Statistical learning over sociolinguistic cues in children and adults

Dr Anna Samara

School of Human Sciences
University of Greenwich

UNIVERSITY of GREENWICH

Overview of my research interests

- 2014. PhD in Psychology, Bangor University (UK)
  shedding light onto the learning processes involved in literacy acquisition
  - *Artificial lexicon experiments* investigating whether statistical learning processes, akin to those implicated in spoken language acquisition (Saffran et al. 1996) contribute to children's ability to pick up on probabilistic spellings (e.g. Samara et al., 2019)
- 2014-2019. Postdoctoral fellow on projects using similar methods with primary school children to investigate…
  - Sociolinguistic learning (Samara, Smith, Brown, & Wonnacott, 2017)
  - Grammar learning (Samara, Wonnacott, & Ambridge, in prep)

UNIVERSITY of GREENWICH

Today's talk

- Four experiments using semi-artificial language (child appropriate) methods to investigate the sociolinguistic development in 5- 6-year-olds (and adults)
  - Experiment 1: deterministic sociolinguistic cues
  - Experiment 2: probabilistic sociolinguistic cues
  - Experiment 3: unreliable sociolinguistic cues
  - Experiment 4: generalizations over sociolinguistic cues

UNIVERSITY of GREENWICH

Learning of variation in language (1)

- Successful language acquisition involves learning that language exhibits variation at all levels – phonological, lexical, morphological, syntactic etc- which is generally constrained
  - e.g. regular past tense realized e.g., [t] (e.g. liked) vs. [d] (loved) vs. [ɪd] (hated) depends on phonological features of final segment of the stem
  - Much interest in children's ability to learn linguistic (deterministic) conditioning of this sort

UNIVERSITY of GREENWICH

Learning of variation in language (2)

- Variant preference depends *probabilistically* on extralinguistic social (e.g. speaker) characteristics:
  - Speaker gender, social group, ethnicity, age etc.
  - Social context (formal vs. informal)
- Variation at all levels, including words (e.g. *dinner* vs. *supper*, *bup* vs. *ban* vs. *roll*), and accents (e.g. pronouncing bath, grass and dance with a short vs. long vowel) etc.
- Increasing awareness that language learning does not take place in a sociolinguistic vacuum!

UNIVERSITY of GREENWICH

Insights from natural languages: adults

- Extensive naturalistic and experimental evidence that adults' usage of linguistic variants conditioned by extra-linguistic characteristics and or social context characteristics (e.g. formal vs. informal environment)
  - Local dialect variants used by adult speakers to index (consciously or unconsciously) their identity in social media (e.g. #VoteAye) (Shomark et al. 2017)
  - Phonological and grammatical variables (e.g. '*she play(s)*'; *walking* /ɪŋ/ vs *walkin'* /ɪn/) conditioned by careful vs. relaxed speech (Trudgill, 1974)

UNIVERSITY of GREENWICH

Insights from natural languages: children

- Research subject to methodological difficulties
  - 'appropriate' time-window? E.g., early enough to examine the earliest development of grammatical forms but late enough for the child to be verbal enough (Roberts, 1994)
- Theoretical argument that child language is monostylistic until age 12 (Labov, 1964)
- Role of input unclear given that child-directed speech tailored to child's social group
  - E.g. children's acquisition of gendered patterns of variation might reflect gendered CDS (Foulkes et al., 2005)

UNIVERSITY of GREENWICH

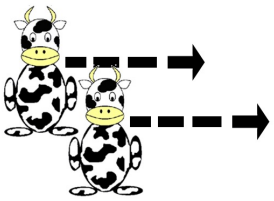
Insights from natural languages: children

Have they *learned* the conditioning, or are they spoken to differently and/or spending more time with their own gender?

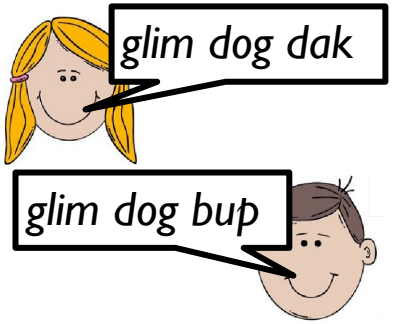
UNIVERSITY of GREENWICH

Artificial language learning

- ✓ Complete control over input to learning
- ✓ Insights over the the learning processes at work
- ✓ Can be adapted for use with children, infants, and adults



