Language Acquisition

discussion of papers

Psycholinguistics

LING/PSYC 27010

Autumn 2016

timleffel@uchicago.edu
lefft.xyz/psycholingAU16

agenda for today

1. housekeeping
2. lecture/discussion
   - Markman (1990)
   - Smith & Yu (2008)
   - Snedeker & Gleitman (2004)
3. wrapping up language acquisition

format of paper discussion days

1. for each paper in the reading list:
   - general questions/reaction (briefly)
   - research questions and motivation for the study
   - methodology, experimental paradigm/design
   - results (mostly non-technical)
   - interpretation of results, relation to theory (quick overview, then discussion)
2. zooming out –
   - what picture do the papers paint jointly?

how do learners figure out what words refer to?

the mapping problem

indeterminacy of translation,
underdetermination of meaning
(*gavagai* scenarios, Quine)

Markman (1990)

general questions or remarks about the paper?

Markman (1990) et seq: key results

language learning involves certain **biases and heuristics** which are **not** general-purpose learning strategies

- type assumption: new words refer to a type of thing, not a particular thing.
- whole object assumption: new words refer to whole objects, not just their substance, color, or parts.
- mutual exclusivity bias: new words don't refer to things that already have names

Smith & Yu (2008)

general questions or remarks about the paper?

Smith & Yu (2008): questions/motivation

indeterminacy of translation, underdetermination of meaning
(*gavagai* scenarios, Quine)

Markman-style paradigm does not capture the **repetitive** and **simultaneous** nature of *gavagai*-type encounters with word-object pairings

Smith & Yu (2008): questions/motivation

main question

how do children learn the mapping between objects in the world and their linguistic labels? ("gavagai")

"The human learning environment is data rich. If human learners possess the right learning mechanisms, they may mine this complexity and in so doing solve the problem of referential uncertainty."

Smith & Yu (2008): methodology/design

Smith & Yu (2008):

- production-comprehension gap suggests infants have knowledge about word meaning not captured by one-off novel word tasks
- introduce conditions more similar to actual primary linguistic data
  (but still with novel words)

Smith & Yu (2008): methodology/design

design:

training phase: infants view a series of individually-ambiguous trials that **collectively** provide definitive evidence for six specific word-object mappings

test phase: infants exposed to a single word with two possible referents, one of which can be deduced to be the correct referent given generalization over the training trials

Smith & Yu (2008): methodology/design

design (cont):

task: preferential looking – compare looking time to targets versus competitors in test trials

logic: preferential looks to the target indicate infants' association between the word and its referent (the target)

