Input, Uncertainty and the Early Course of Language Development

Colin Bannard
Department of Linguistics
University of Texas at Austin

Some recent findings concerning children's statistical learning of language

- Young children can use "statistical learning" to
  - segment sequences into useful units
    (8 months; Saffran et al, 1996)
  - identify non-adjacent dependencies
    (15 months; Gomez, 2002)
  - assign sounds to categories based on their distribution (12 months: Gomez and Lakusta, 2004)

And yet...

- Even in their 3rd year children fail to utilize basic generalizations in disambiguating "who did what to whom" for novel verbs, e.g.
  - SVO word order in English (Akhtar et al, 1997)
  - Case marking in German (Dittmar et al, 2008)

Statistical learning in language development

- Implicit measurement of knowledge using preferential listening booth
- Almost all conducted using artificial languages

Ecological validity?

- Statistical learning experiments can tell us a great deal about what infants can do with controlled input we also need to look look closely at the language they hear:
  - it offers unique challenges
  - it also offers shortcuts

Inference vs decision-making

- Inference = estimation of probabilities given some experience and/or prior belief
- Decision making = use of those probabilities
- We need to understand not just what information a child infers but what use they put it to and why!

Outline of studies

Study 1: Do 2 and 3 year olds reuse sequences of words taken directly from the input?

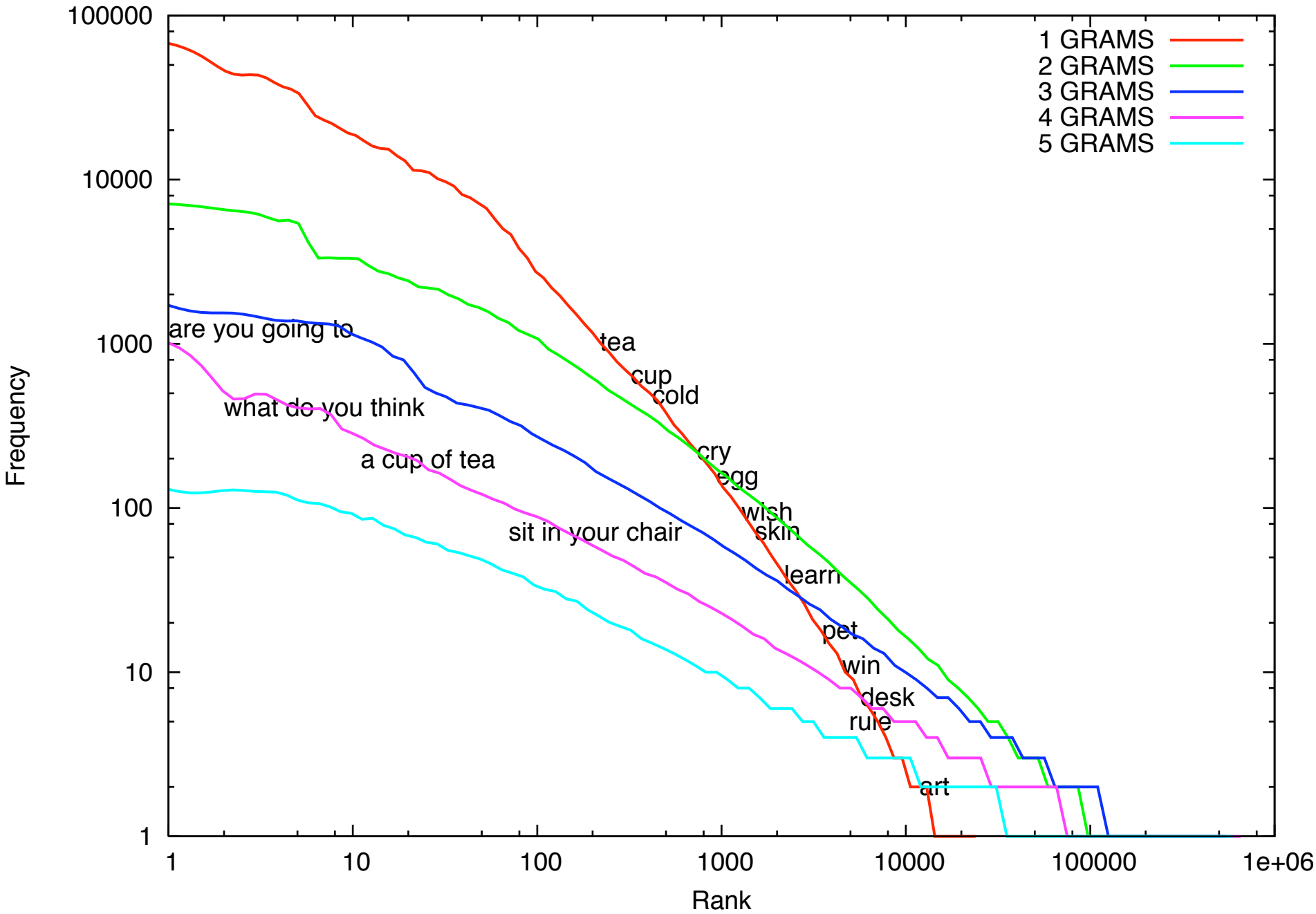
Study 2: When will 3 year olds imitate novel linguistic material?

Study 3: Can we use corpora to predict when children will exploit generalizations?

Language learning as language reuse

Bannard and Matthews (2008, *Psychological Science)*

What children hear



(In 1.8 million words of caregiver's speech)

Would we not expect kids to learn multiword units along with words?

How can we test this?

- Extract matched pairs of frequent and infrequent sequences from a corpus of caregiver's speech
- Use a repetition task to measure processing difficulty

If the children have no dedicated representation for the sequence we should see no frequency effect

	f(chunk)	f(bigram)	f(word)
a drink of milk	4.04	6.68	5.06
a drink of tea	2.39	6.93	5.84

back in the box	3.74	7.31	6.57
back in the car	1.61	7.10	6.03

a piece of cheese	3.85	6.81	4.93
a piece of food	0	6.60	4.21

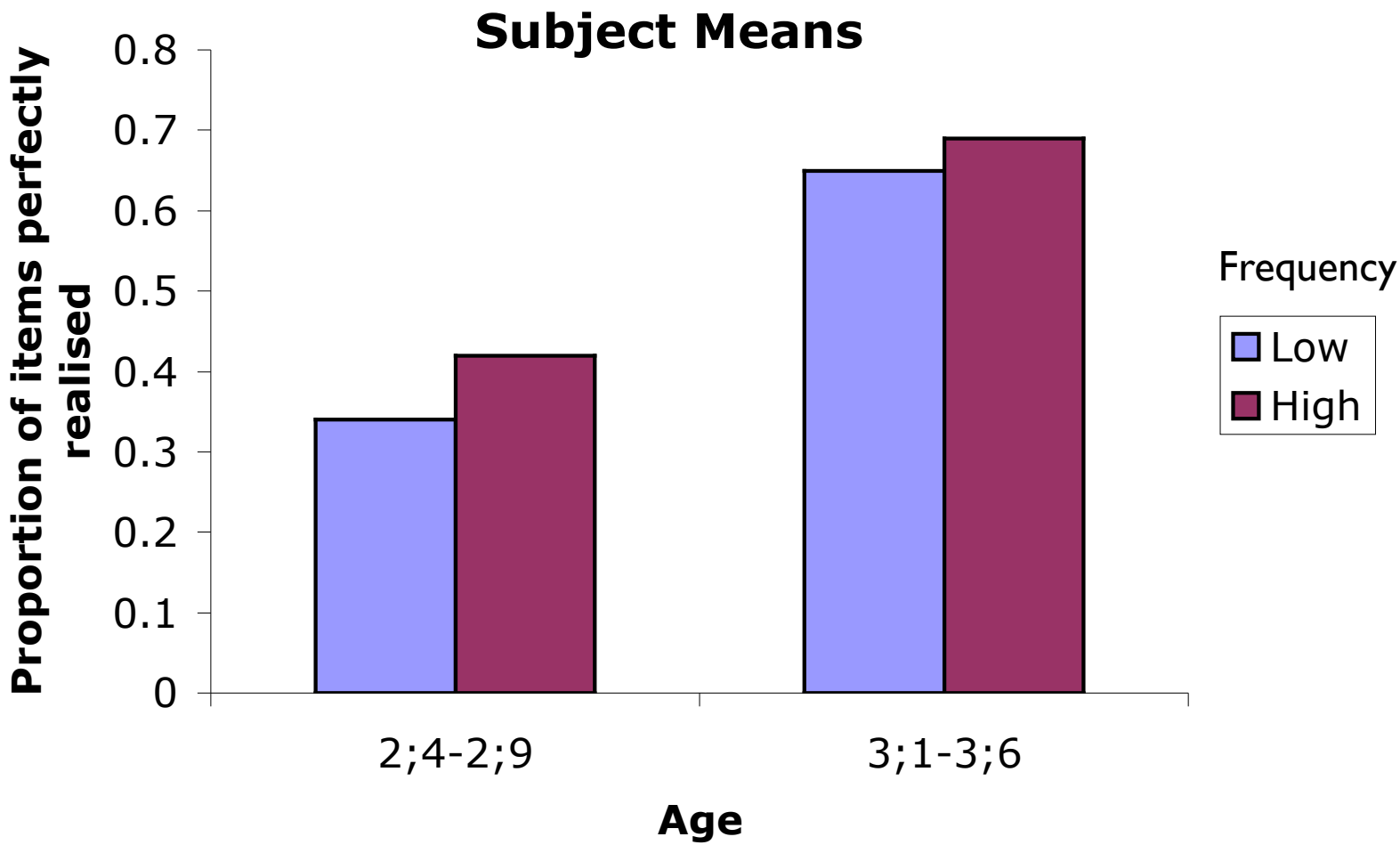
Our study

- 38 children in 2 age groups: 2;6, 3;6
- All subjects in both conditions
- 13 pairs of items
- Look at performance over first 3 words of sequences

Hypothesis

- Children will show a performance advantage for the frequent sequences:
  - Fewer errors
  - Shorter duration

