Learning Gender Classes in Italian via a Computerized Word Learning Game: A study with 7 year olds

Dr Elizabeth Wonnacott, Dr Helen Brown & Miss Lydia Gunning

Department of Language and Cognition,

Division of Psychology and Language Sciences

30 March 2017

Background: Grammatical Gender

- Many languages divide words into “gender”(masculine, feminine, neuter)
- Gender classes are notoriously difficult for second language learners where the first language doesn’t contain gender classes
- In contrast: seems to present relatively little difficulty in first language acquisition
  - 3 year olds use gender predictively pattern (e.g. Lew-Williams & Fernald 2007)

la pelota

Faster to look at the target word when different gender from the foil

el zapato

la galleta

“Encuentra la pelota. ¿La ves?”

Find the ball. Do you see it?

Background: Grammatical Gender

- Many languages divide words into “gender”(masculine, feminine, neuter)
- Gender classes are notoriously difficult for second language learners where the first language doesn’t contain gender classes
- In contrast: seems to present relatively little difficulty in first language acquisition
  - 3 year olds use gender predictively pattern (e.g. Lew-Williams & Fernald 2007)
- Second language: even highly proficient adults may not show this pattern (e.g. Gruter, Lew-Williams & Fernald, 2012)

Statistical Learning Literature

- Focused on the problem of how the language learning system extracts gender classes from the input without explicit teaching.

Statistical Learning Literature

environment 1

un…

word1
word2
word3
word4

environment 3

une…

word5
word6
word7
word8

environment 2

le…

environments….

environment 4

la…

environments….

Statistical Learning Literature

environment 1

der__-er

CLASS ONE
word1
word2
word3
word4

environment 3

die_-e

CLASS TWO
word5
word6
word7
word8

environment 2

dem__-em

environments….

environment 4

der_-er

environments….

Various experiments with computational and human (adult) learners

Statistical Learning Literature

- Mintz (2002): Learners hear sentences from an artificial language:

"bol nex jiv" and *"choon pux wug"*

environments	CLASS ONE
bol___jiv	nex
zim___noof	kwob
poz___fen	zich
	pren

environment	CLASS TWO
choon___wug	pux
	daik
	fend
	plif

- Test: Judge grammaticality of **novel** combinations

poz zich fen ↑ correct class

poz plif fen ↑ incorrect class

Statistical Learning Literature

- Gomez (2002): Infants exposed to three word strings like

"pel wadim rud" and *"dak kickey tood"*

pel___rud
dak___tood
vot___jic

Manipulate number of possible words in the middle slot
3, 12 or 24

"frames" →

wadim
kickey
fengle
coomo..

Infants only discriminate when hear the frames with 24 middle words
→ *benefit from more varied examples*

- Test: Judge **novelty** of sentences

pel wadim rud

pel wadim jic

Our Experiment

- **Native English speaking 7** years olds learning **number** and **gender marking** in Italian noun-phrases (no previous experience of Italian)
- Can children pick up on gender marking from a **computerized word learning game** (no explicit teaching of gender)
- Do they benefit from “variability” i.e. is it helpful to exemplify the gender classes more example nouns
- Also want to look for any evidence of automatized processing (as seen in native speakers)

Italian definite noun-phrases

masculine singular:	il	$noun_{[masculine]}$ + o	**il** lett**o**
masculine plural:	i	$noun_{[masculine]}$ +i	**i** lett**i**
feminine singular:	la	$noun_{[feminine]}$ +a	**la** ballen**a**
feminine plural:	le	$noun_{[feminine]}$ +e	**le** ballen**e**

"frame"

Variability Manipulation

12-nouns (HV): 6 masculine 6 feminine

4-nouns (LV): 2 masculine 2 feminine

Methods

Participants

- Thirty 7-8 year olds (M=7yrs, 10 months)
  - ➢15 in 12-nouns (HV)
  - ➢15 in 4-nouns (LV)

Day 1
- ➢ exposure (84)

Day 2
- ➢ exposure (168)

Day 3
- ➢ exposure (84)
- ➢ tests

Day 4
- ➢ exposure (168)