population: 12- and 14-month children

Smith & Yu (2008): methodology/design

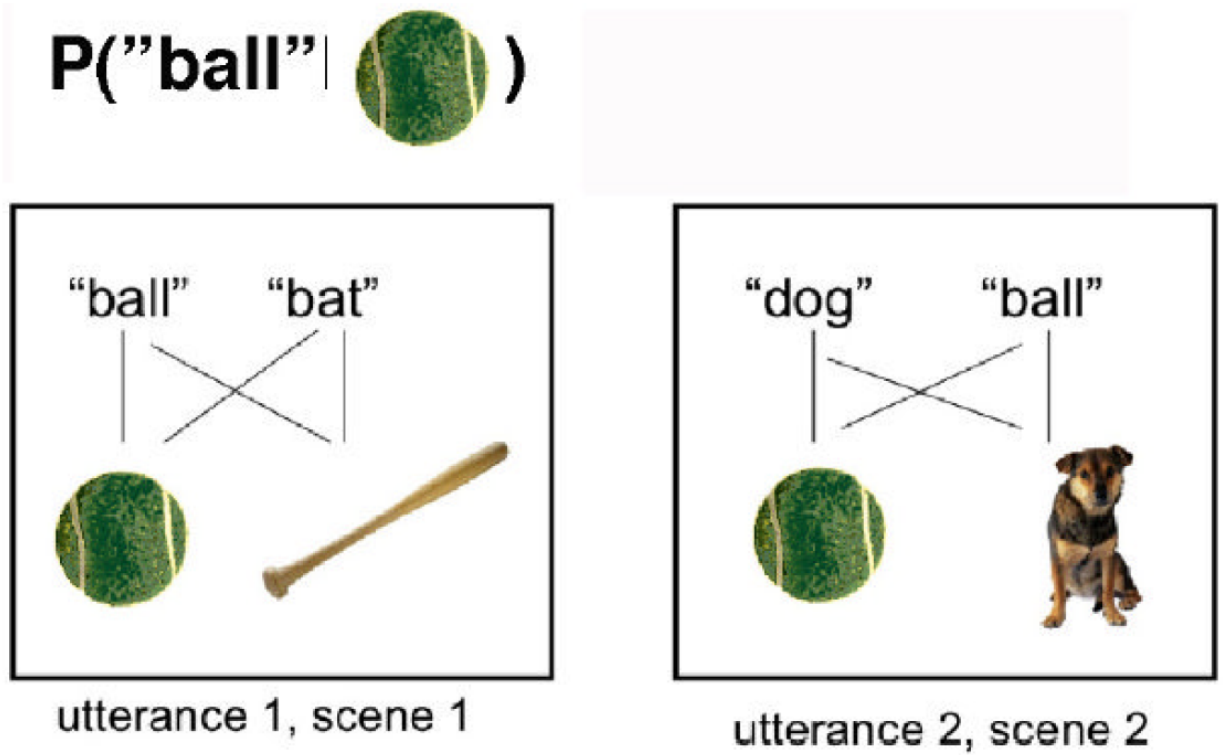


Figure 1.
Associations among words and referents across two individually ambiguous scenes. If a young learner calculates co-occurrences frequencies *across* these two trials, s/he can find the proper mapping of "Ball" to BALL.