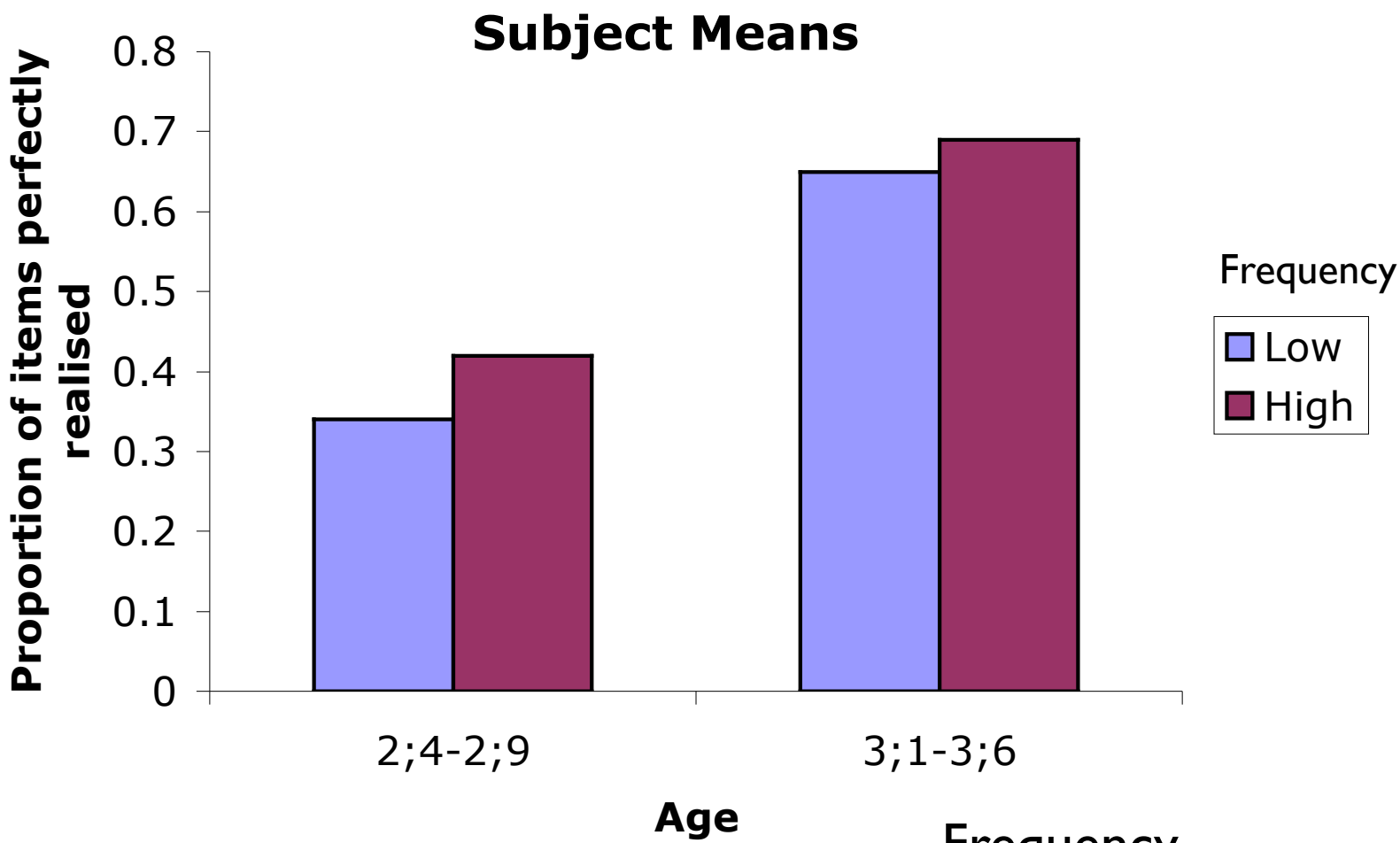
Correct repetitions



Age

F_1 (1,36) = 14.344, p < 0.001

Correct repetitions



Frequency

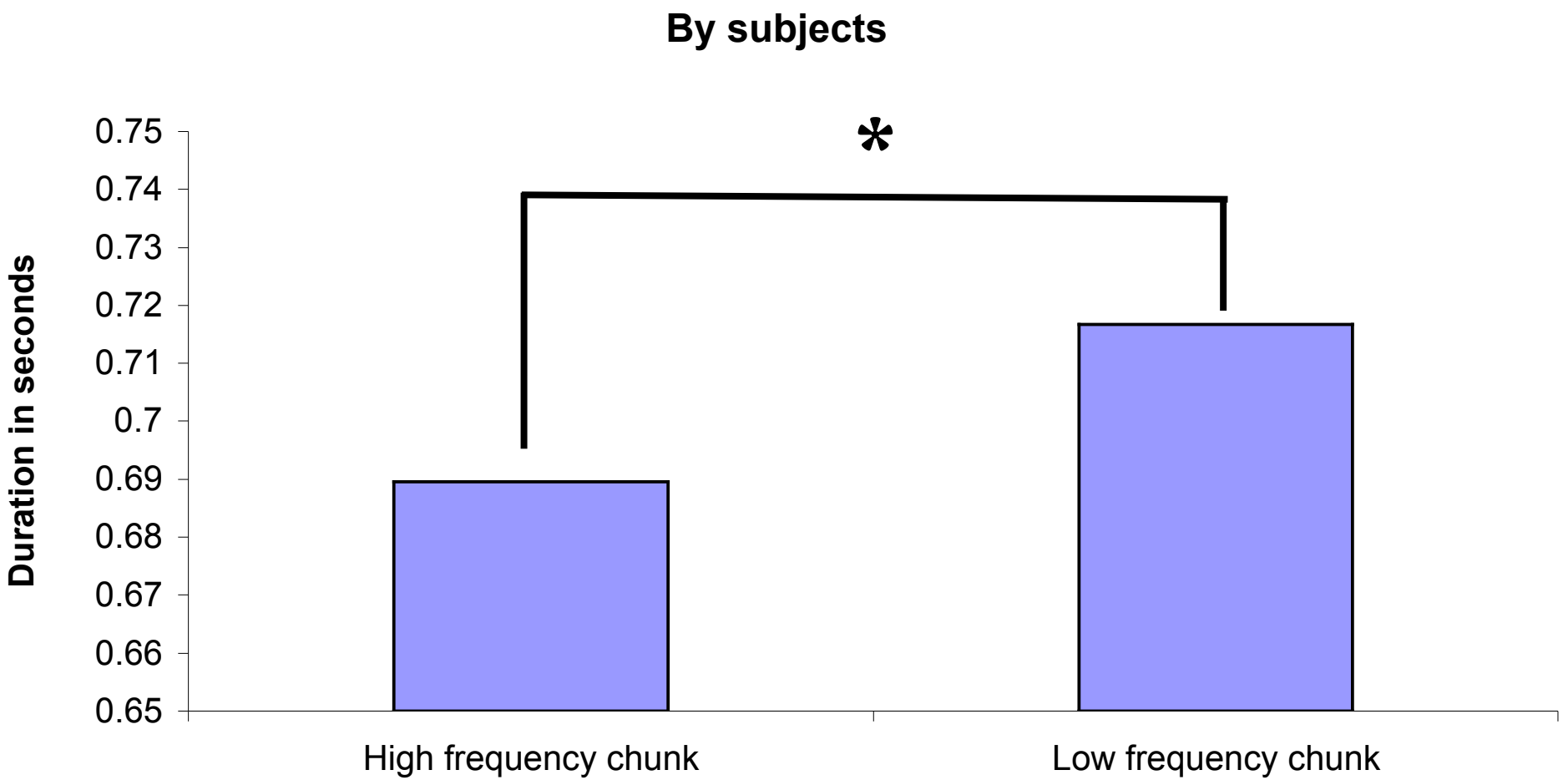
F_1 (1,36) = 6.358 p = .016

F_2 (1,24) = 3.561 p =. 071

Duration coding

- Excluded items that contained repetitions, deletions or insertions
- Measured time taken to produce first three words of chunks (the part of the utterance that was identical across conditions)

Duration


By subjects
*
0.75
0.74
0.73
0.72
0.71
0.7
0.69
0.68
0.67
0.66
0.65
Duration in seconds
High frequency chunk
Low frequency chunk

So...

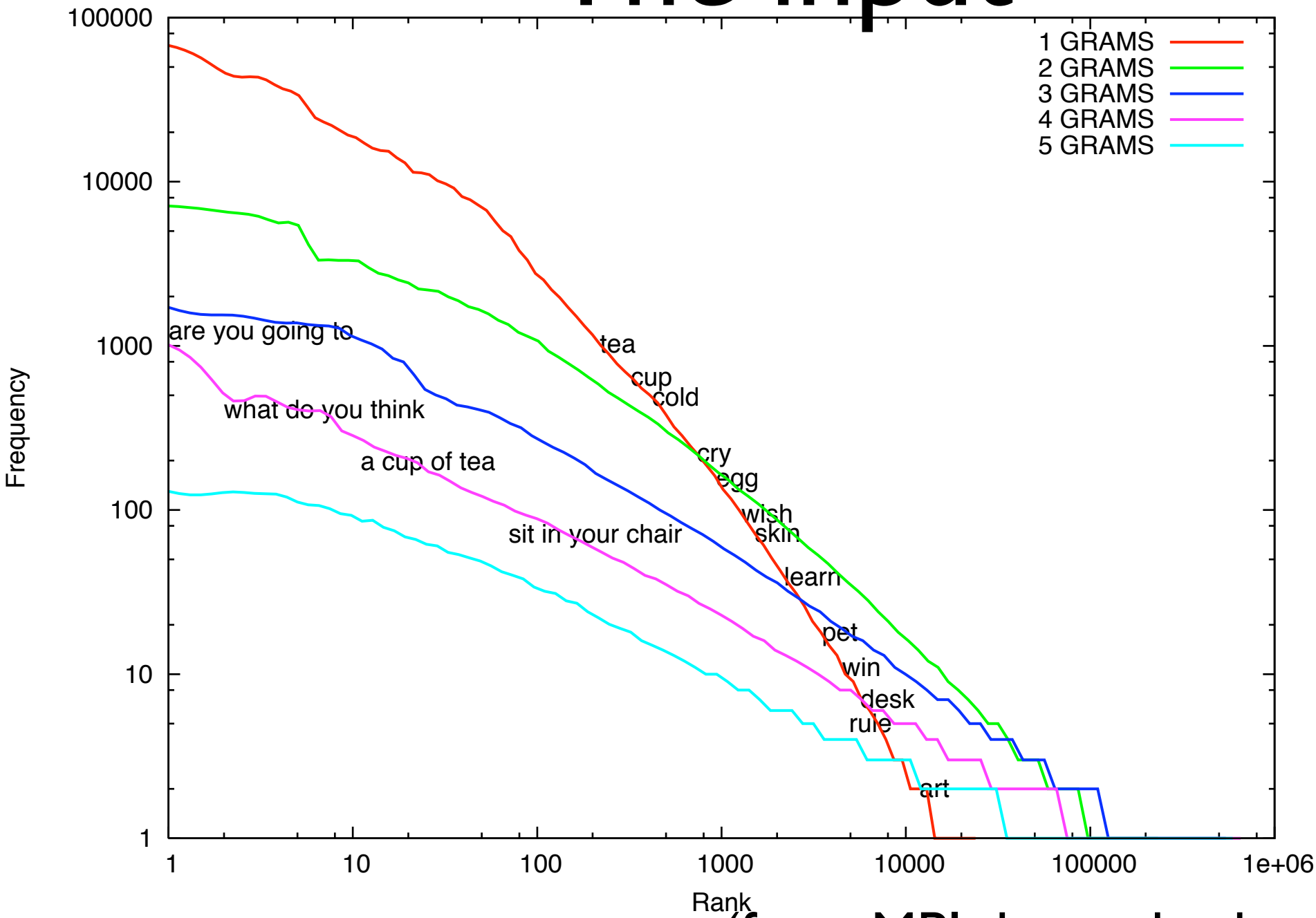
- 2 and 3 yr olds find it easier to produce frequent sequences of words than carefully matched infrequent counterparts
- 2 and 3 yr olds show memory for frequently occurring sequences of words

but
why?

