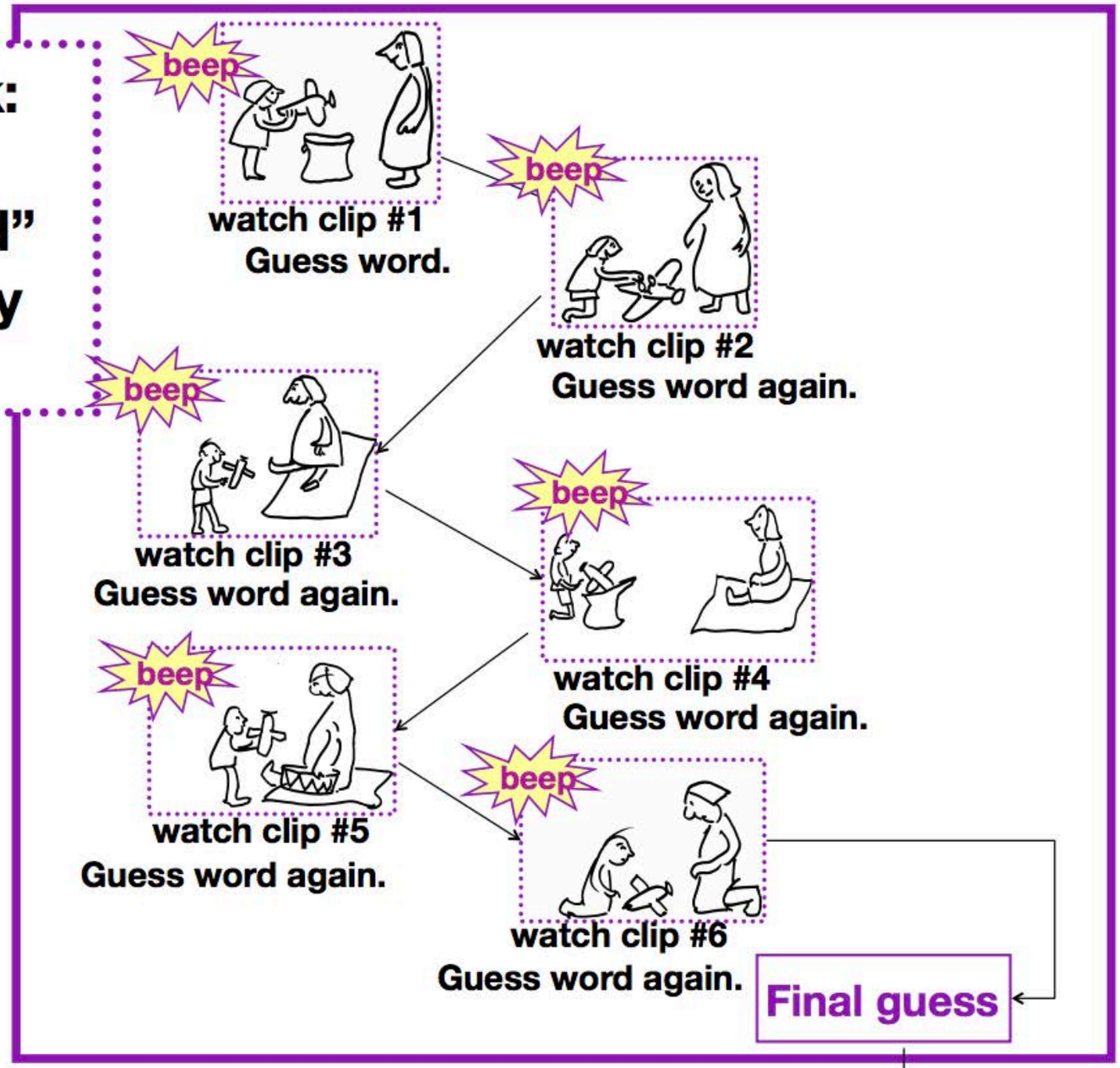
Snedeker, Gleitman, & Brent (1999)

- Human Simulation Paradigm
  - asked adult speakers (who are presumably “cognitively mature”) to view scenes of what mothers are saying to their children and see which words they could learn
  - all audio removed; “beep” at the critical word
- replicated in Medina et al. (2011)

Snedeker, Gleitman & Brent (1999)

- Stimuli preparation
  - Videotape English speaking mothers playing with their 18- to 24-month-old children
  - Transcribe video tape for mothers' 24 most frequent nouns and 24 most frequent verbs.
  - For each of the most frequent words, randomly select 6 uses of the word.
  - Edit each instance for 40 second clips. Audio was removed and a beep is sounded at instant word uttered.

Subject's Task: Identify the "mystery word" represented by the beep.