glim cow fip



ooshra trunko tid buzzo

beznim

UNIVERSITY of GREENWICH

Insights from artificial languages

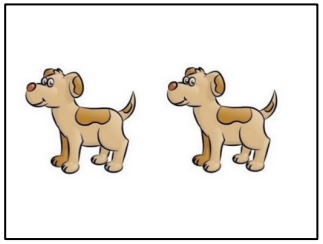
- Emerging evidence that adults pick up on novel socially conditioned variation (Rácz et al., 2017; Sneller & Roberts, 2018) BUT no work with children
- Adults adept at learning deterministic (lexically) conditioned variation (Hudson & Newport, 2009; Smith & Wonnacott, 2010) and 6-year-olds can too (Wonnacott, 2011)
- Adults can acquire probabilistic morphosyntactic variation but children cannot (Hudson Kam, 2015)
  - Not socially conditioned

UNIVERSITY of GREENWICH

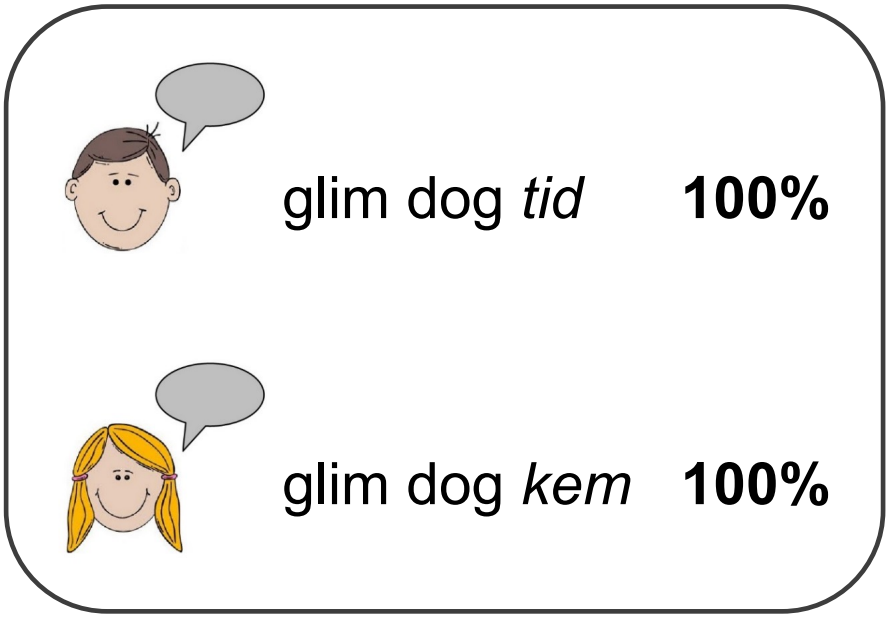
Samara et al. (2017). *Cognitive Psychology*

- Learning sociolinguistic variation in a semi-artificial language (English nouns and novel verbs and function words)
- Variation in usage of particles which follow nouns: two alternating particles “gos”/”kem

glim	*dog*	*gos*
[THERE ARE TWO]	DOG	PARTICLE
glim	*dog*	*kem*



Conditioning is based on speaking identity



Ability to pick up on speaker identity cues fundamental to tracking sociolinguistic variation

Particles occur equally for each of 8 nouns
Both old and new nouns are features in tests