Day 5
- ➢ exposure (84)
- ➢ tests

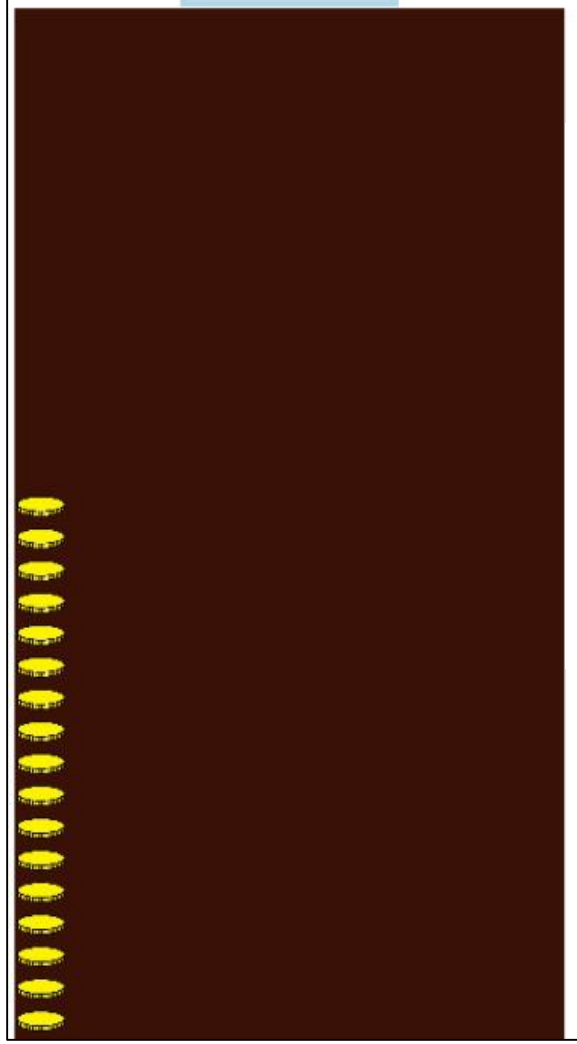
12-nouns: 12 nouns * 7
4-nouns : 4 nouns * 21