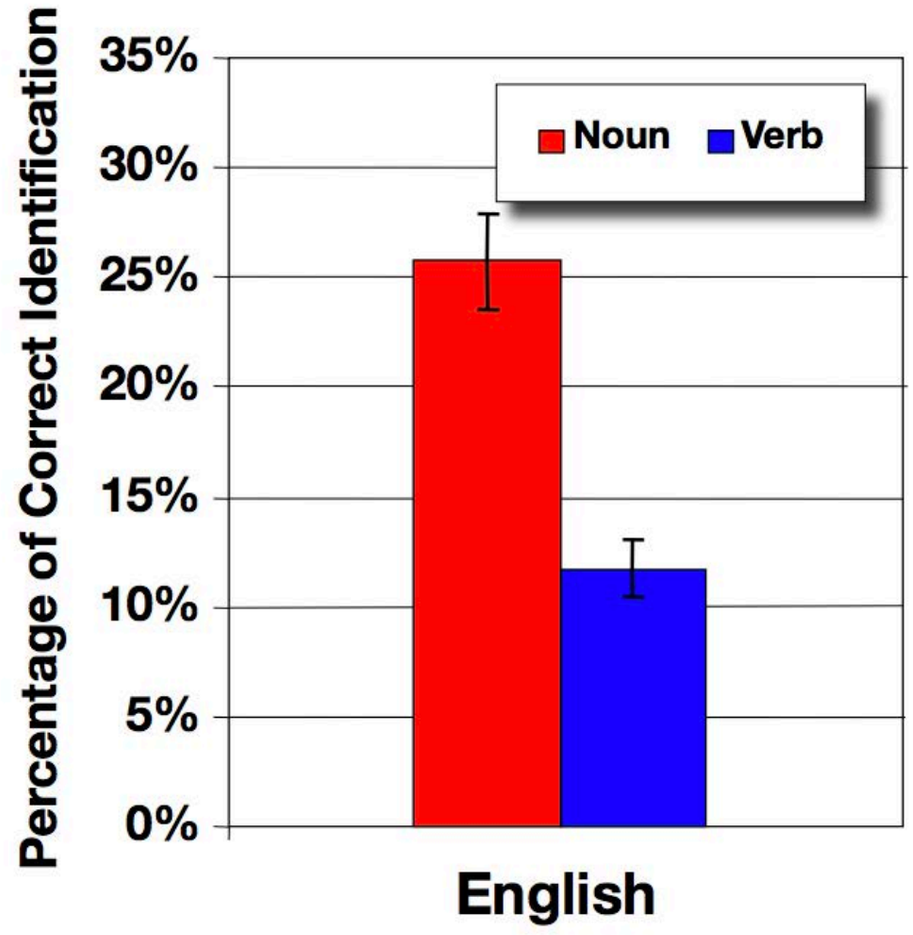


On to next mystery word

Snedeker, Gleitman & Brent (1999)

- Generally quite difficult (~15% accuracy rate overall)

- Nouns easier to identify than verbs



preponderance of nouns in early vocabularies

- Vocabularies of children with 50 or less words are heavily concentrated on experiences child has: names for people, food, body parts, clothing, animals, household items.

- Braginsky, Yurovsky, Marchman, & Frank 2015: large-scale analysis over tens of thousands of children in English, Spanish, Norwegian, & Danish confirming an “over-representation of nouns” in early vocabularies

Referential gems

- Trueswell et al. 2016
  - Corpus study of 360 parent-child interaction videos (40s) that were used as test items in the HSP
  - Predict accuracy of word-identification on HSP by features of the scene

Referential gems

- Only 7% of the videos yield target word-identification rates above 50%

- All of them were nouns

Referential gems

- What characterizes these nouns:

    i. increased likelihood that the target referent appears immediately before word onset

    ii. increased Parent Attention to target, sharply rising 1–3 seconds before word onset

    iii. increased Parent Gesture/Presentation of the target one second before word onset

    iv. increased Child Attention to the target beginning 3 seconds before word onset if not earlier

    v. decreased Parent and Child Attention to non-target referent objects starting at word onset and persisting about 8 seconds after word onset

Referential gems

- Needless to say, they better be concrete objects; the analysis presupposes this!

  - e.g. “increased likelihood that the target referent **appears** immediately before word onset”

Mischaracterization?

- As some of you pointed out in your comments on the readings, making a big deal about ease of “noun”-learning reflects an Anglo-centric bias.

- Ultimately, it’s probably not so much noun-ness that matters, but concreteness (Gentner, 1982; Gleitman & Gleitman, 1997)

- In languages like English, words for concrete things just so happen to be nominal.

Beyond nouns

- But even restricting our attention to languages like English, there is the question of how the child moves beyond an initially concrete, largely nominal vocabulary

- Surely verbs are also acquired. But if observational cues don’t help here, then what does?