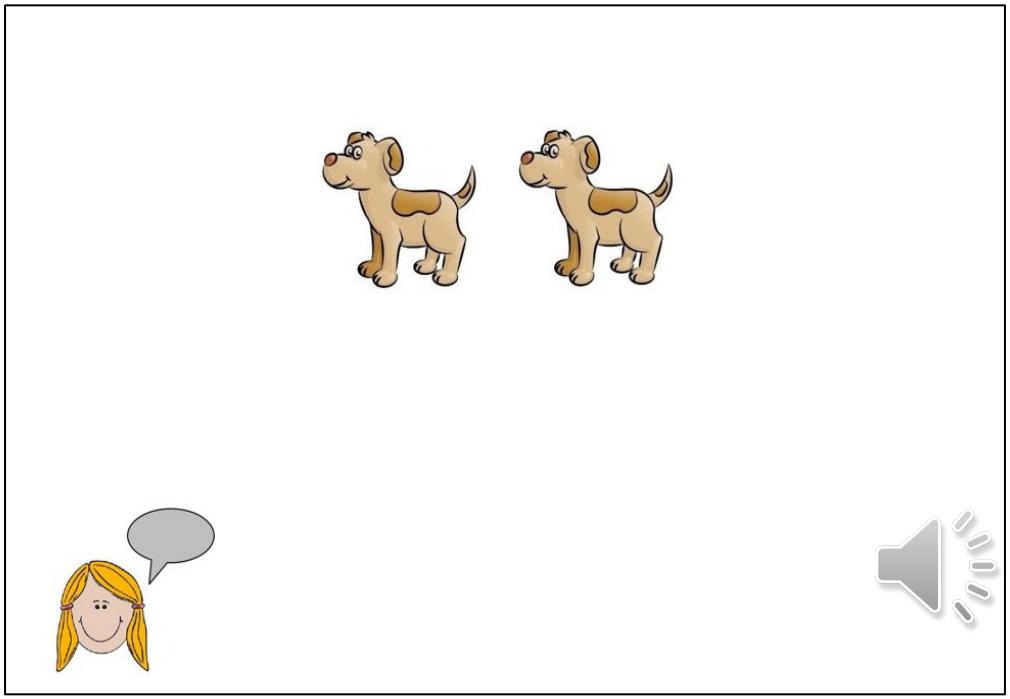
UNIVERSITY of GREENWICH

Methods

Session 1	Session 2	Session 3	Session 4
1. Noun practice	1. Noun practice	1. Noun practice	1. Noun practice
2. Training	2. Training	2. Training	2. Training
3 .Production			3 .Production
			4. 2AFC