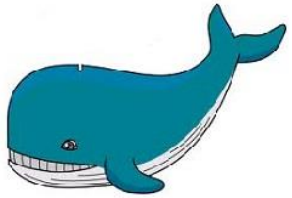
TODAY

2 AFC test
Speeded noun ID task
Production test

Training

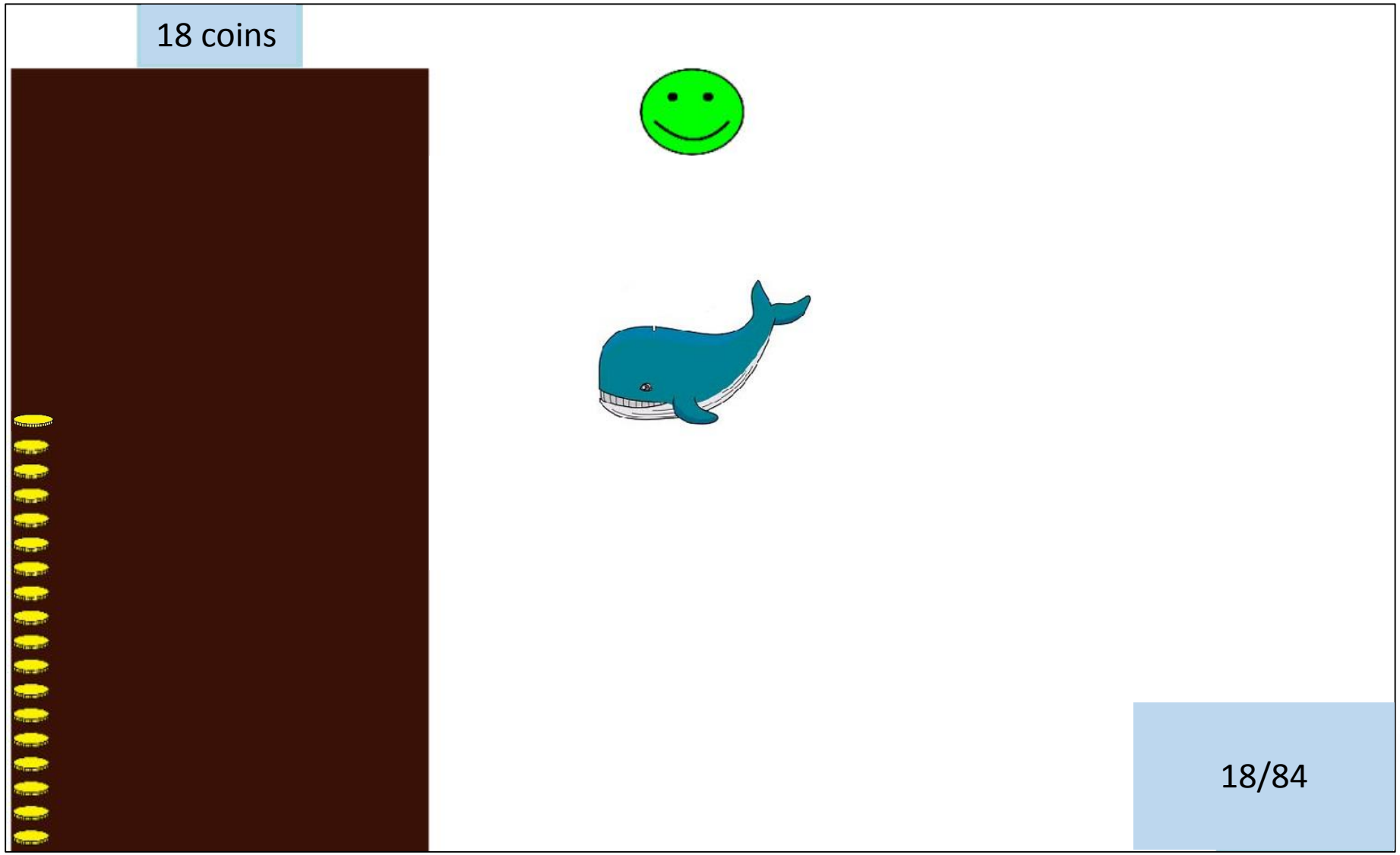
17 coins





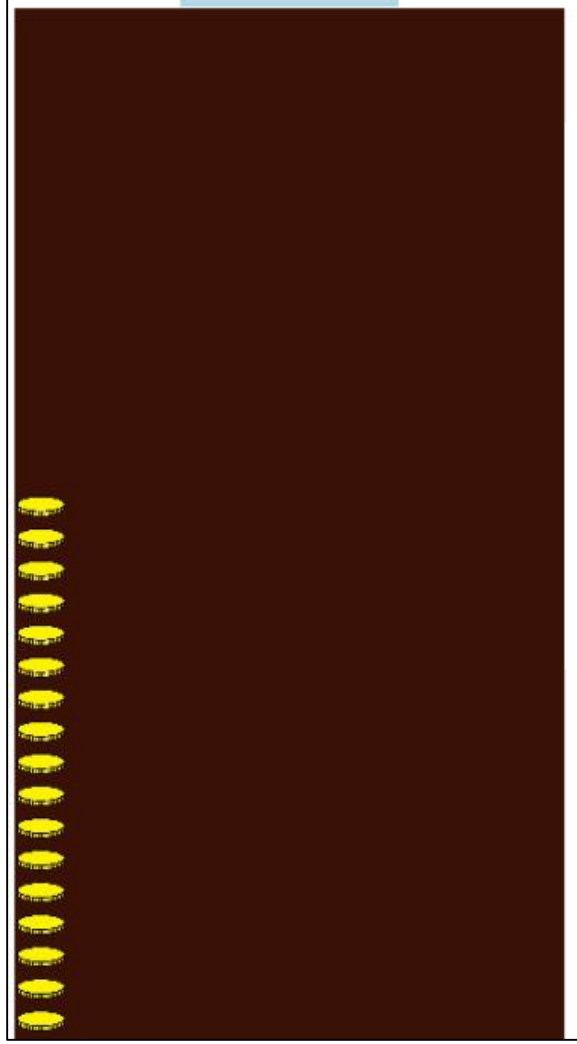
17/84