Bootstrapping syntax

- Gleitman 1990 *et seq.*
  - The early largely nominal vocabulary helps the infant figure out something about the L1

Bootstrapping syntax

- Early nouns as “seed words”
  - Equipped with rich enough prior knowledge, early nouns could aid in identifying surface-distributional properties of L1, e.g. basic word order, category-specific functional skeleton etc.
    - if you (i) know the words ‘Mommy’ and ‘ball’, (ii) you know that every clause is built around a verbal element, and (iii) you observe Mommy acting on a ball, you might deduce from “Mommy kicked the ball” that English is SVO.
    - if you (i) know the word ‘ball’ and (ii) know that content words form constituents with category-specific functional morphology, you might deduce from “the ball” that *the* combines with noun phrases in English

Syntactic bootstrapping

- The starting idea:
  - Where word-to-world mapping cannot work, sentence-to-world mapping still might
    - ““The structure of the sentence that the child hears can function like a mental zoom lens that cues the aspect of the scene the speaker is describing” (Gleitman & Gleitman, 1992)”
  - there are principled connections between syntactic structures and meaning, such that the range of structures can be informative for deducing which phonological objects goes with which concept

Gleitman 1990

“Speaking more generally, certain abstract semantic elements such as ‘cause, ‘transfer' and ‘cognition' are carried on clause structures (subcategorization frames) rather than (or in addition to) item-specific information in the lexical entries of verbs… It follows that the subcategorization frames, if their semantic values are known, can convey important semantic information to the verb learner"

p. 24

Existence proof from adults

- Compare 3 options for word learning

Gillette et al., 1999; Kako & Gleitman, 2004; Snedeker & Gleitman, 2004

Existence proof from adults

1. Learn from Scenes

- Child relies on situational context alone

- Can learn only very concrete words: object labels

Gillette et al., 1999; Kako & Gleitman, 2004; Snedeker & Gleitman, 2004

Existence proof from adults

1. Learn from Scenes

2. Learn from Nouns

   - Object labels provide richer representation of linguistic context

   - Utterance = set of known nouns

   - Child may learn concrete relational words like spatial prepositions (ex: “near”), some function words and some verbs this way

Gillette et al., 1999; Kako & Gleitman, 2004; Snedeker & Gleitman, 2004

Existence proof from adults

1. Learn from Scenes

2. Learn from Nouns

3. Learn from the syntactic structure (“frames”)

   - Utterance is represented as a syntactic structure + function words

   - This might be necessary for the child to learn more abstract words

Gillette et al., 1999; Kako & Gleitman, 2004; Snedeker & Gleitman, 2004

Snedeker & Gleitman 2004

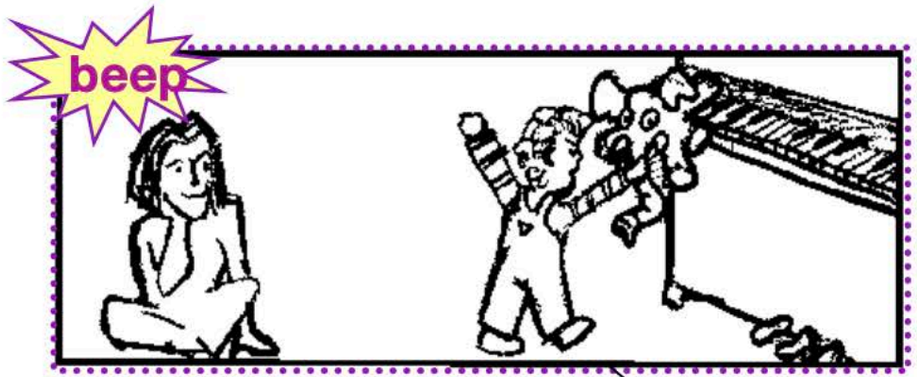
- Targets
  - Videotaped interactions of 4 mother-child pairs
  - 24 most common verbs chosen as targets
  - for each target, 6 instances randomly selected

- 7 groups based on type of identification information
  1. Scenes
  2. Nouns
  3. Frames
  4. Scenes + Nouns
  5. Scenes + Frames
  6. Nouns + Frames
  7. Scenes + Nouns + Frames

Gillette et al., 1999; Kako & Gleitman, 2004; Snedeker & Gleitman, 2004

Scenes condition

Example “mystery verb”: “play”



Guess word.

Guess word again.

Etc….

Final guess

Task: Subjects guess mystery verb from watching 6 instances of word use in video clips. The video clips are silent except beeps replace the moments the mystery word were uttered.

Nouns condition

Example “mystery verb”: “play”

1. elephant, piano

Guess word.

2. mommy

Guess word again.

3. I, it, you

Guess word again.

4. it, you

Guess word again.

5. drums

Guess word again.

6. music, you

Final guess

Task: Subjects shown the nouns co-occurring with the mystery verb in 6 sentences, the same sentences as those in the video clips with the beeps.

Frames condition

Example “mystery verb”: “play”

1. Can kax SIRN the bussit?

Guess word.

2. Noggle SIRN?

Guess word again.

3. Can po SIRN while lo nirp nu?

Guess word again.

4. Lo are gonna SIRN nu?

Guess word again.