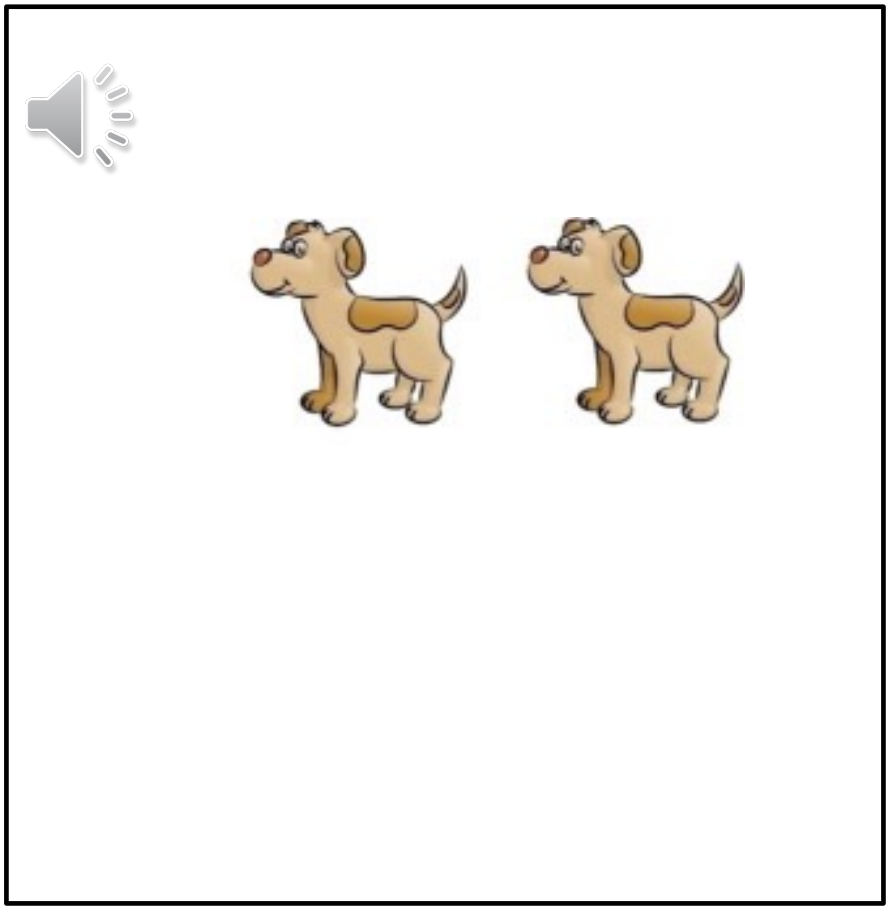
Training methods

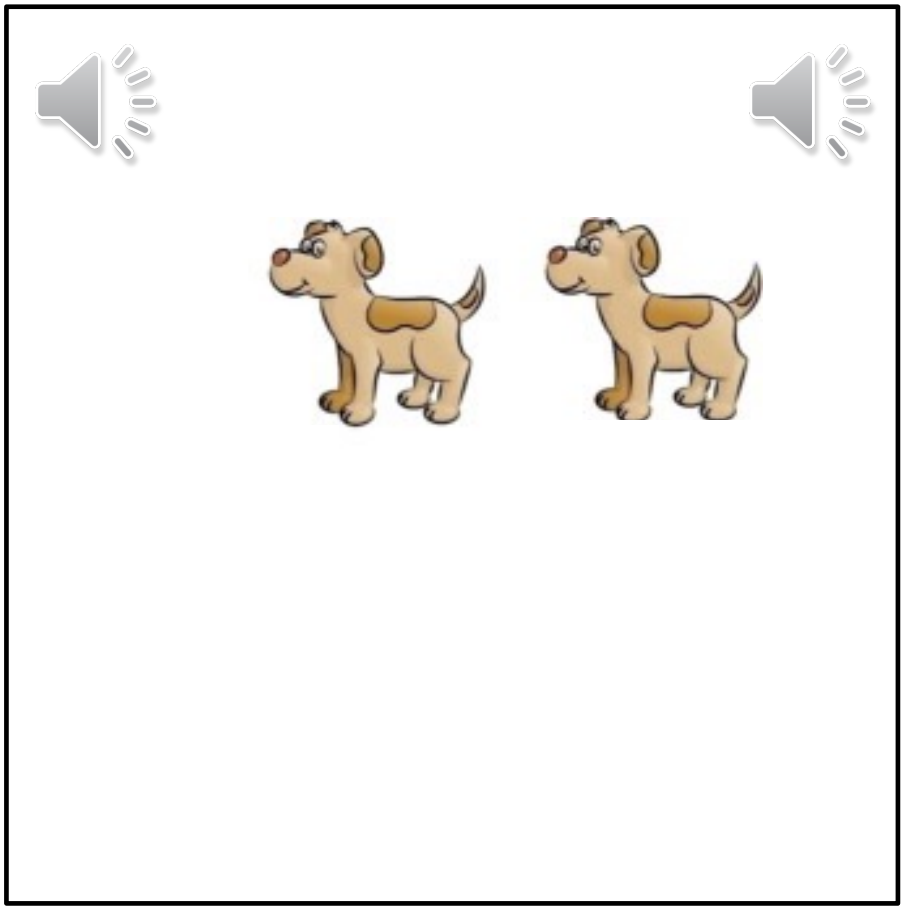
- Over four sessions (days)….