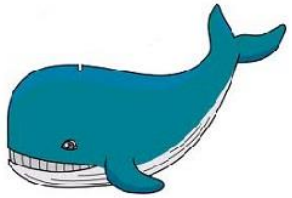
Training



Training

17 coins

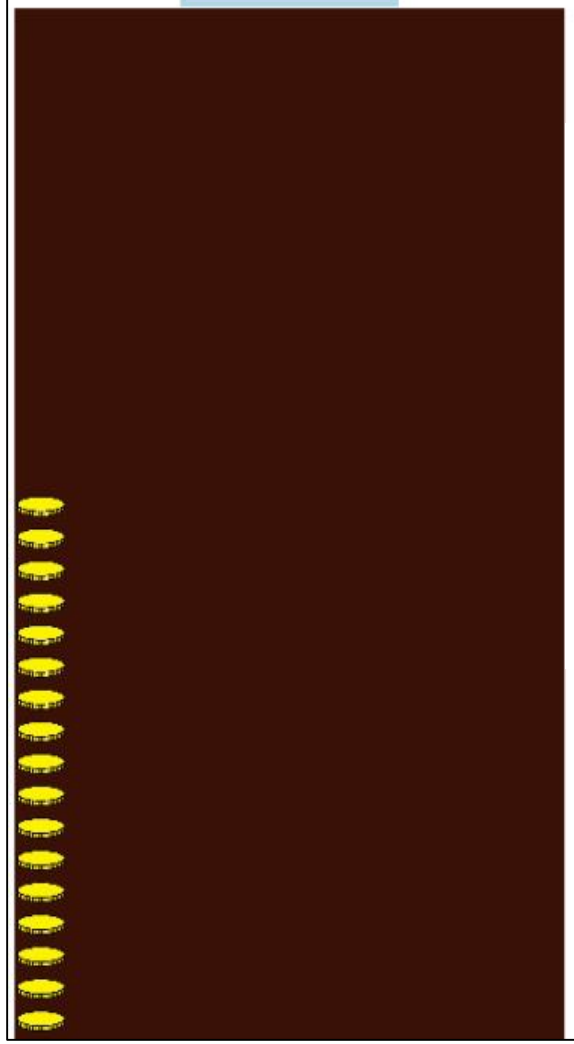


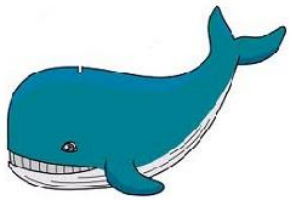


17/84

Training

17 coins

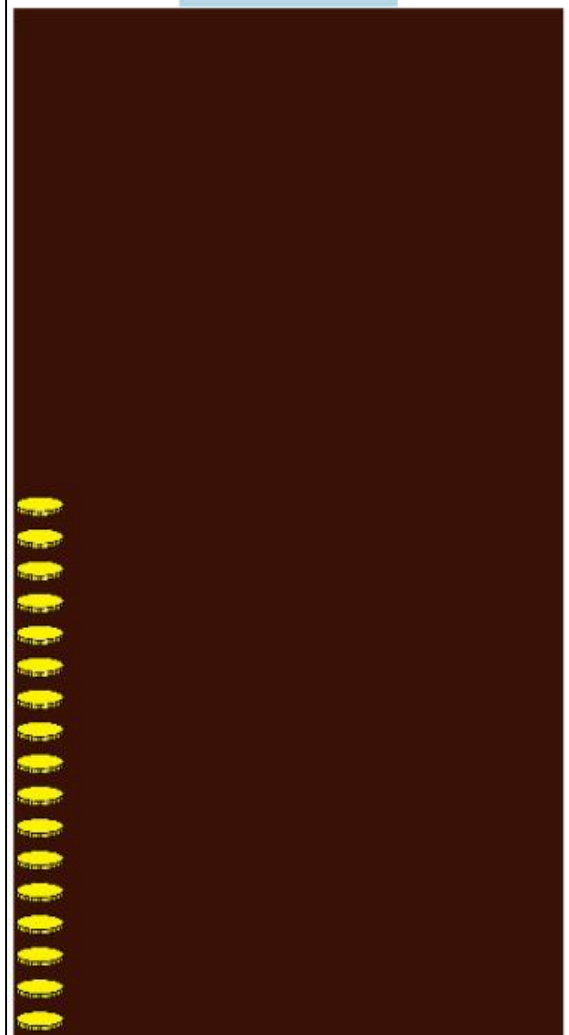