5. SIRN the neps.

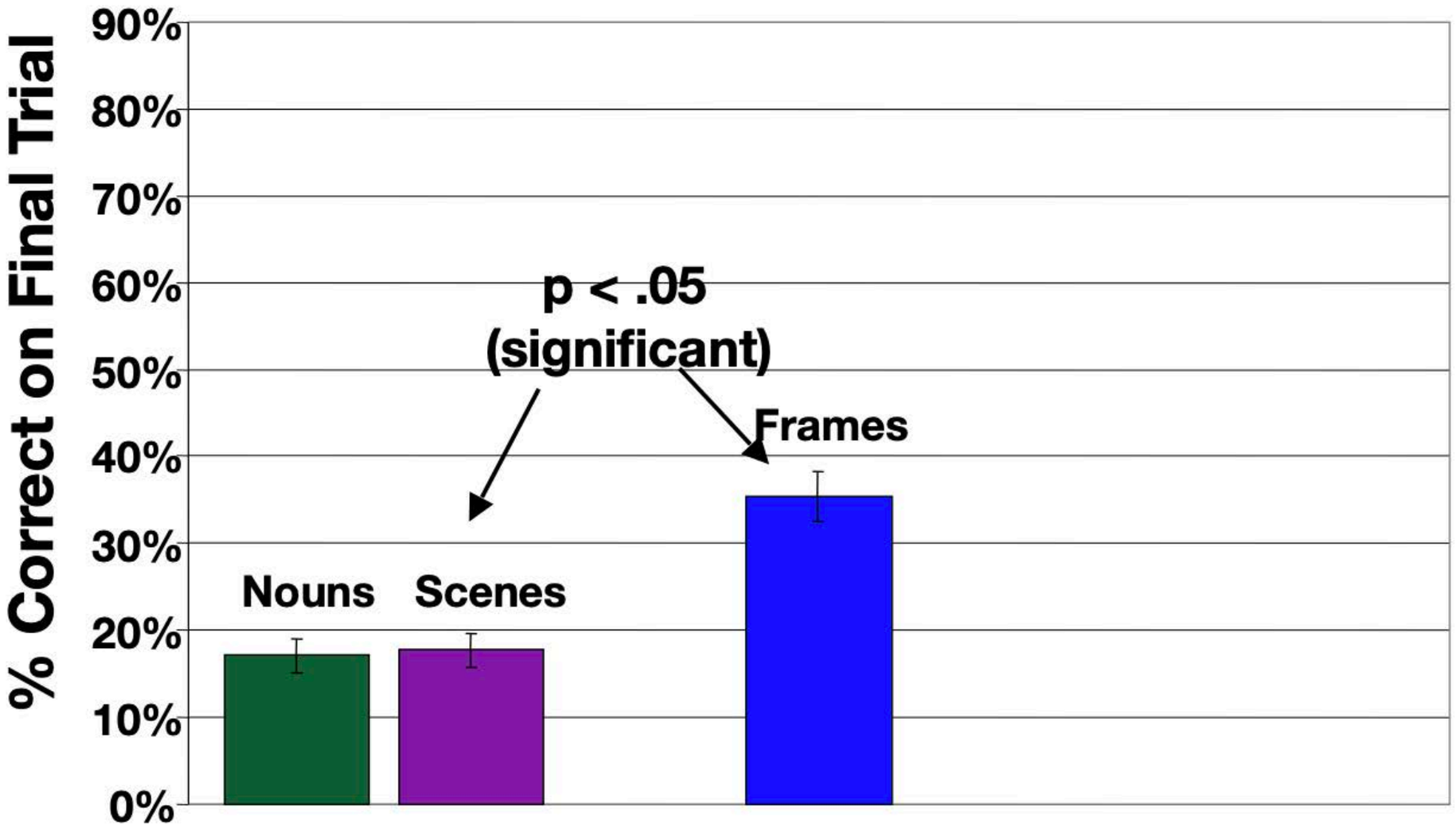
Guess word again.