Segmentation

- Children must identify and learn their basic operating units from the input
- There is evidence kids do this by tracking transitional probabilities (e.g. Saffran et al, 1996, Thiessen and Saffran, 2003):
  - Sequences with high probability transitions are units
  - Low probability transitions are unit boundaries

The input



(from MPI dense database)

Segmentation

- Learners that do not oversegment must undersegment
- Modelling work has consistently produced multiword segments (Brent, 1997; Cairns et al, 1997; Kit and Wilks, 1999)

Probabilistic segmentation produces multiword chunks

A rational argument for "low-scope grammars"

$$\frac{P(\mathcal{H}_1 \mid D)}{P(\mathcal{H}_2 \mid D)} = \frac{P(\mathcal{H}_1)}{P(\mathcal{H}_2)} \frac{P(D \mid \mathcal{H}_1)}{P(D \mid \mathcal{H}_2)}$$

- A learner will, in absence of a strong counter bias in distribution of priors, prefer low-scope models
- An effective reasoner should always prefer a model (grammar/lexicon) that spreads itself less thinly over the space of possible data

To put it more pragmatically...

- The child is trying to produce language like competent adult speakers
- They have concrete adult models
- Wouldn't it be better to imitate the adults' speech acts directly rather than make potentially flawed inferences "beyond the data"?

Imitation in language learning

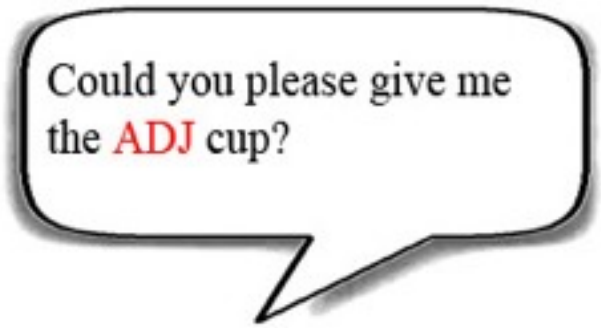
Bannard, Klinger and Tomasello (under revision)

Our question

- When will children imitate redundant linguistic material produced by an adult?
- To what extent is this determined by their perception of the adult speaker's communicative goals?

Procedure

Procedure



Procedure

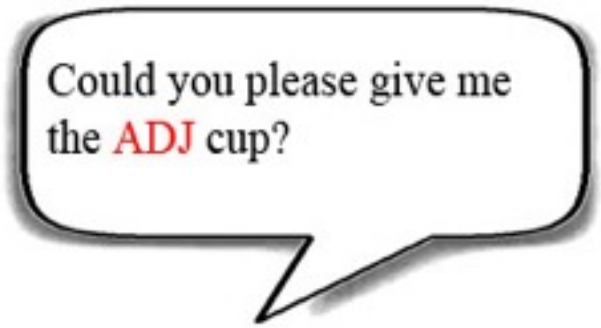
Procedure

Procedure

Procedure

Procedure

Procedure



Procedure

Procedure

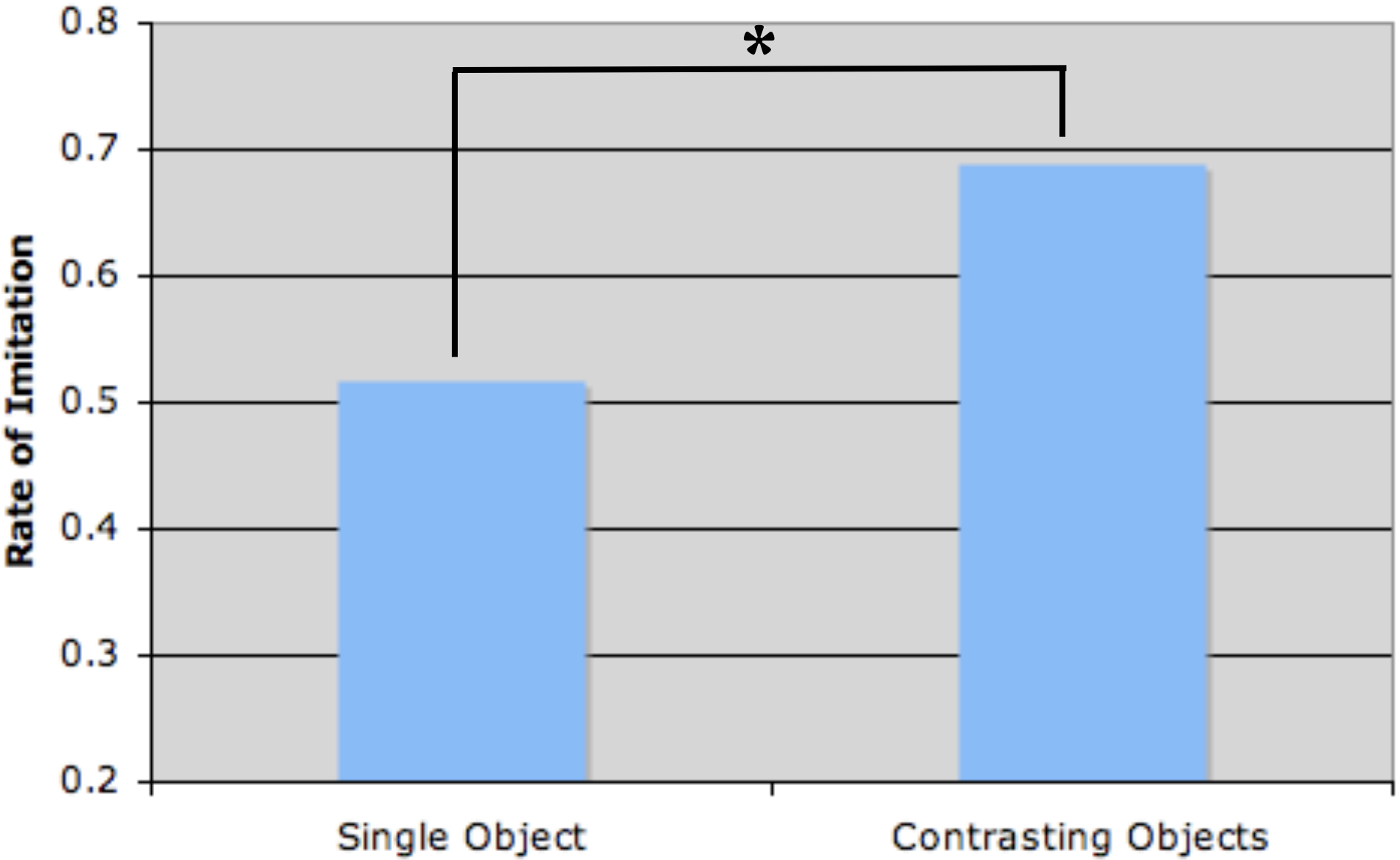
Procedure

Procedure

Will children imitate the novel adjective when there is no choice of items, i.e. when there is no communicative motivation for its use?