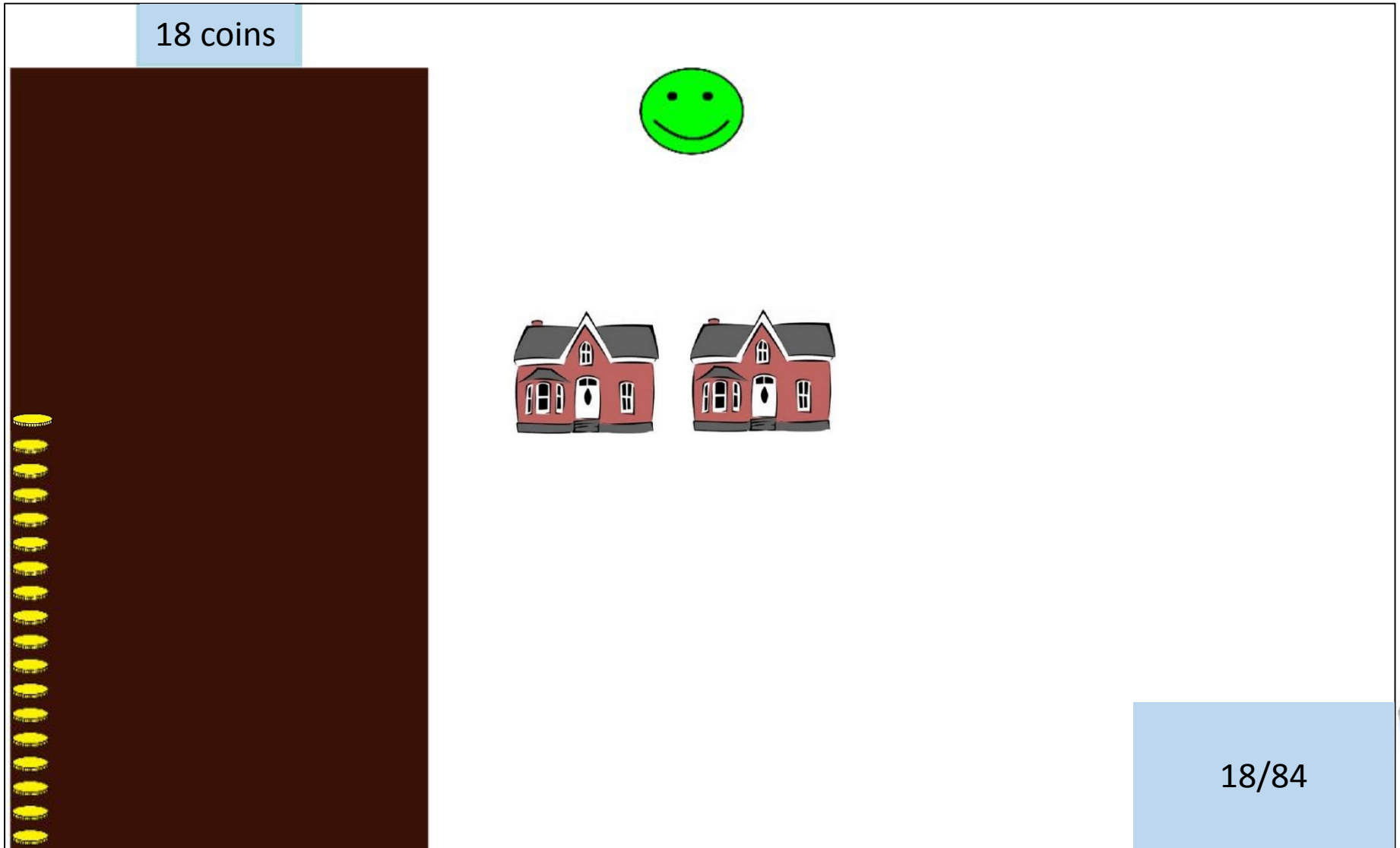
18/84

Training

17 coins



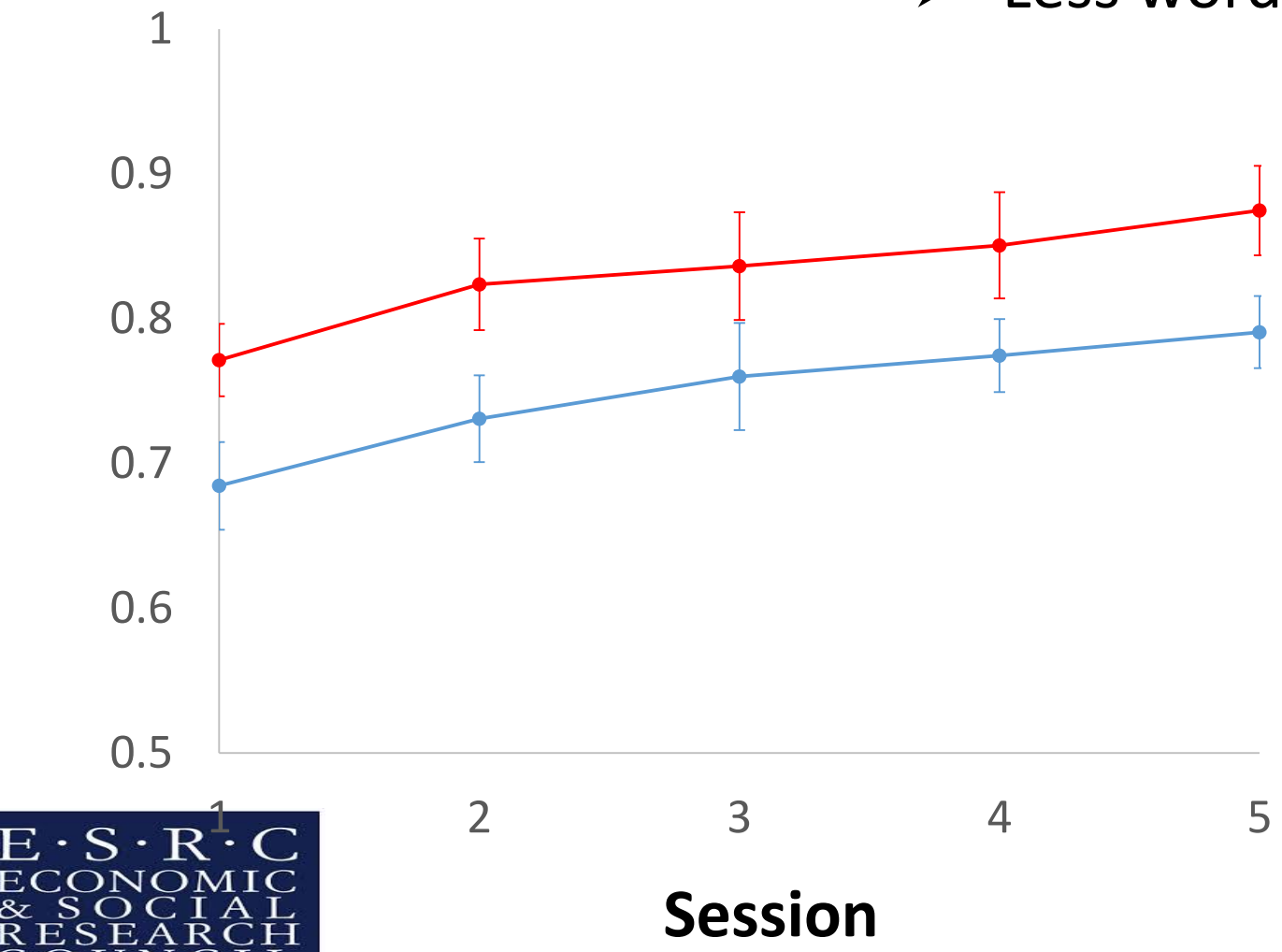
Training



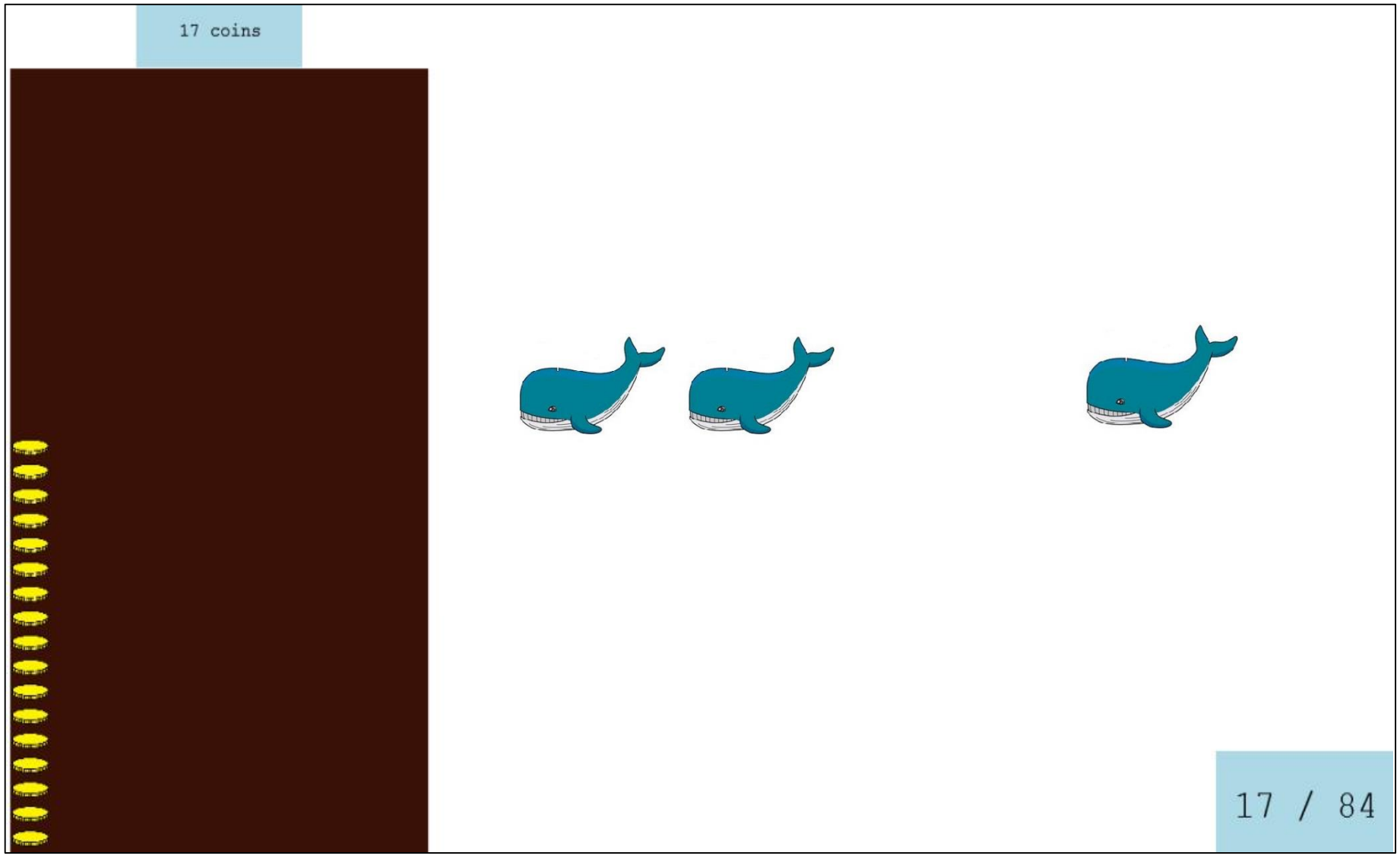
Training Data

Better learning in low variability condition