6. Lo SIRN tuggy wilm.

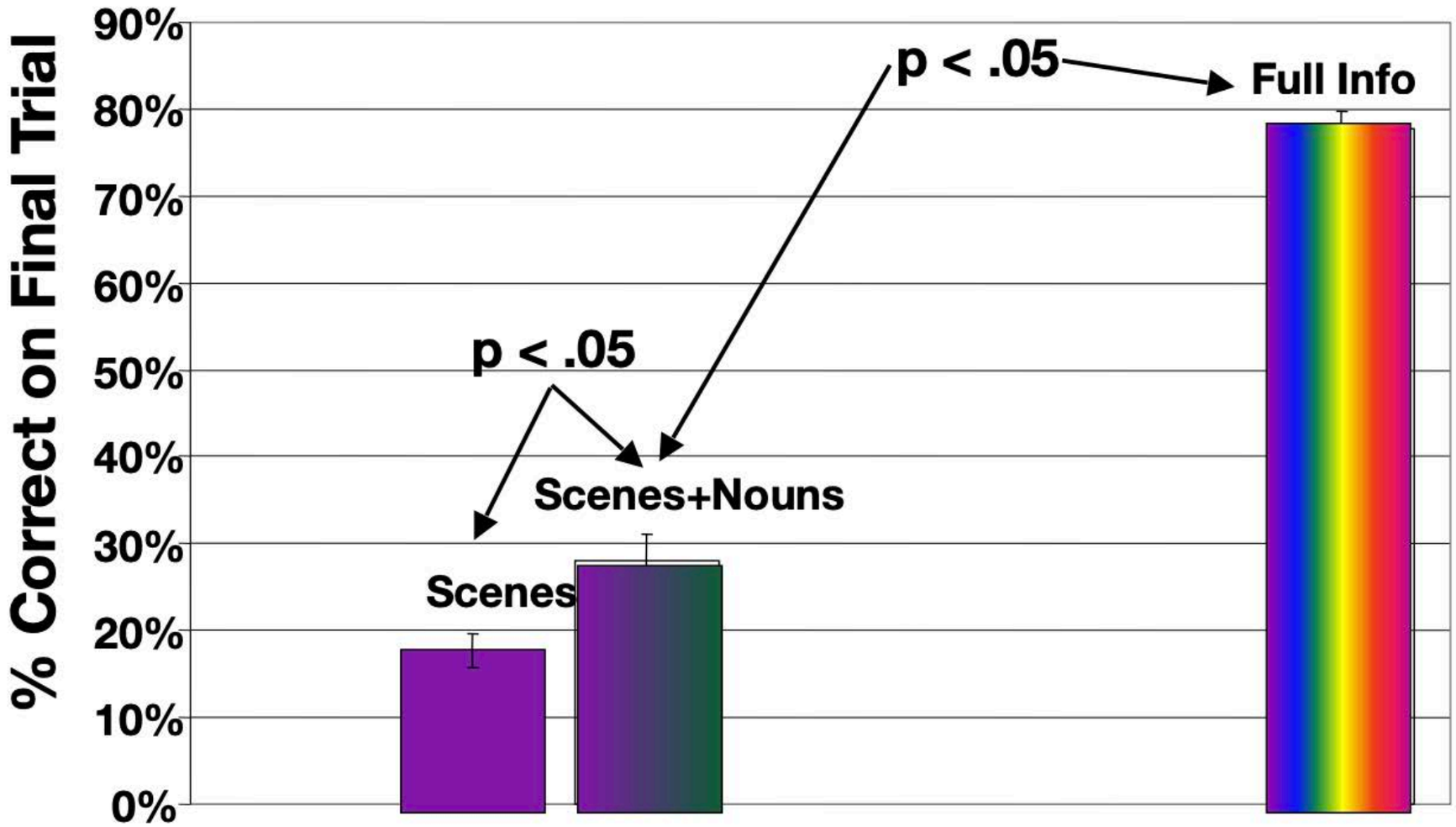
Final guess

Task: Subjects guess the mystery verb from the 6 sentence frames. The sentence frames are constructed by replacing words in the 6 utterances with nonsense words.

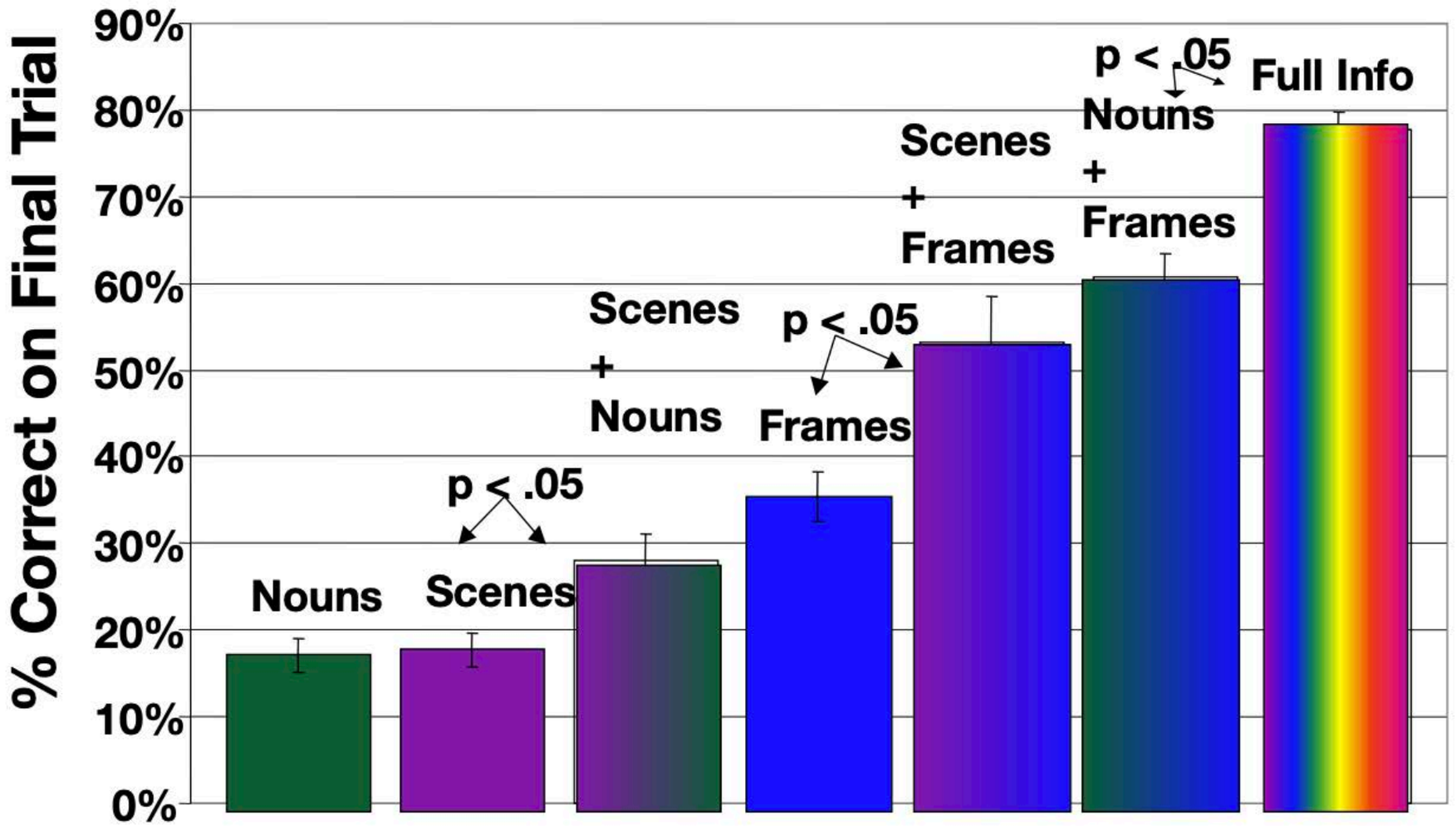
Rate of correct identification varies w/ condition