Smith & Yu (2008): methodology/design

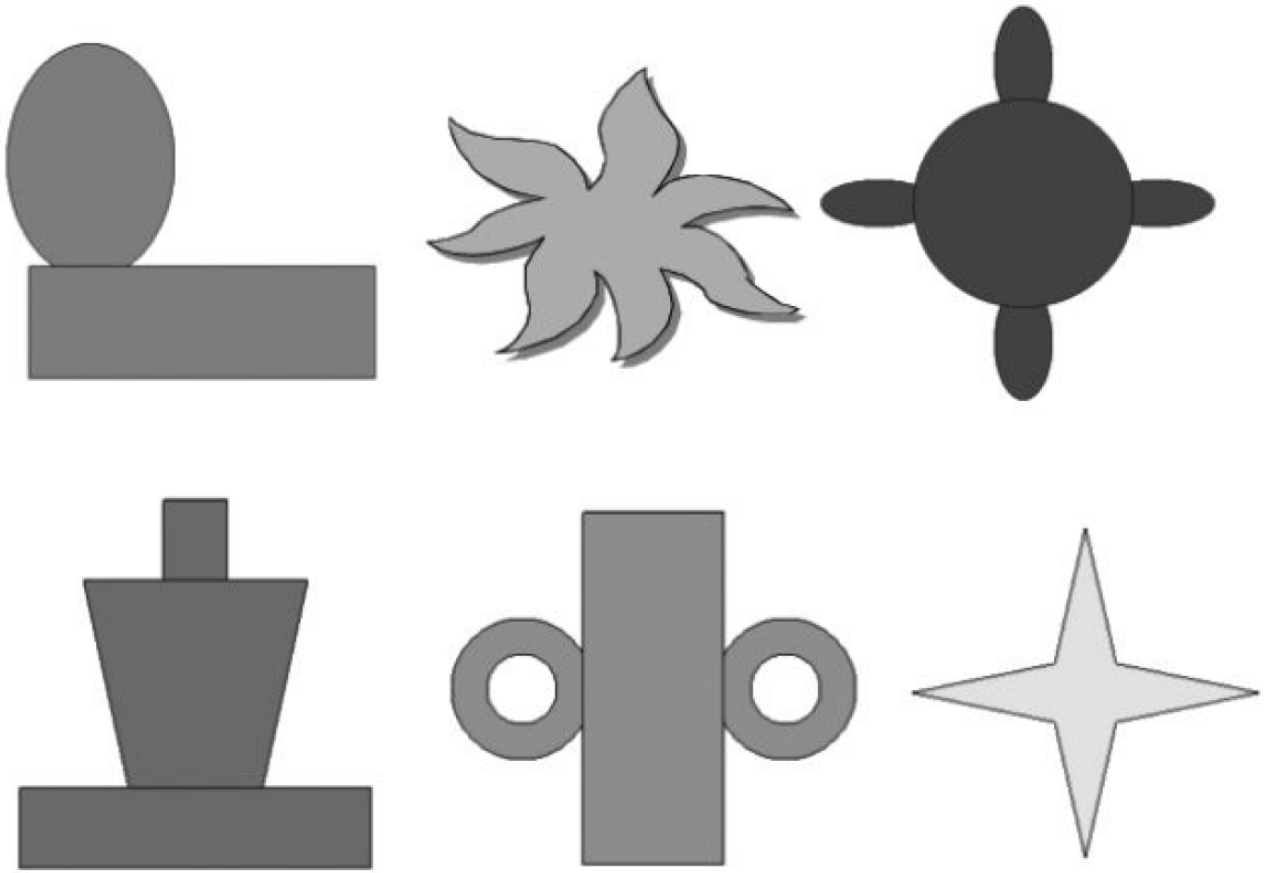


Figure 2.
The six stimulus shapes.