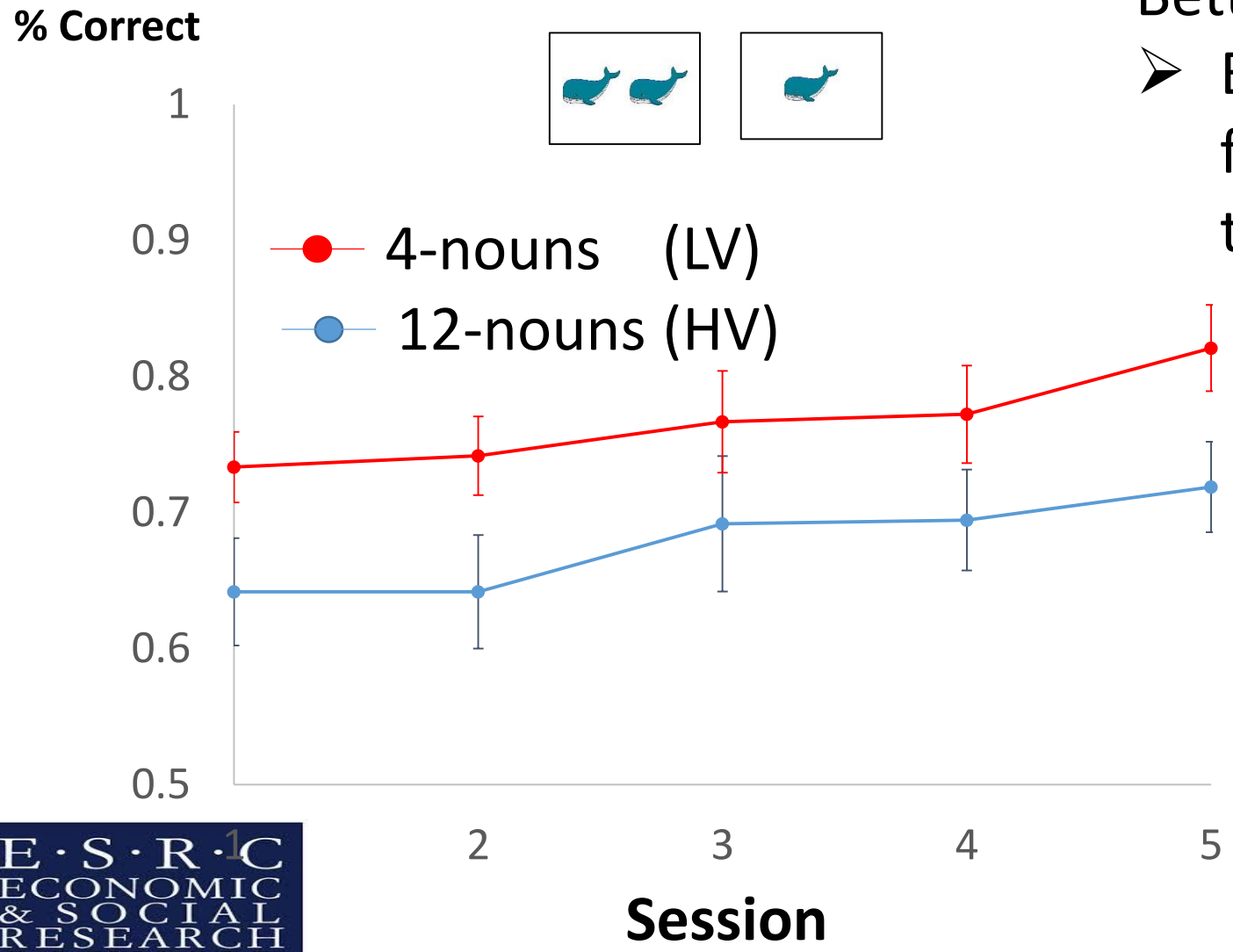
- Less words to learn



Training



Training Data (2)



Better learning in the 4-nouns condition

- Easier to learn the singular/plural forms in the context of two nouns than in the context of six nouns

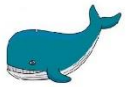
BENEFIT FROM LESS VARIED EXAMPLES

*Easier to see the link between the number **marking** and number **semantics** when repeatedly hear with the same lexical items*