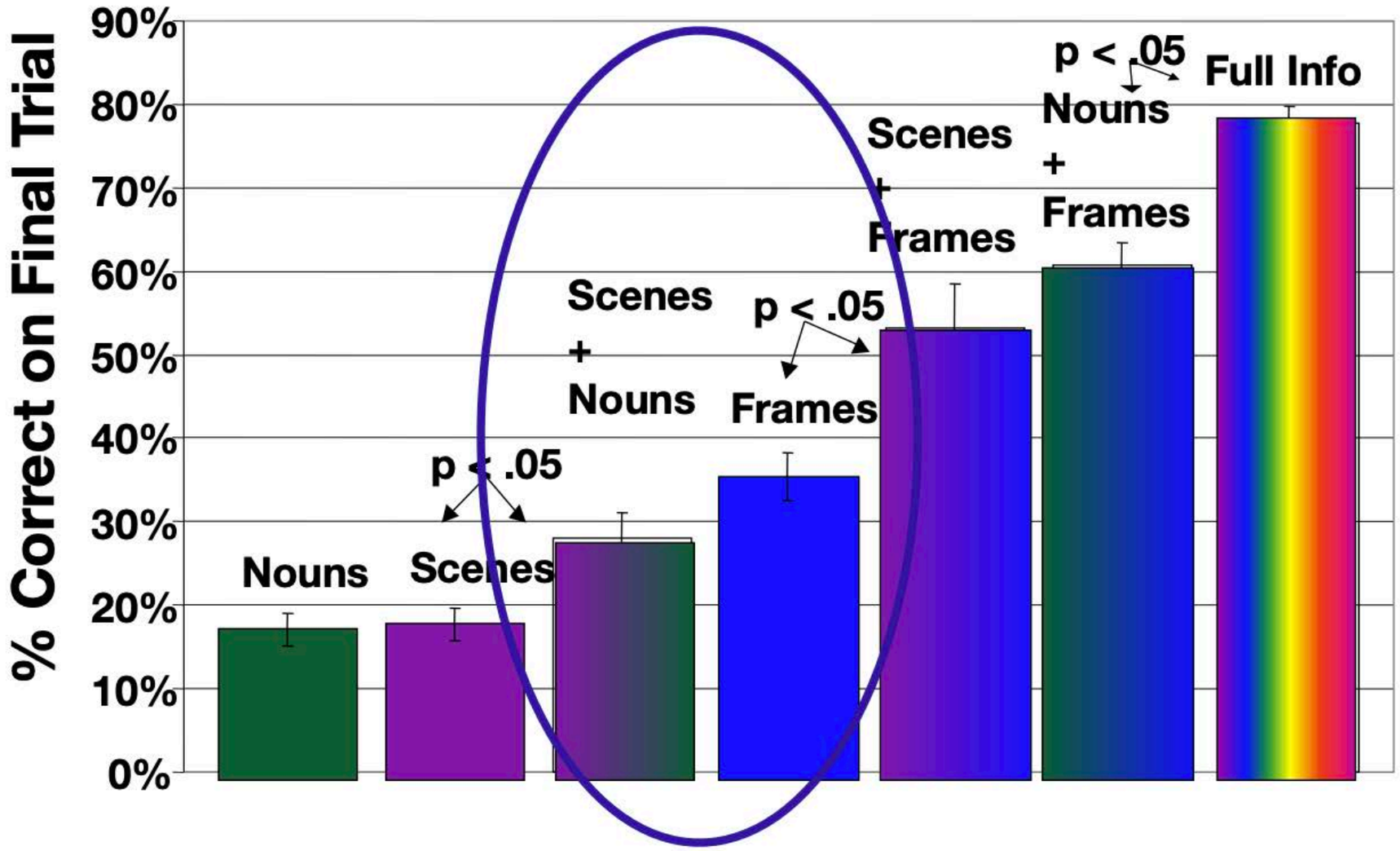
Rate of correct identification varies w/ condition