Smith & Yu (2008): results

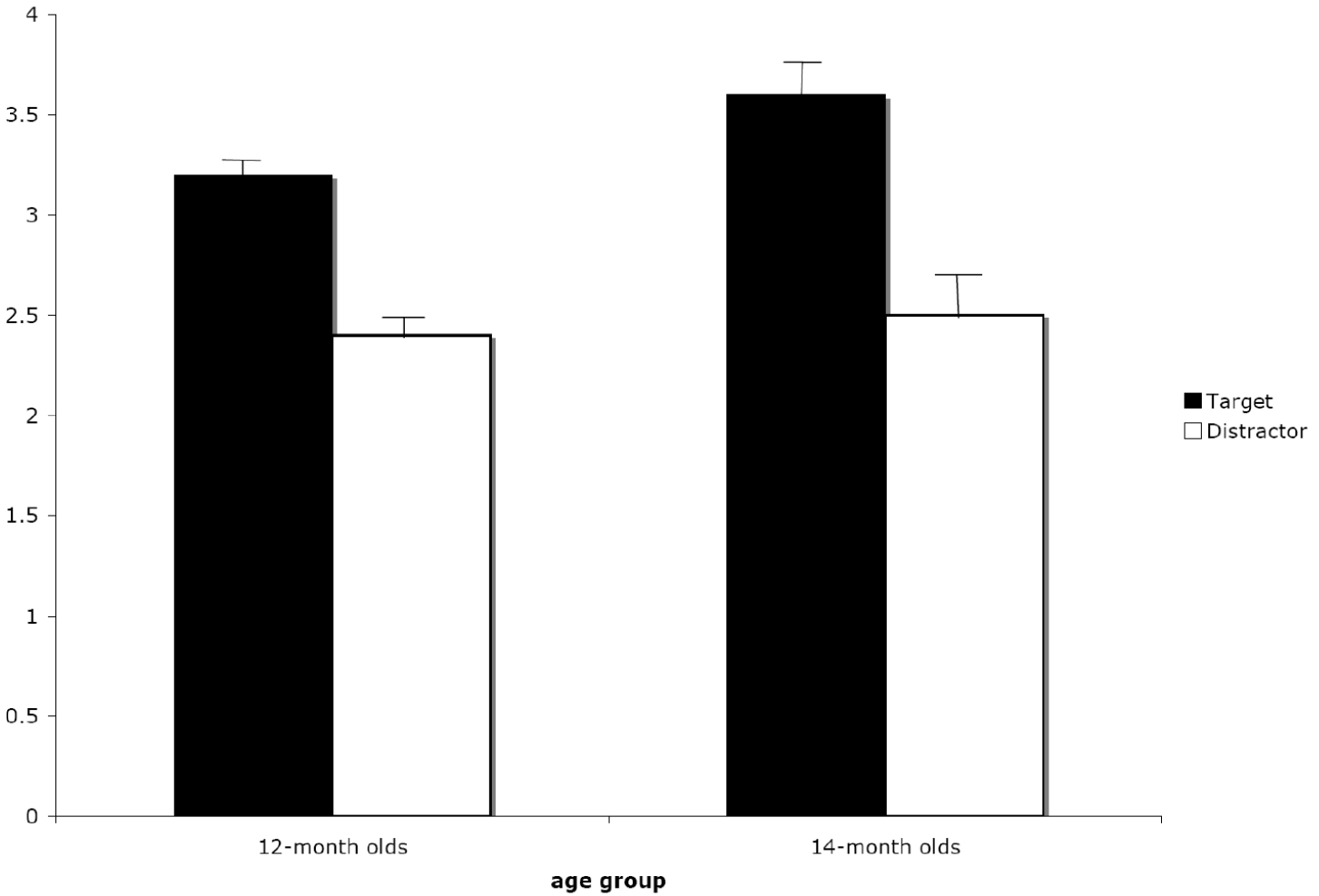


Figure 3.
Mean looking time to target and distracter per 8 sec test trial (and standard error of the mean) for younger and older infants.

Smith & Yu (2008): results

children looked reliably longer at targets than at distractors*
(**main effect of target/competitor**)

the difference in looking time varies across words
(**word** $\times$ **target/competitor interaction**)

preference for target stronger for 14mo than for 12mo children
(**age** $\times$ **target/competitor interaction**)

Smith & Yu (2008): interpreting results

main take-away

statistical regularities in word-object associations used in word-learning, even by 12-14mo children (here, co-occurrence frequencies)

but what is the mechanism? two possibilities are:

- "hypothesis-testing" theory
- associative learning theory

relation to nativism debate?
authors' interpretation?

Smith & Yu (2008): useful quotes

"infants can figure out multiple word-referent mappings from a system of experienced associations."

"the mechanisms responsible for the present results may be relevant to making use of the complexity in natural learning environments, evaluate the regularities in the data set as a whole, and determine the underlying mappings. Such mechanisms could even benefit from increased complexity in the data set."

"[the mechanisms should] help children learn from the regularities that accrue across the many ambiguous word-scene pairings that occur in everyday communication."

Snedeker & Gleitman (2004)

general questions or remarks about the paper?

Snedeker & Gleitman (2004): questions/motivation

starting point:

nouns before verbs

"Even though children hear both verbs and nouns from earliest infancy, their earliest vocabulary is overwhelmingly nominal"

concreteness of verbs

"The earliest verbs are not an unbiased sample of those that appear frequently in the input." (*throw*, *run* before *think*, *know*)

main question: why?

Snedeker & Gleitman (2004): questions/motivation

the cognitive development theory

nouns associated with "simpler" concepts, hence learned first. vocab development reflects cognitive development (same w concrete/abstract verbs)

the linguistic development theory

verbs more linguistically complex than nouns, hence acquired later bc they require richer ling knowledge

Snedeker & Gleitman (2004): methodology/design

strategy: use adults to "simulate" the process of word learning from early primary linguistic data

$\Longrightarrow$ resulting data used to estimate the learning function of child language acquisition

motivating intuition: linguistic complexity should be reflected in adult language processing as much as it is in child language acquisition

Snedeker & Gleitman (2004): methodology/design

experiment format:

human simulation paradigm

scene only

nouns and frames

full information (no verbs)

full vocabulary

task:

identify target word after six (masked) tokens (masking procedure varies across experiments and conditions)

Snedeker & Gleitman (2004): methodology/design

logic: compare "acquisition" of nouns versus verbs in adults, under conditions that mimick those experienced by children learning their first language

"by examining the use of the word in a variety of contexts, the observer can attempt to parse out those properties of the world common to all these encounters"

population: undergraduate English speakers
(note: different kind of generalization from sample than other papers)

Snedeker & Gleitman (2004): results

45% of nouns correctly identified, versus 15% of verbs (scene + beeps condition)

every noun target identified correctly by at least one subject; a third of verbs never correctly identified