Speeded Noun Identification

target & foil DIFFERENT gender

target & foil SAME gender



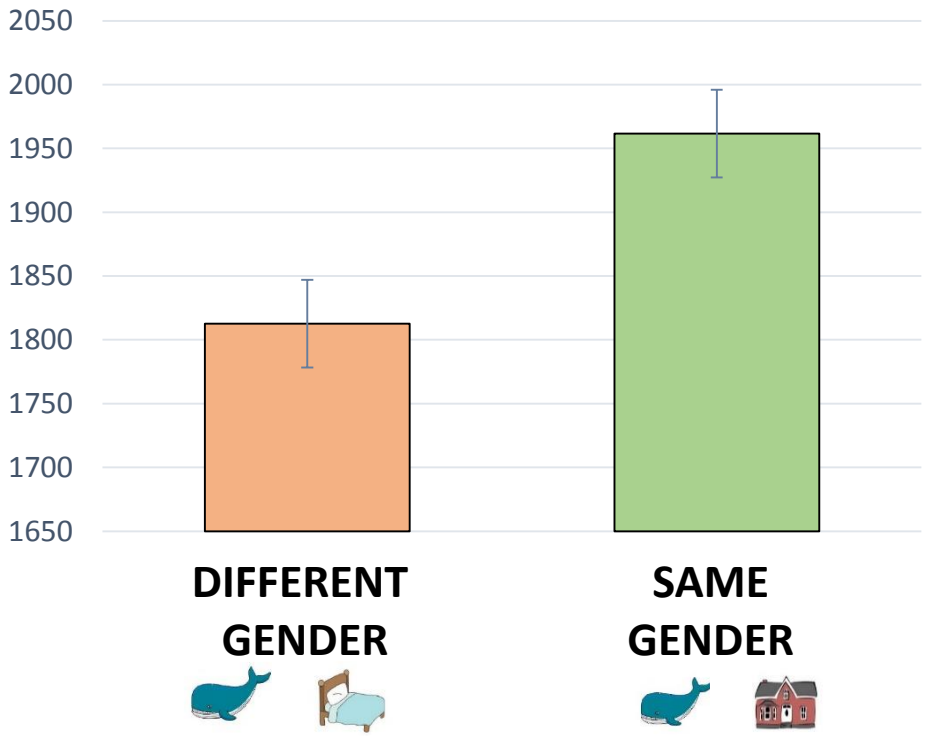
Speeded version of training task

- **QUESTION:** Do they use gender marking in identifying the noun?

PREDICTION:

Should be faster in the different gender trials
since can identify target before onset of noun
(Lew Williams & Fernald, 2007)

Speeded Noun Identification



No effect of variability (collapsed)

TRUE ON DAY 5
BUT NOT DAY 3

- takes more input to see these effects in processing

2AFC Test of Frames

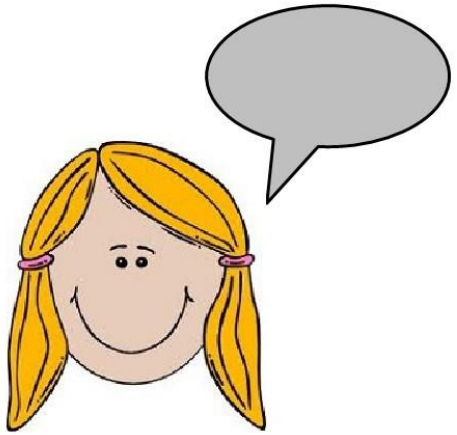
masculine singular:	il	$noun_{[masculine]}$ + o	**il** letto
masculine plural:	i	$noun_{[masculine]}$ +i	**i** letti
feminine singular:	la	$noun_{[feminine]}$ +a	**la** ballena
feminine plural:	le	$noun_{[feminine]}$ +e	**le** ballene

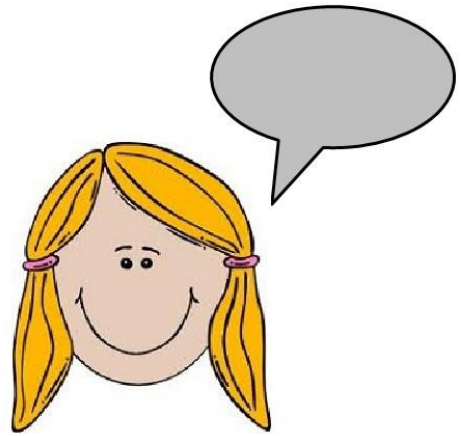
correct noun-phrase

foil

(**determiner** and **vowel** don't match)

Familiar nouns
& unfamiliar nouns





la balena

il balena

2AFC Test of Frames

	FAMILIAR
HIGH VARIABILITY	90%
LOW VARIABILITY	93%

2AFC Test of Frames

	FAMILIAR	UNFAMILIAR
HIGH VARIABILITY	90%	48% *CHANCE*
LOW VARIABILITY	93%	55%

No effect of variability condition

- Strong item learning but no generalization of frames independent of the nouns

Summary

- Strong item level learning
  - High accuracy with familiar items in the 2AFC test
  - Evidence that gender marking is used in identifying the familiar nouns
  - Representations become more robust with increased exposure
- Generalization weak/tentative
  - Unfamiliar nouns test 2AFC at chance
    - ➢ Aren't learning "frames" independent of the nouns
  - Number marking is better learned in the **LOW** variability condition
  - For this game, high variability doesn't lead to generalization (and too much variation may actually hinder learning plurals)

Future Directions

- Age comparisons – compare with 11 year olds.
- How to encourage generalization?
  - Learn singulars first (two training sessions) → move to plurals
  - "Skew" the input – most nouns of one gender, a few "exceptions" the other gender
  - Augment with explicit instruction

Future Directions

- Research using the same “game” exploring aspects of language learning:
  - **Vocabulary learning** – is it better when learning from multiple speakers than one speaker?
  - **Phonetic learning** – do children learn phonetic contrasts better when hear from multiple talkers.
  - **Chinese Tone learning –** can children pick up on lexical use of tone without instruction?

Implications

- Computerized training can provide children with a lot of input – complement classroom teaching
- What type of input is best?

References

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431-436.

Gruter, T., Lew-Williams, C., and Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language research, 28,* 191-215.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science, 18,* 193-198.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30,* 678-686.