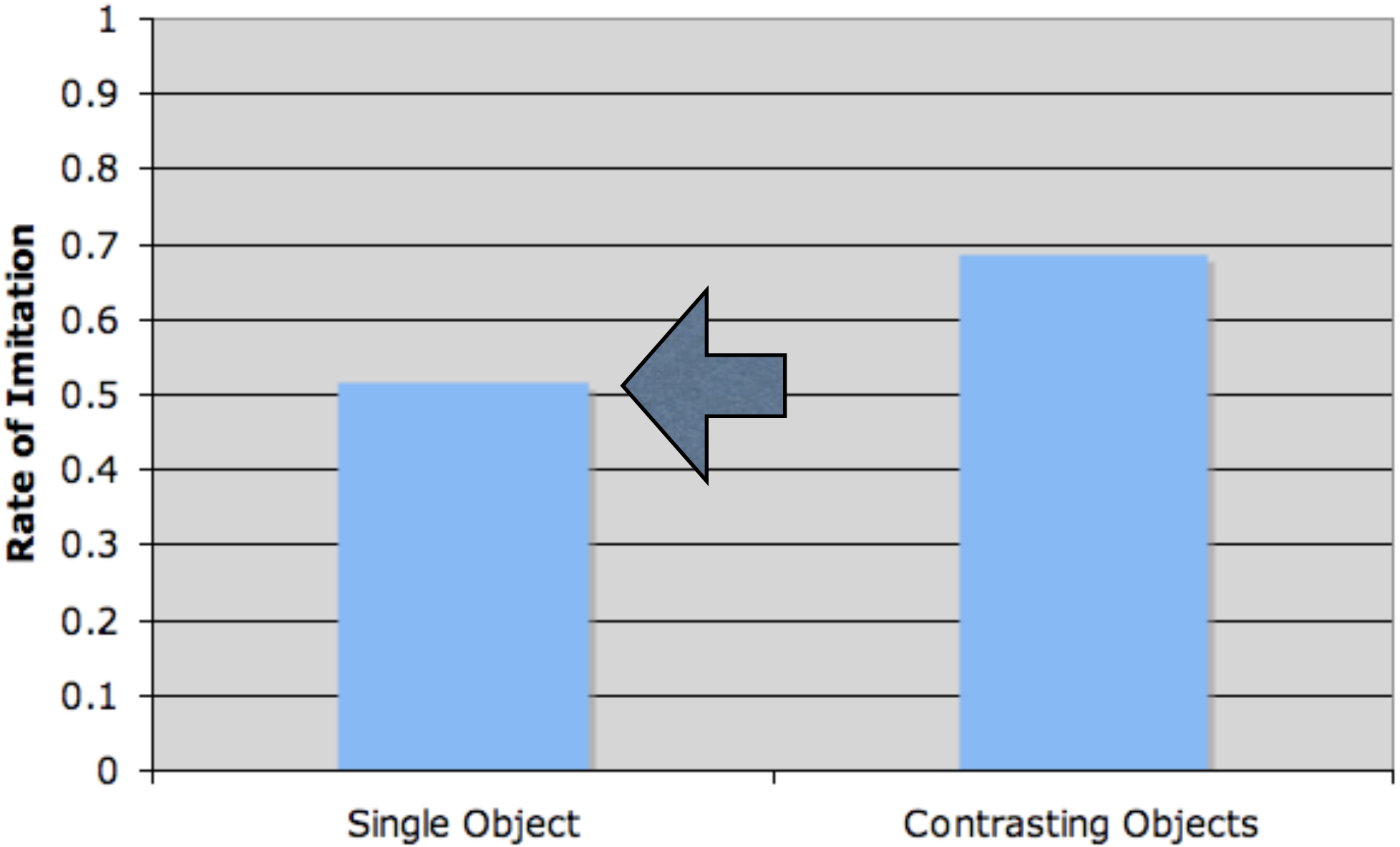
Our study

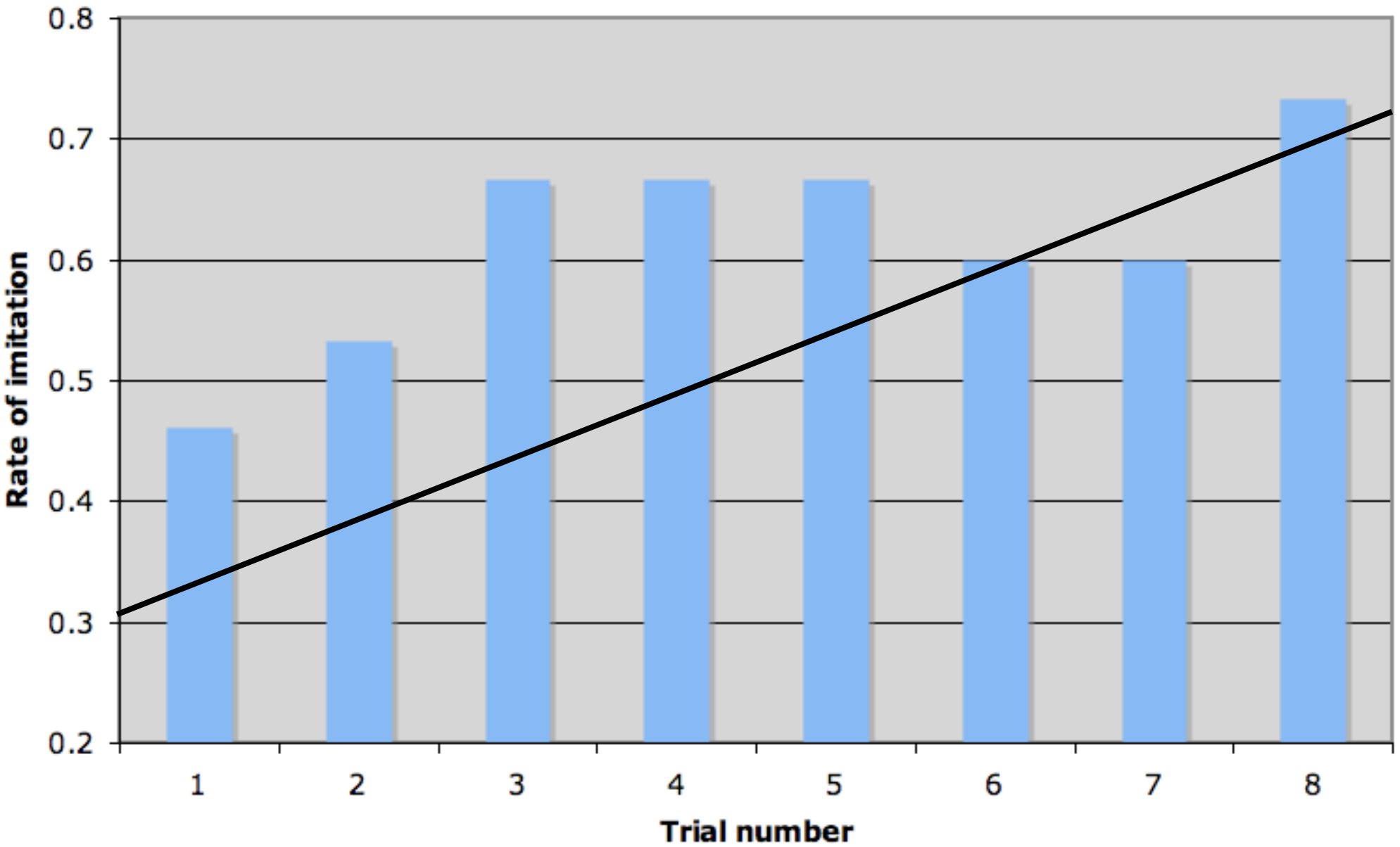
- 16 children at age 3 years & 6 months
- All subjects participated in both conditions
- 8 trials per child
- 8 novel adjectives, e.g. *dilsig*, *fatsig*


0.8
0.7
0.6
0.5
0.4
0.3
0.2
Rate of Imitation
*
Single Object
Contrasting Objects

So children pay attention to communicative goals in their imitation.

But...


Rate of Imitation
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Single Object
Contrasting Objects


0.8
0.7
0.6
0.5
0.4
0.3
0.2
Rate of imitation
1
2
3
4
5
6
7
8
Trial number

Are many children simply employing the minimum cost strategy?

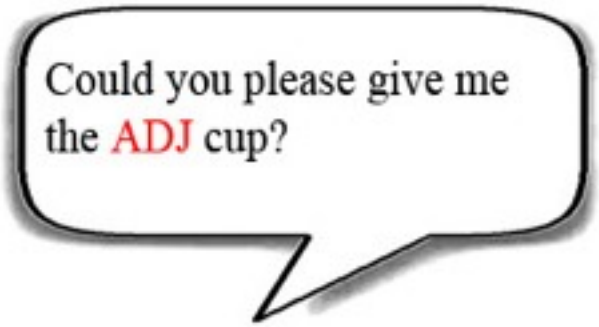
- The child is trying to produce language like competent adult speakers
- They have concrete adult models
- Wouldn't it be better to imitate the adults' speech acts directly rather than make potentially flawed inferences "beyond the data"?

Experiment 2

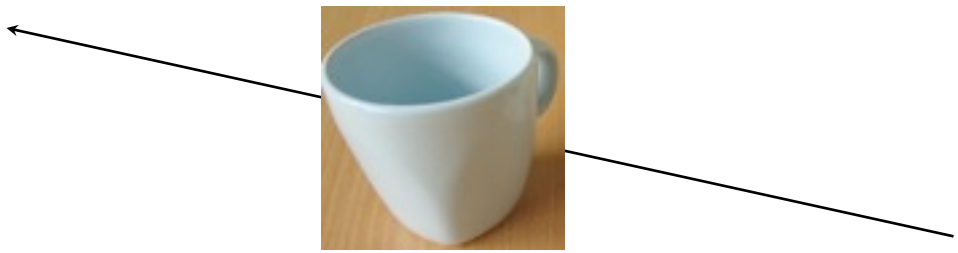
- Is children's tendency to imitate redundant linguistic material reduced when the word is marked as accidental?
- Same requesting game as in experiment one
  - A control condition identical to that in experiment one
  - An "accidental" condition in which after producing the novel adjective, experimenter one produced a gesture to indicate that the word had been uttered accidentally

Procedure

Procedure



Procedure



Procedure

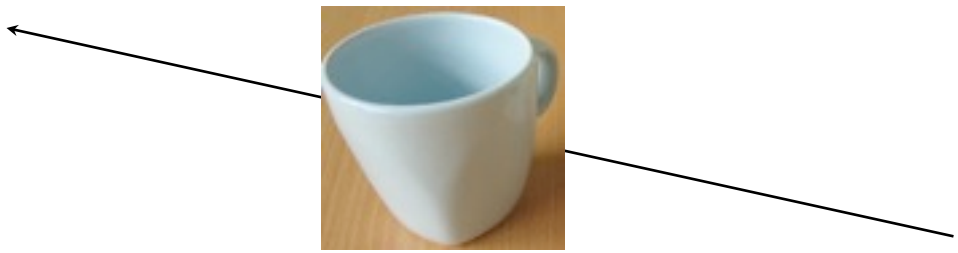
Procedure

Procedure

Procedure

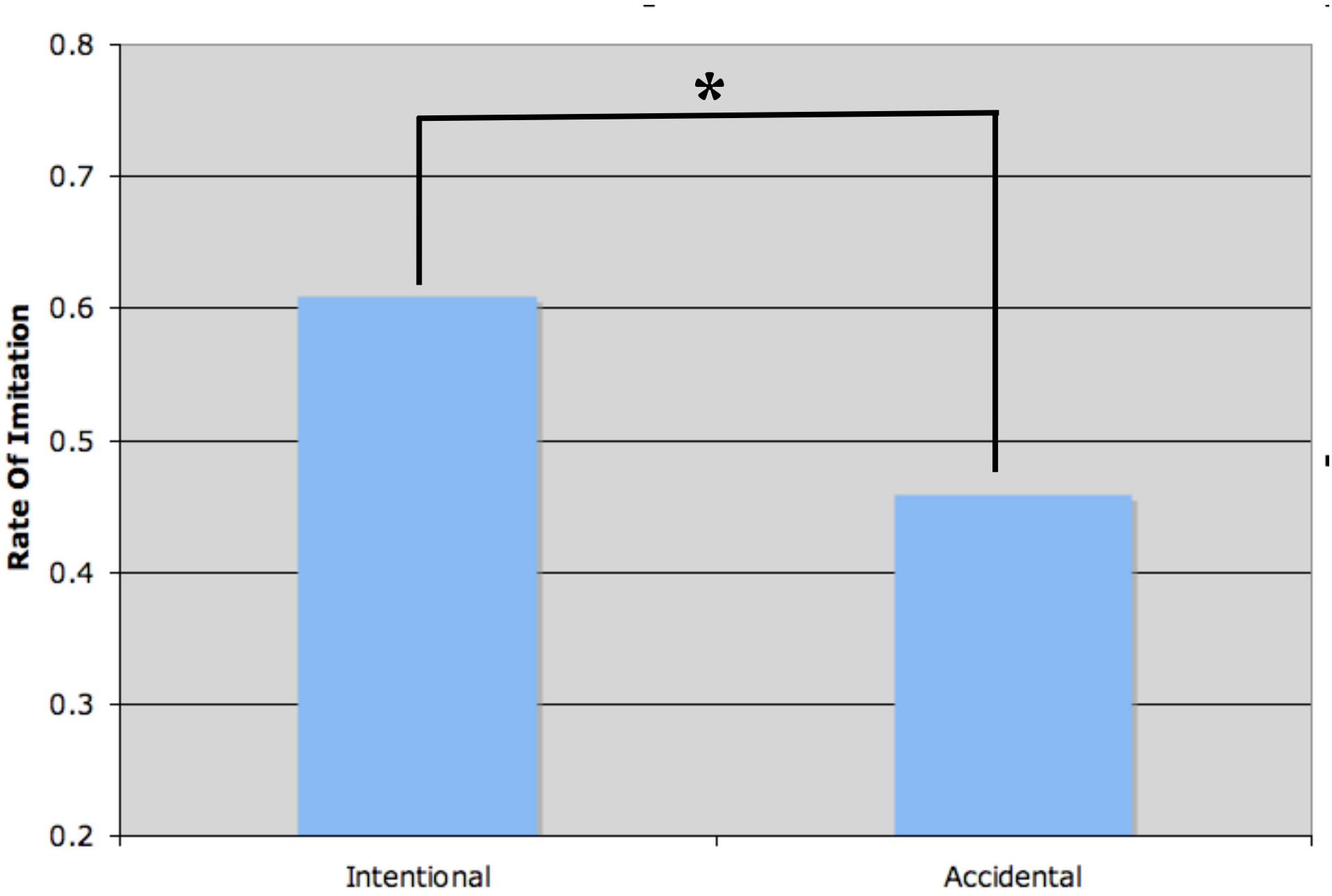
Could you please give me the ADJ <GESTURE> cup?