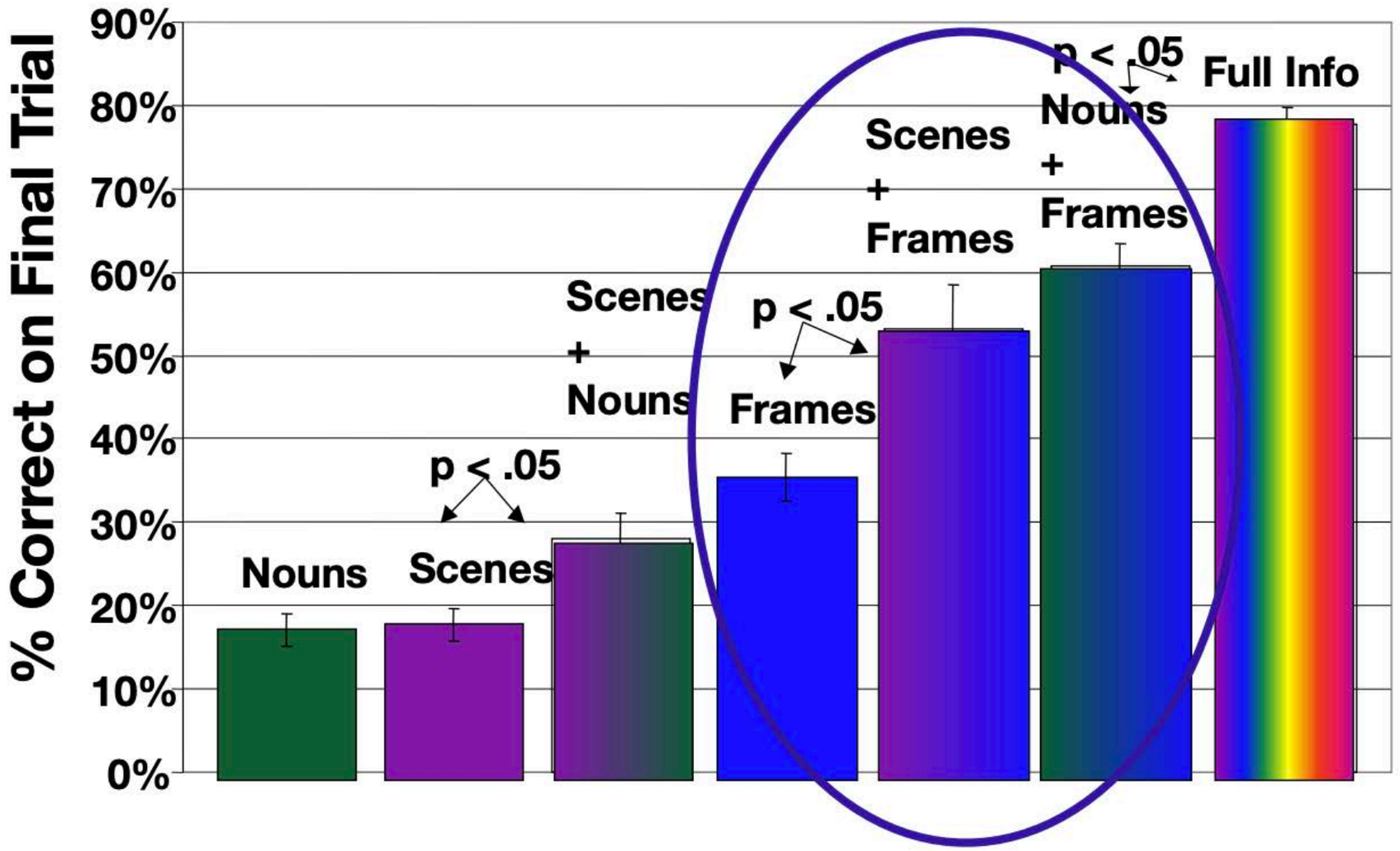
Rate of correct identification varies w/ condition