UNIVERSITY of GREENWICH

Tests



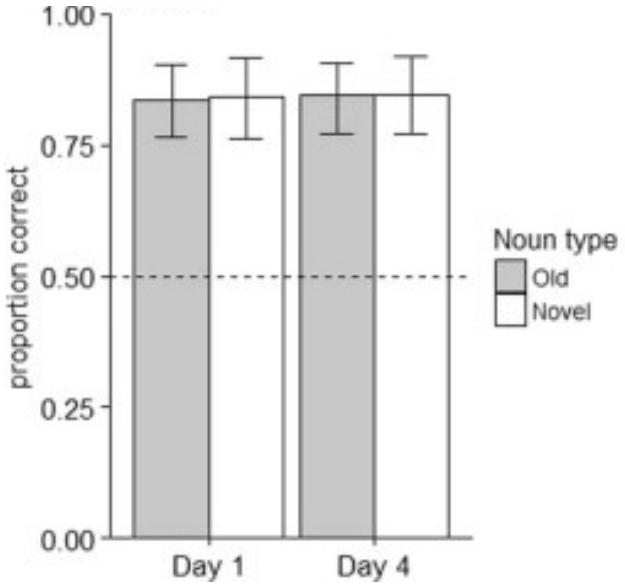


UNIVERSITY of GREENWICH

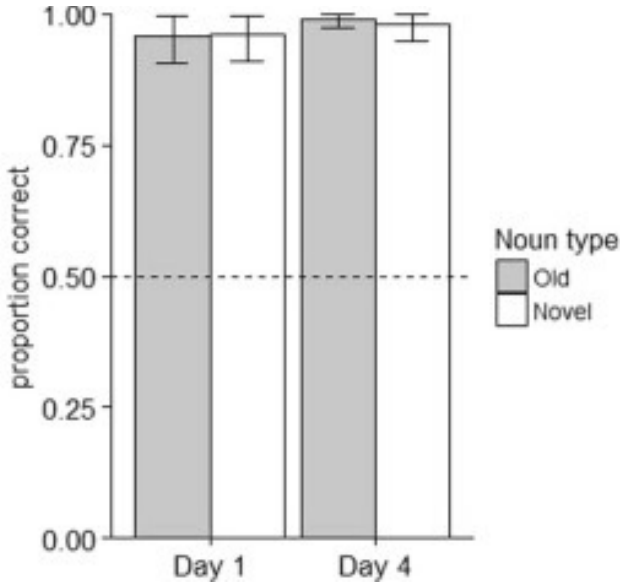
30 5- 6-year-olds, 30 adults

Results: Experiment 1 (production)

children



adults



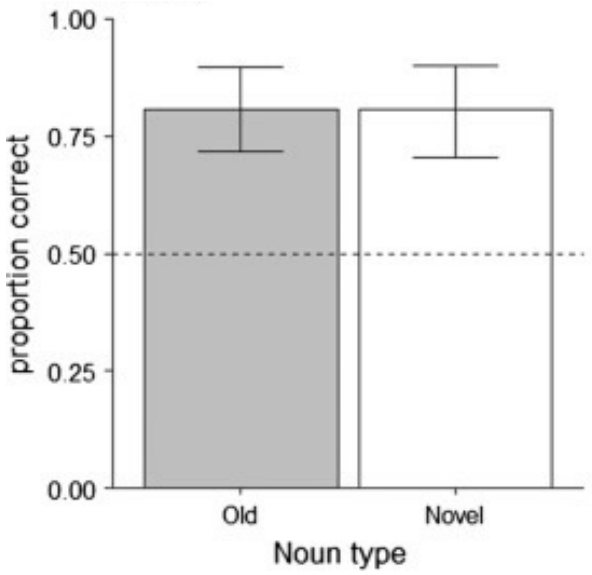
Above chance performance in both age groups

UNIVERSITY of GREENWICH

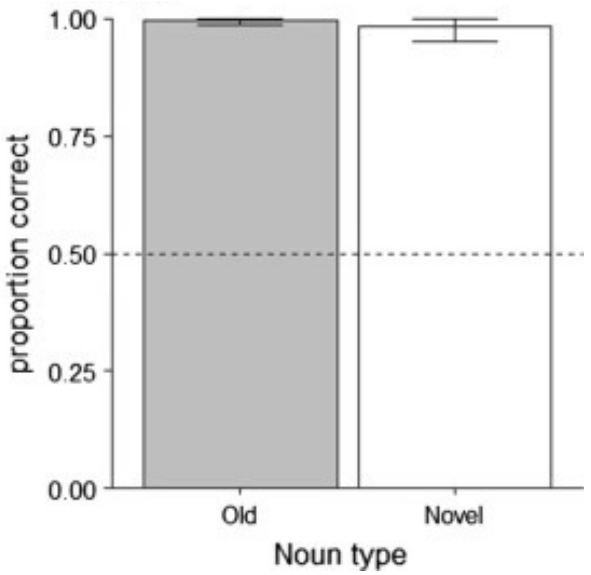
30 5- 6-year-olds, 30 adults

Results: Experiment 1 (2afc)

children



adults



Above chance performance in both age groups

EXPERIMENT 2

- Sociolinguistic variation is NOT categorical
  - e.g. linguistic variation associated with gender involves gender-preferential rather than gender exclusive differences (e.g. Labov, 1966)