Procedure



Procedure

Procedure

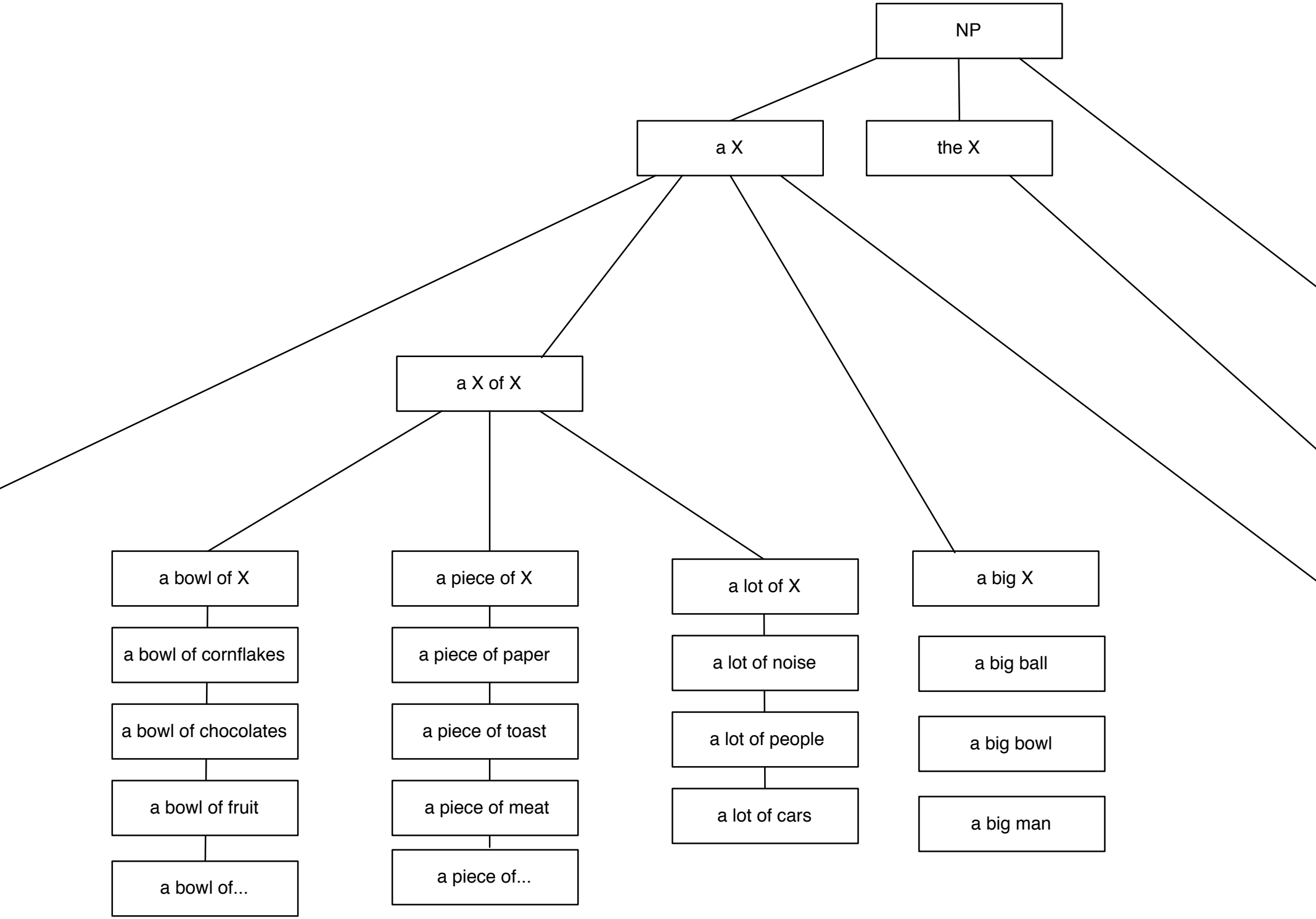

0.8
0.7
0.6
0.5
0.4
0.3
0.2
Rate Of Imitation
*
Intentional
Accidental

In conclusion...

- 3-year-olds are more likely to imitate novel linguistic material when it serves a communicative function
- However they still imitate redundant material about half the time!
- 3-year olds imitate redundant material less when it is marked as accidental (or perhaps more when it is marked by contrast as intentional), suggesting that their copying is strategic even when blind

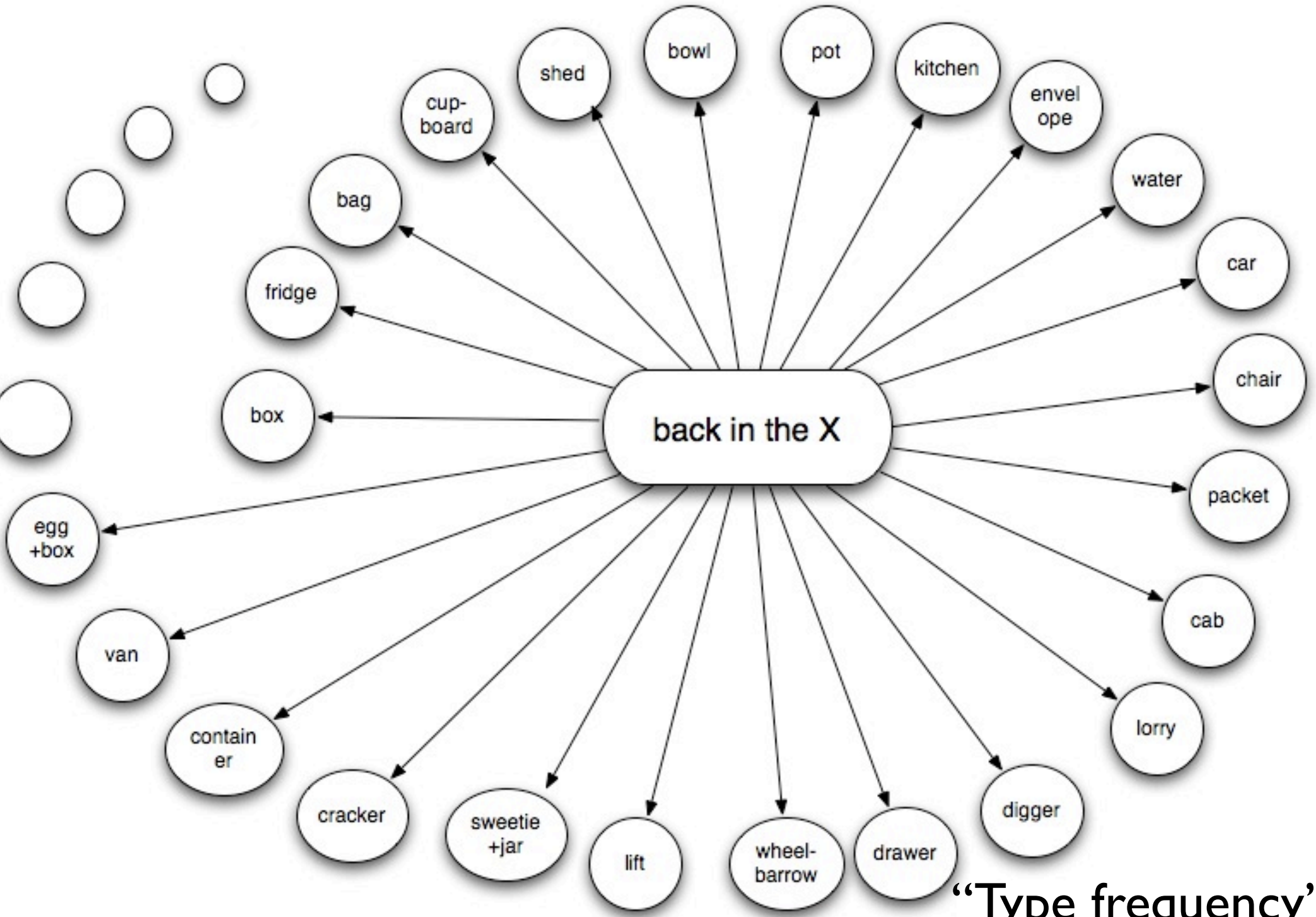
So...

- Children hug the input, reusing exact material heard where available
- But they cannot ONLY imitate.
- A prediction: Children should be more likely to form generalizations where the input does not provide imitation-based shortcuts: where there is uncertainty in the speech stream


NP
a X
the X
a X of X
a bowl of X
a bowl of cornflakes
a bowl of chocolates
a bowl of fruit
a bowl of...
a piece of X
a piece of paper
a piece of toast
a piece of meat
a piece of...
a lot of X
a lot of noise
a lot of people
a lot of cars
a big X
a big ball
a big bowl
a big man

Can we use corpora to predict when children will exploit generalizations?