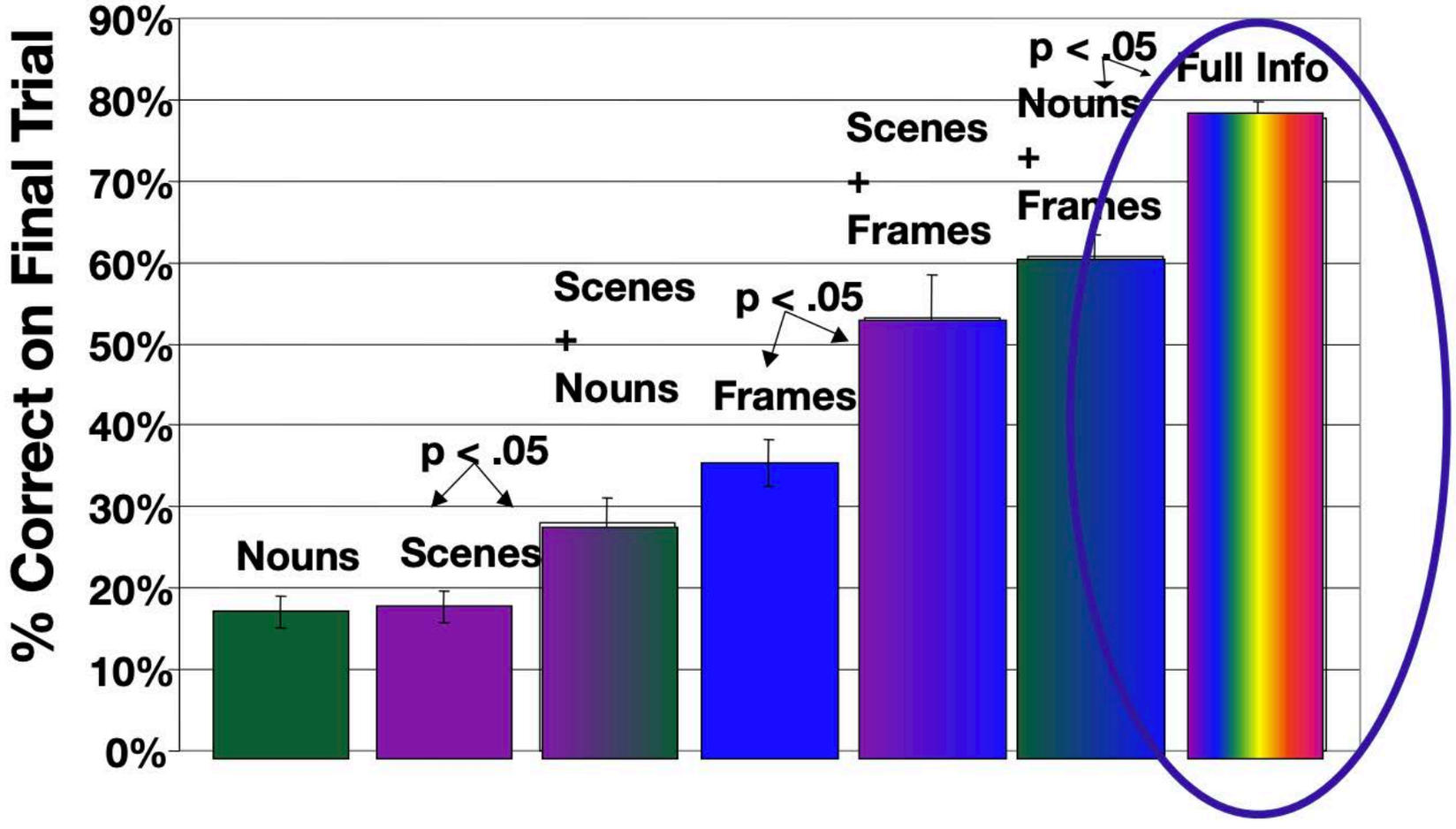
Utility of syntactic frame knowledge: Scenes + Nouns equivalent to Syntactic Frames only



Utility of additional knowledge with Frames: Scenes + Frames equivalent to Nouns + Frames, which is better than Frames alone



Superiority of using all the available information: Scenes + Nouns + Frames is better than all other information type combinations



Upshot

- Identifying verbs isn’t so hard once you have some linguistic background (like knowing some nouns and some notion of syntactic structure) and informative situational context (scenes)

What we haven’t yet seen

- A full-fledged theory of learning this way:
  - What exactly are the “bootstraps”?
  - What are the syntactic primitives? semantic/conceptual primitives? mapping functions?

Next week

- Evidence (?) for children “bootstrapping” verb meanings from syntax [read **two**]
  - Lidz, Gleitman and Gleitman (2003)
  - Fisher (1996)
  - Yuan and Fisher (2009)
  - Fisher et al. (2020)
- Evidence that syntax might indeed aid in the acquisition of some aspects of content words [read **one**]
  - Gordon (1985)
  - Shi and Melançon (2010)
- Fun read [optional, but highly recommended]: Chapter 3 of Fodor's (1998) *Concepts*