glim dog *tid* **75%**

glim dog *kem* **25%**

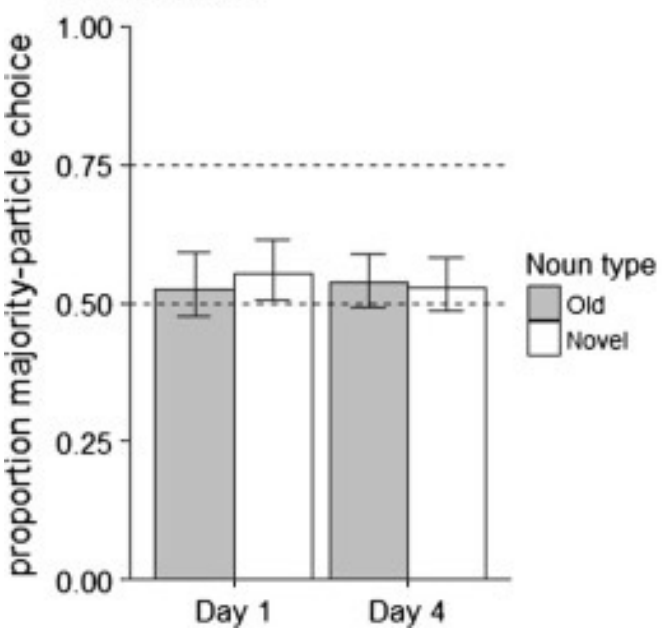
glim dog *kem* **75%**

glim dog *tid* **25%**

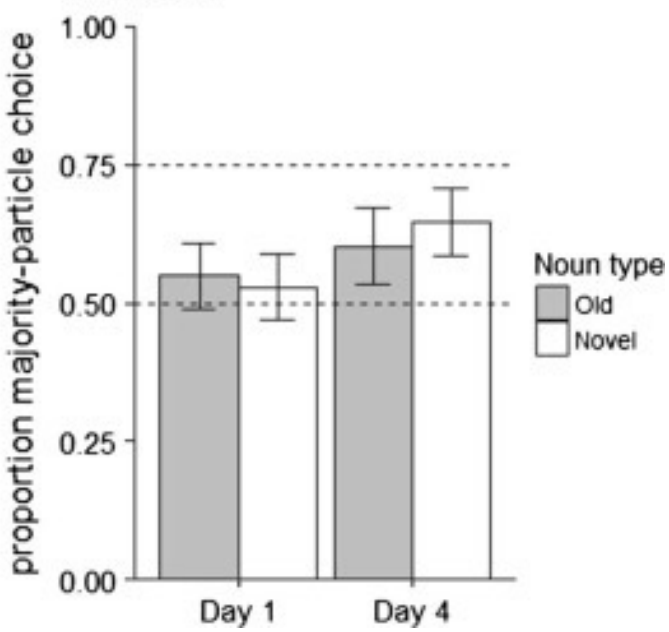
All other aspects identical to Experiment 1

Results: Experiment 2 (production)