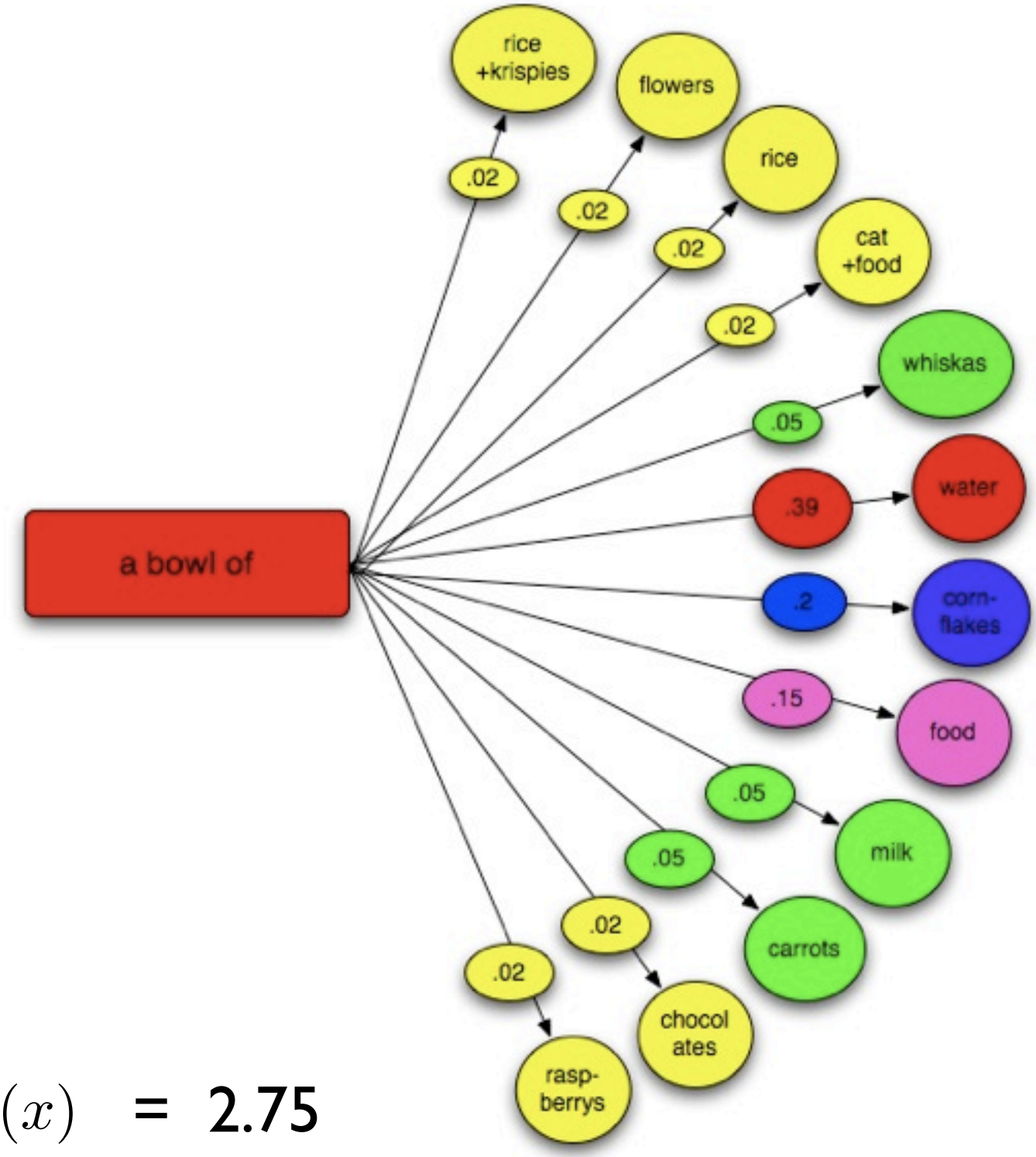
Matthews and Bannard (2010, *Cognitive Science*)