Snedeker & Gleitman (2004): verbs study

conditions:

scenes only*	nouns only	frames only
scenes + nouns*	scenes + frames	nouns + frames
full information*		

verb classes:

action verbs

fall, stand, turn, play, wait, hammer, push, throw, pop

mental verbs

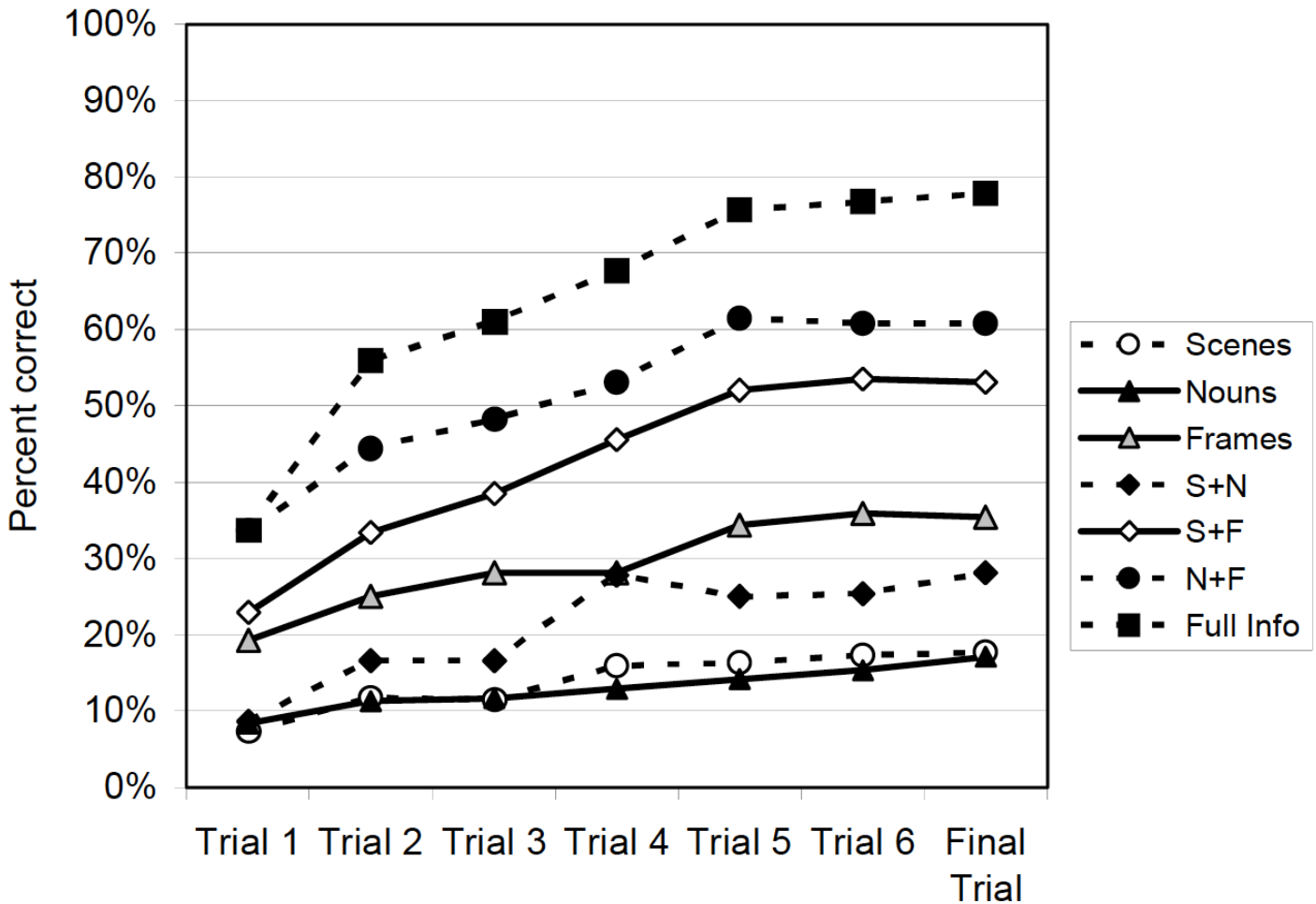
know,like,see,say,think,love,look,want

light verbs

come, do, get, go, have, make, put

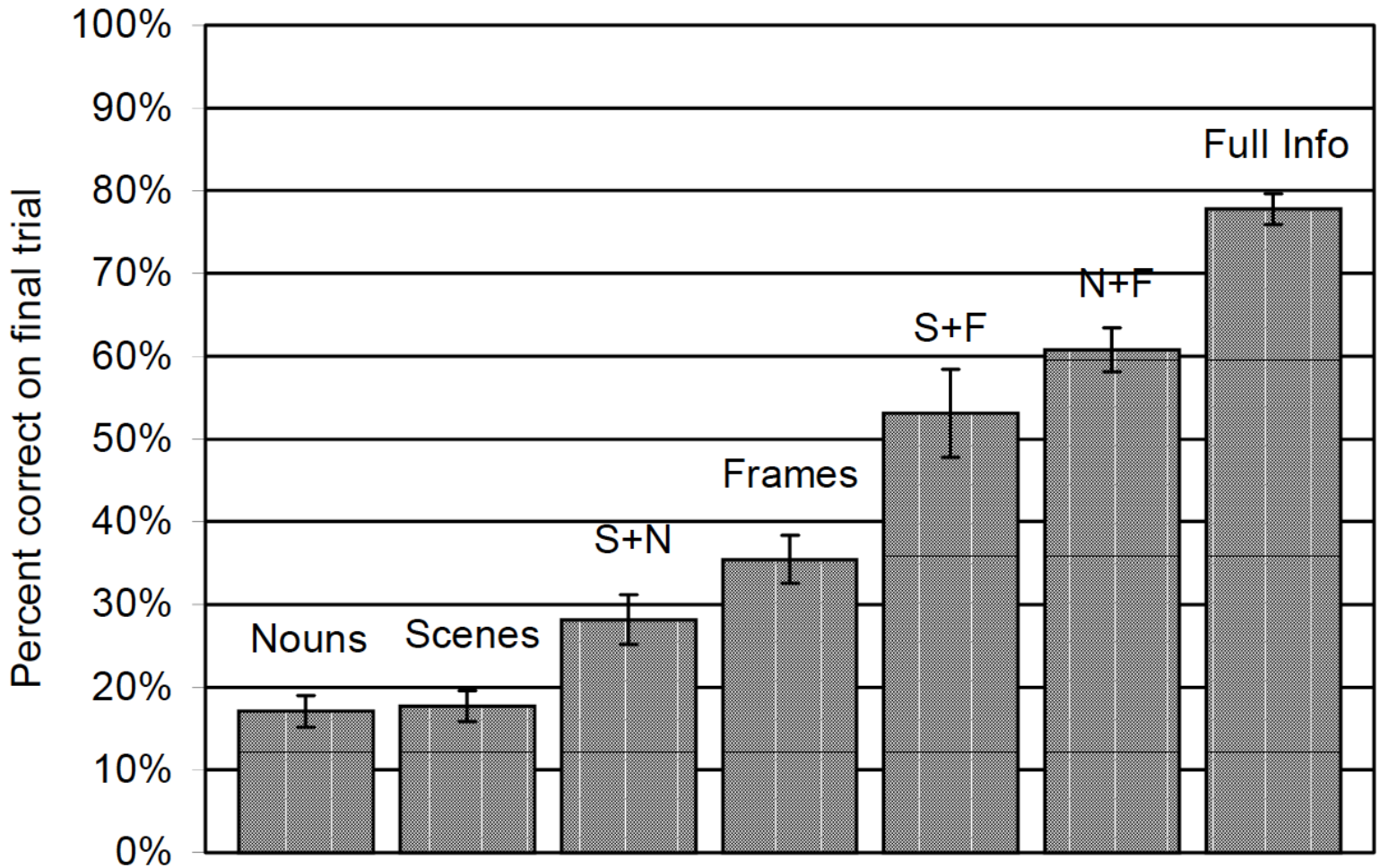
Snedeker & Gleitman (2004): results

Figure 1: Performance across trials for each information condition (verb targets).