children



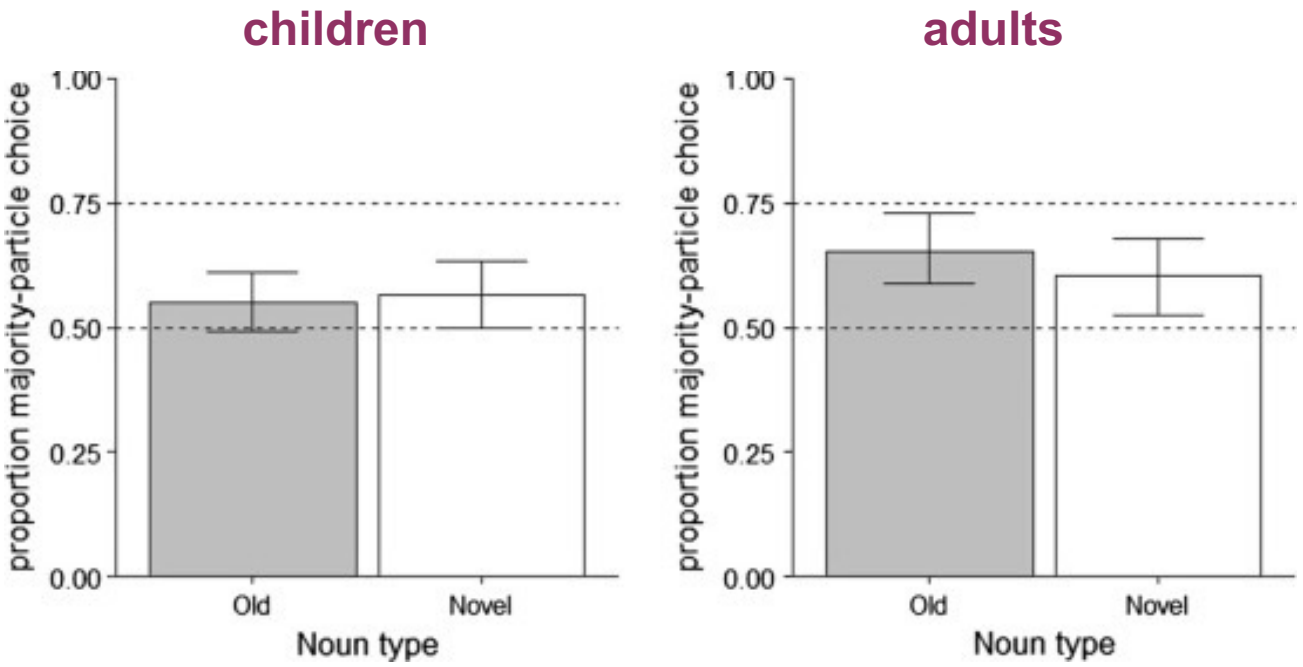
adults



Only adults (day 4) significantly above chance

30 5- 6-year-olds, 30 adults

Results: Experiment 2 (2afc)



Day 4 only: across noun types, both above chance (though it's hard…)

EXPERIMENT 3

- Linguistic variation is hardly ever unpredictable (i.e. unconditioned) in natural languages
  - What do children do to unconditioned variation and how does it compare to what adults do?

glim dog *tid* **50%**

glim dog *kem* **50%**

glim dog *kem* **50%**

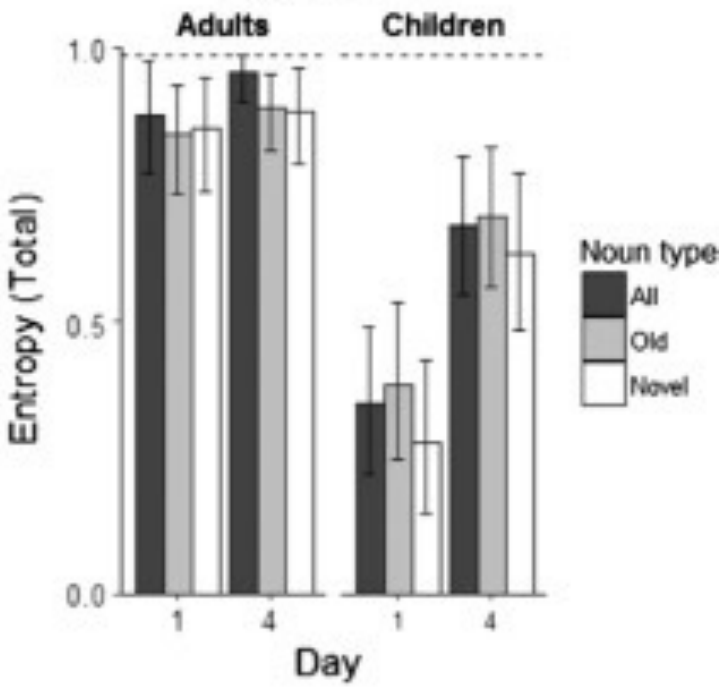
glim dog *tid* **50%**

All other aspects identical to Experiments 1/2

$$Entropy = -\sum P(i)\log_2 P(i)$$

Score of 0 is ‘monostylistic’, 1 is using particle equally probably (as in input)

Experiment 3

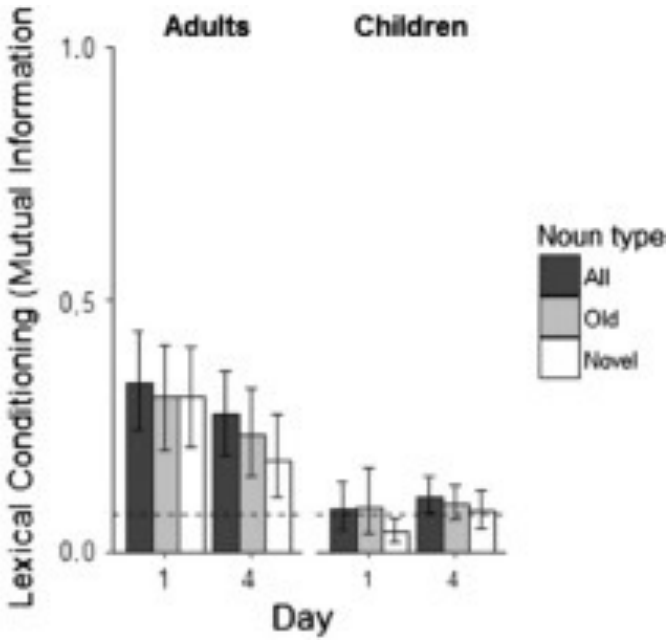


Adults are more variable in productions than children

$$Mutual\ \ Information\ (MI) = total\ \ entropy - conditional\ \ entropy\ \ of\ \ particle\ \ given\ \ noun$$