"Type frequency" = 89

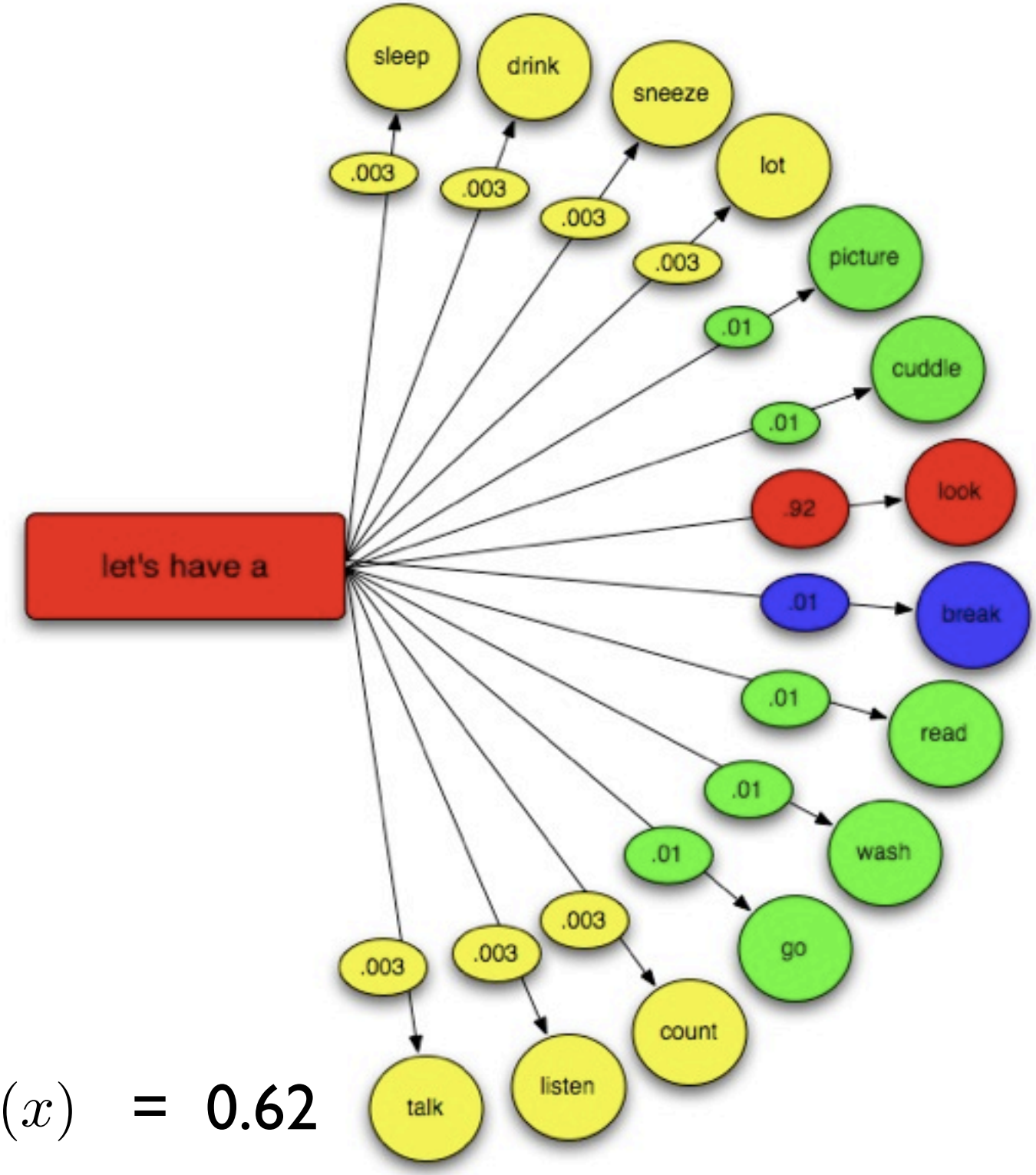
Slot entropy

- A measure of uncertainty over the set of words seen in a particular position



slot entropy =

$$\sum_{x \in X} p(x)\ log_2\ p(x) \quad = \quad 2.75$$

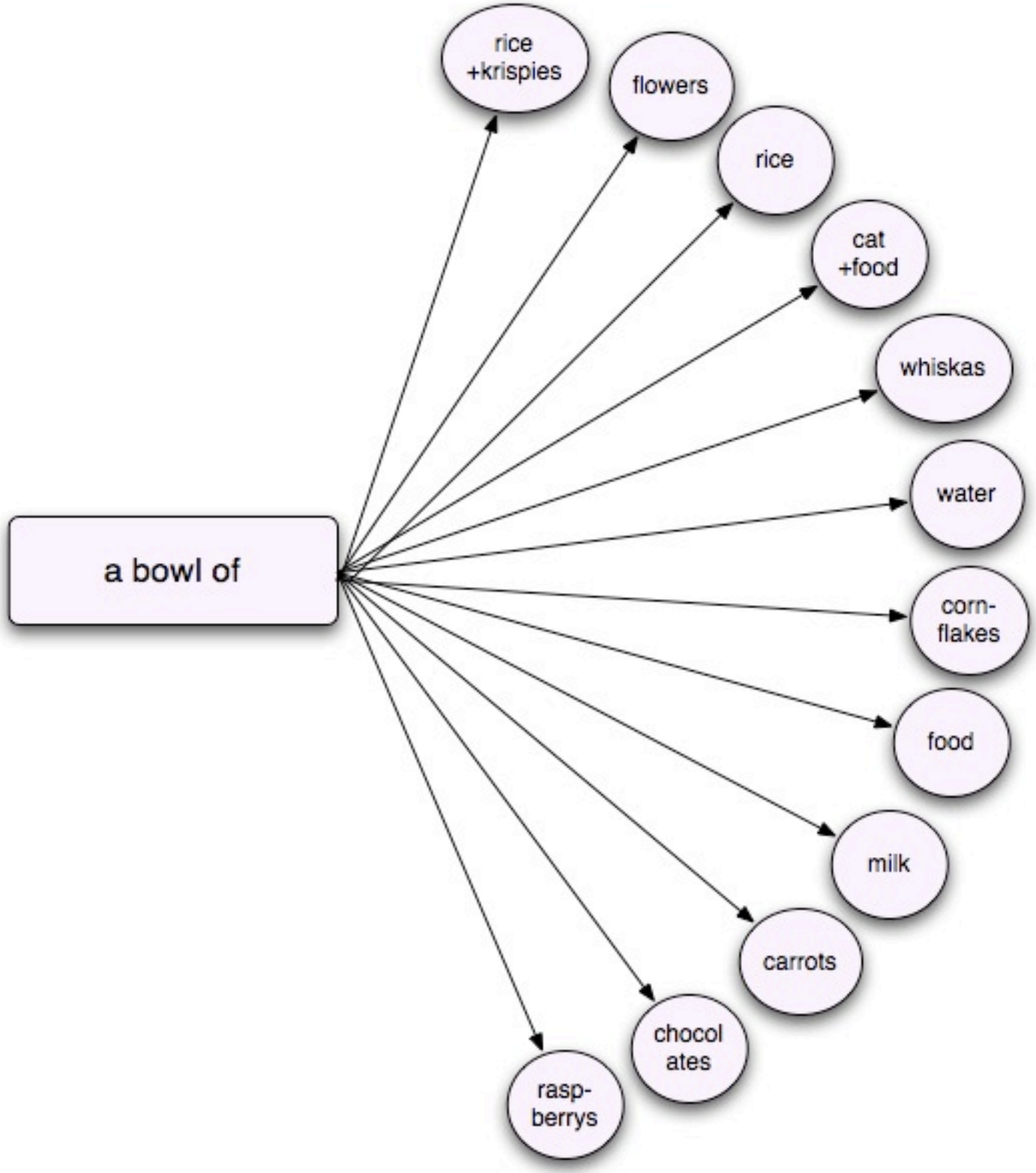


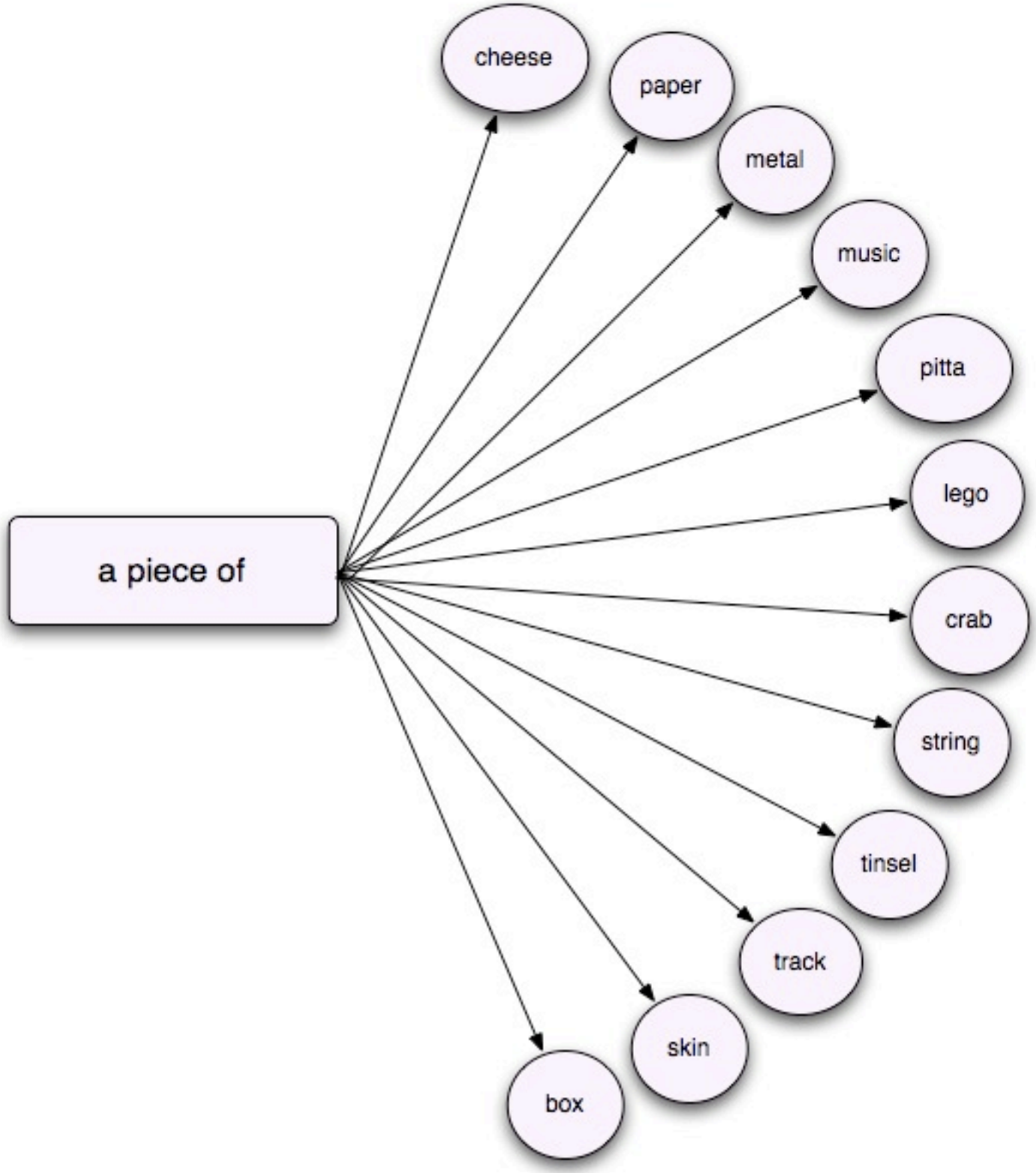
slot entropy =

$$\sum_{x \in X} p(x)\ log_2\ p(x) \quad = \quad 0.62$$

A second prediction

- Children will extract a schematic pattern if they encounter many similar things in the relevant position

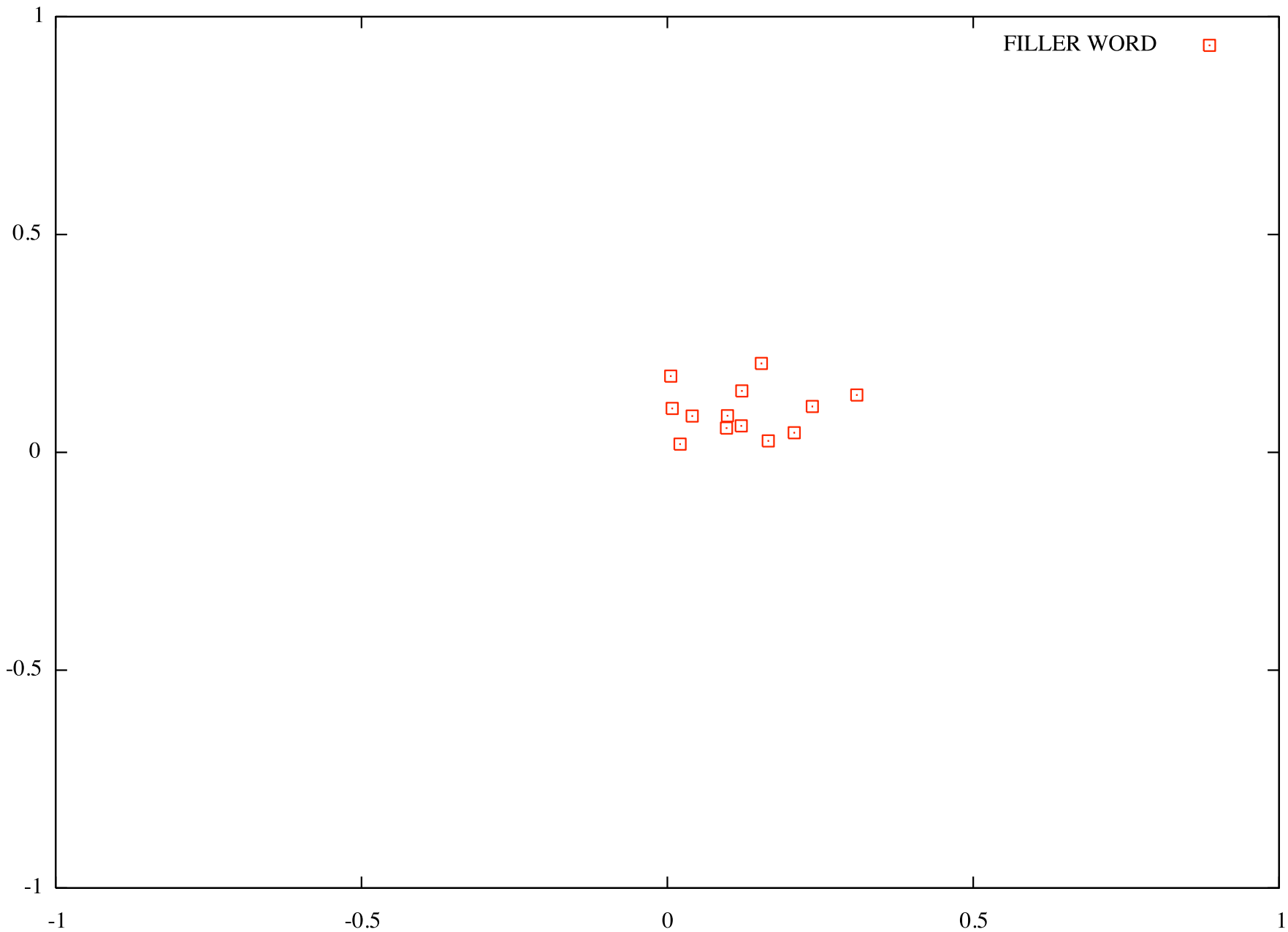

a bowl of
rice +krispies
flowers
rice
cat +food
whiskas
water
corn-flakes
food
milk
carrots
chocol ates
rasp-berrys


a piece of
cheese
paper
metal
music
pitta
lego
crab
string
tinsel
track
skin
box

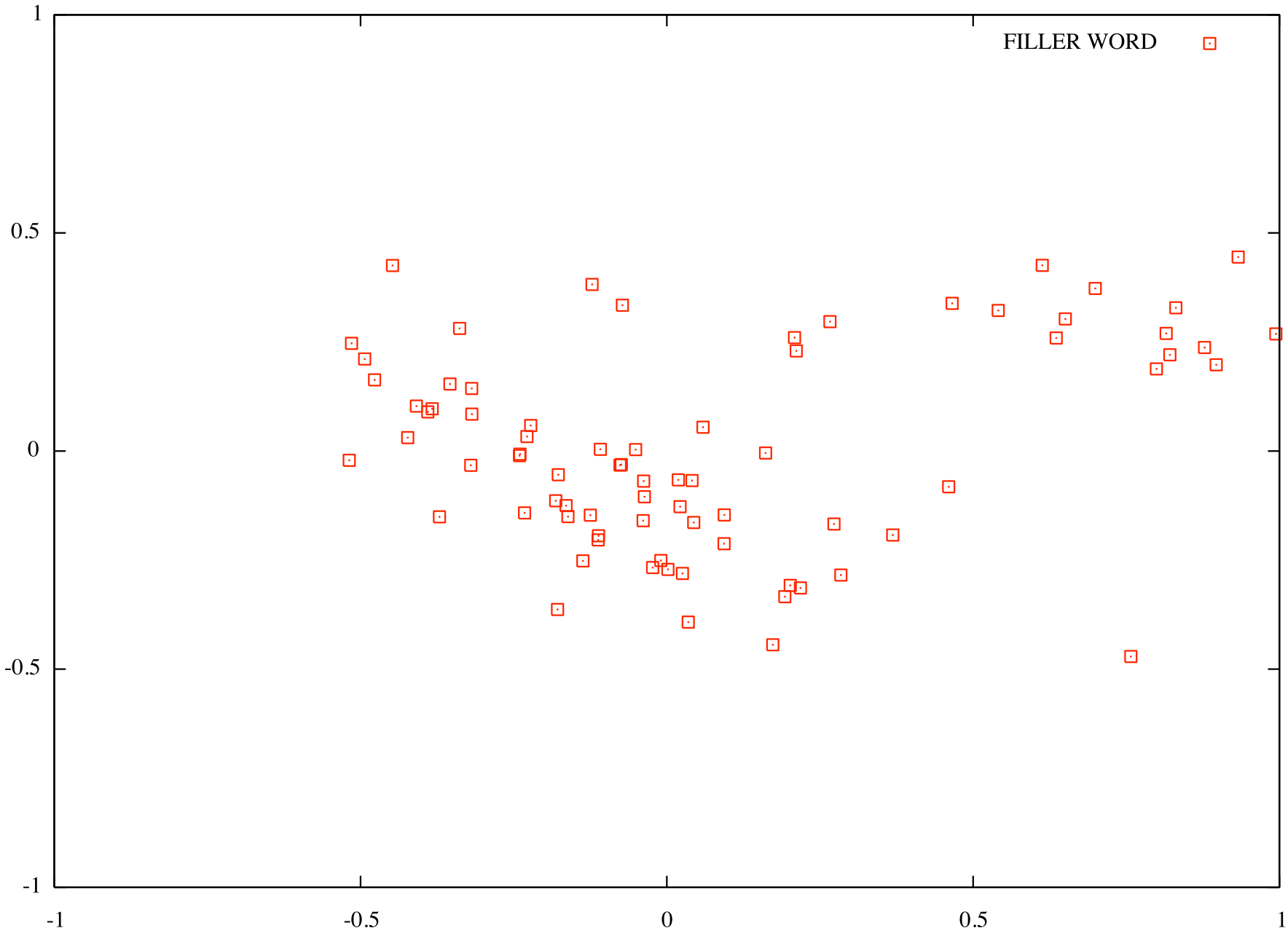
How we operationalise "similarity"?