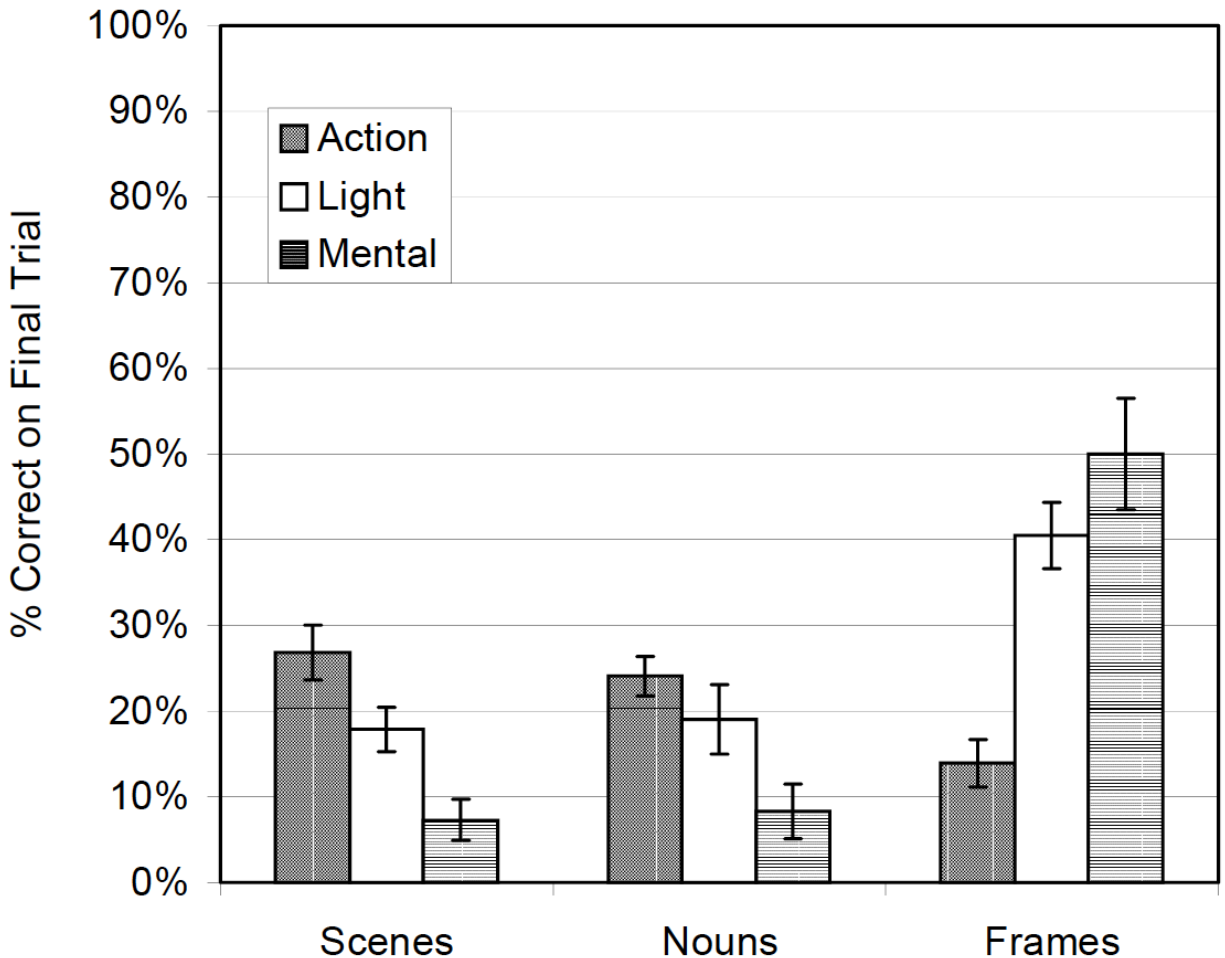
Snedeker & Gleitman (2004): results

Figure 2: Performance on the final trial for all conditions.