- Use information we know is available to children - co-occurrence statistics in input
- Items are taken to be similar if they occur in similar contexts over our corpus
- We know that children and young as 12 months form categories of items that are distributionally similar to one another (Gomez & Lakusta, 2004)

A BOWL OF X



A PIECE OF X



Our study

- A repetition study using 4 word sequences
- Children asked to repeat novel instantiations of patterns which vary according to the factors described, e.g.
  - a piece of meat / a piece of brick
  - a bowl of biscuits / a bowl of flowers
- We coded the child's production of the first three words of the sequence for errors

Our study

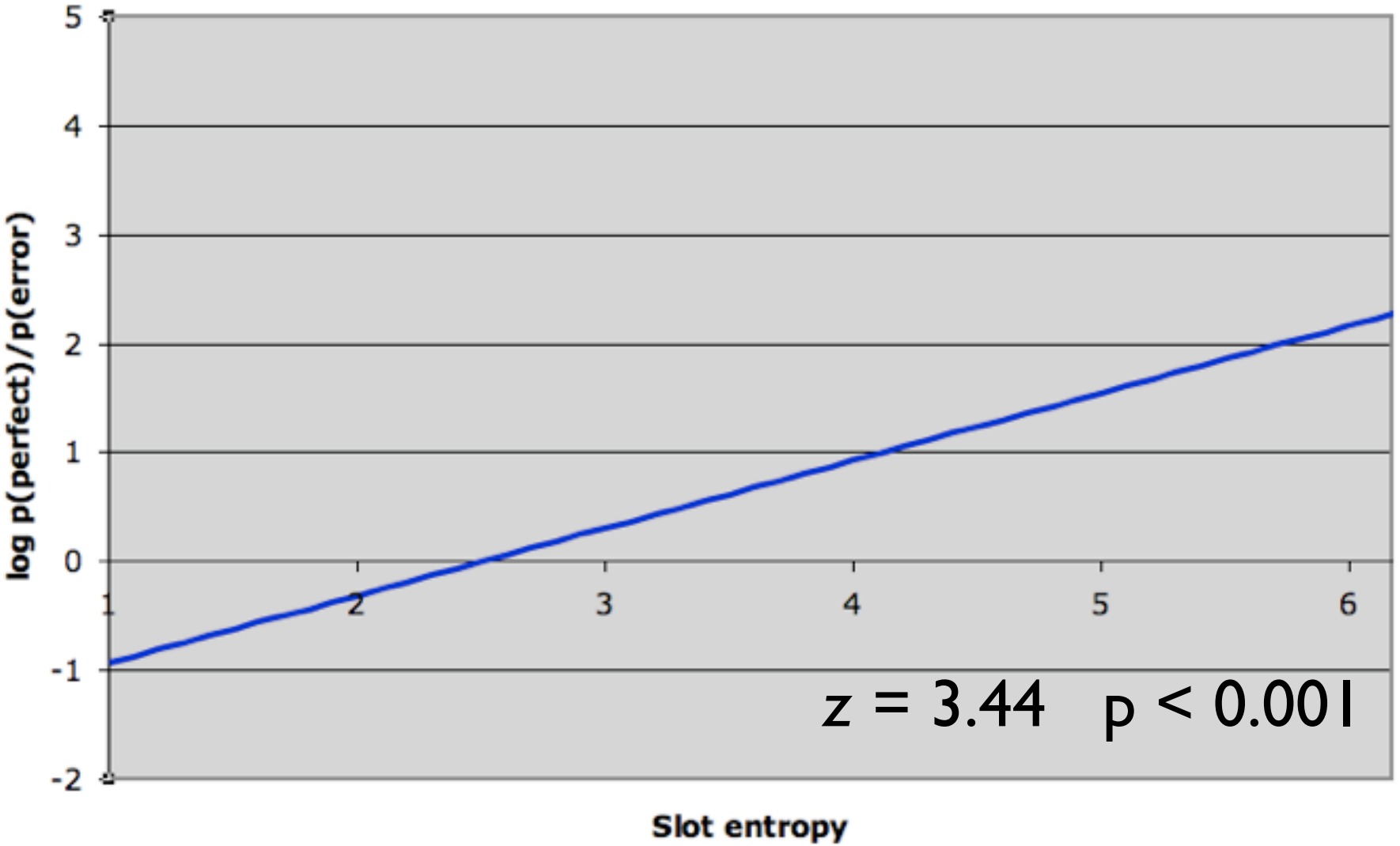
- 59 children: 2;6, 3;6
- Within-subjects
- 9 schematic patterns (18 items in total)

Hypothesis

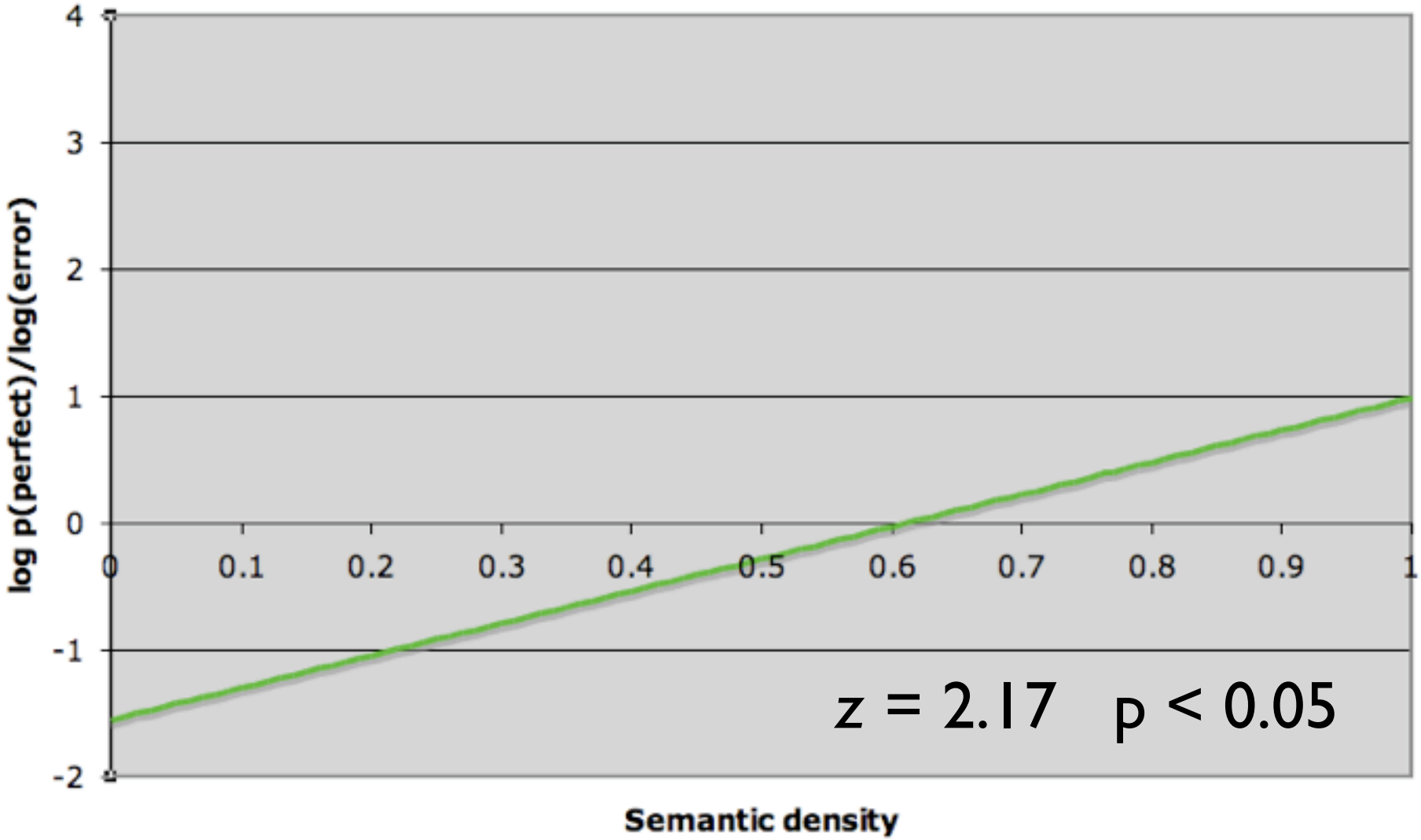
- Children should be less likely to make errors when:
  - the target utterance instantiates a pattern that has high "slot entropy"
  - the target utterance instantiates a pattern that has high "semantic density"

Analysis

- Mixed effects logistic regression model
  - Success of realization as outcome variable
  - Subject included as a random effect
  - Slot entropy and slot density included as fixed effects
  - Frequencies of all target phrases and component word and sequence frequencies reduced to 4 dimensions using PCA and included as covariates



Probability of error free production increases with slot entropy



Probability of error free production increases with semantic density

The emerging picture

- Children are highly strategic in their early language productions:
  - They represent and reuse frequently encountered sequences of words
  - They imitate adult utterances in order to achieve known communicative ends, even when blind to the mapping between means and end
  - They are more likely to form and exploit generalizations where they encounter uncertainty, suggesting a preference for reuse where available

Thanks to

Danielle Matthews, Elena Lieven, Michael Tomasello, Joern Klinger, Jess Butcher, Roger Mundry, Anna Roby, Manuel Schrepfer, Elizabeth Wills, Kathrin Greve, Michael Ramscar. And of course all the children who take part in our studies.