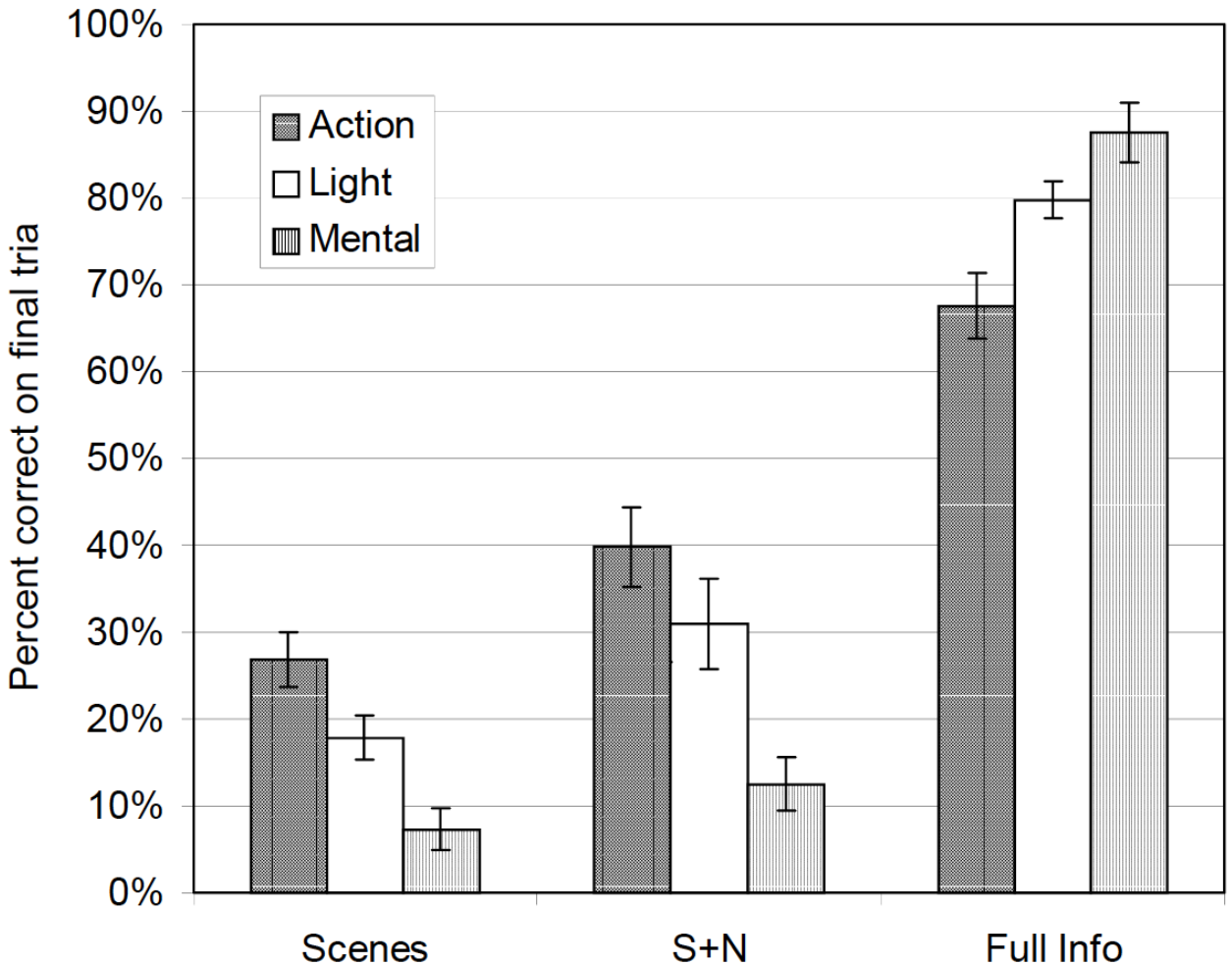
Snedeker & Gleitman (2004): results

Figure 3: Different types of verbs require different information sources.



Snedeker & Gleitman (2004): results

Figure 4: Performance for each verb class improves with increased representational resources.



Snedeker & Gleitman (2004): interpreting results

syntactic bootstrapping

S&G conjecture "a learning device that uses primitive quasi-structural noun-to-verb information for purposes that go beyond acquiring the specifics of verb meaning: as a bootstrap into the language-specific clause level syntax"

Snedeker & Gleitman (2004): interpreting results

think of the informationally impoverished conditions as the earliest stages of language acquisition

think of the informationally rich conditions as later stages of linguistic development

early $\Longrightarrow$ noun-dominant; verbs concrete-dominant

later $\Longrightarrow$ less noun-dominant; all kinds of verbs

even later $\Longrightarrow$ clause-level syntax

Snedeker & Gleitman (2004): interpreting results

all kinds of info is useful for acquiring language...

including lang-specific info that accrues over time

co-occurrence frequencies, syntactic context (frame), environment, etc., are all factors that aid in the acquisition of more complex linguistic structures

this is the essence of **bootstrapping**

wrapping up

how do these results bear on the theoretical issues introduced on Tuesday?

wrapping up

key takeaway points

- various biases and heuristics are employed in word-learning
- domain-general statistical learning strategies detectable even during infancy
- **L**anguage-specific factors do seem to affect learning difficulty (hence acquisition process)
- diverse range of methodologies can shed light on acquisition process

the nativism debate rages on!