Higher scores is more conditioning

Experiment 3

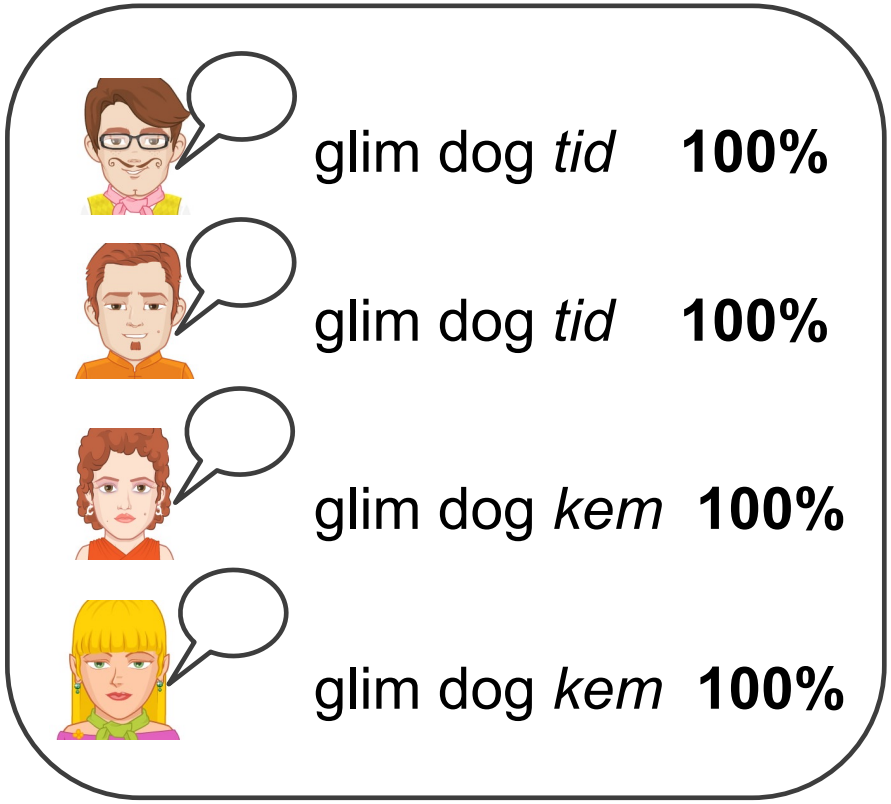


Adults (not children) lexicalize unrestricted variation

EXPERIMENT 4

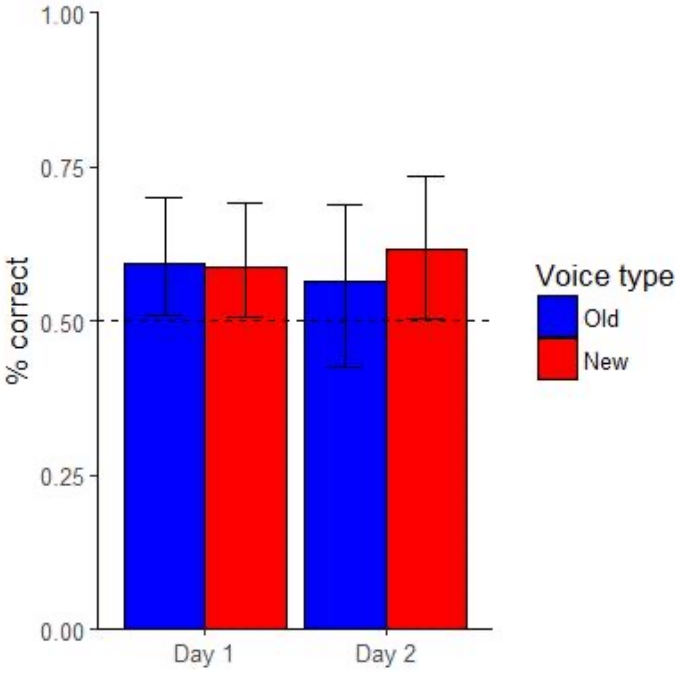
- Sociolinguistic variation is based on broader macro-social categories like gender, which encompasses multiple speakers.
  - Can children reliably acquire variation in linguistic particles based on speaker gender rather than speaker identity?

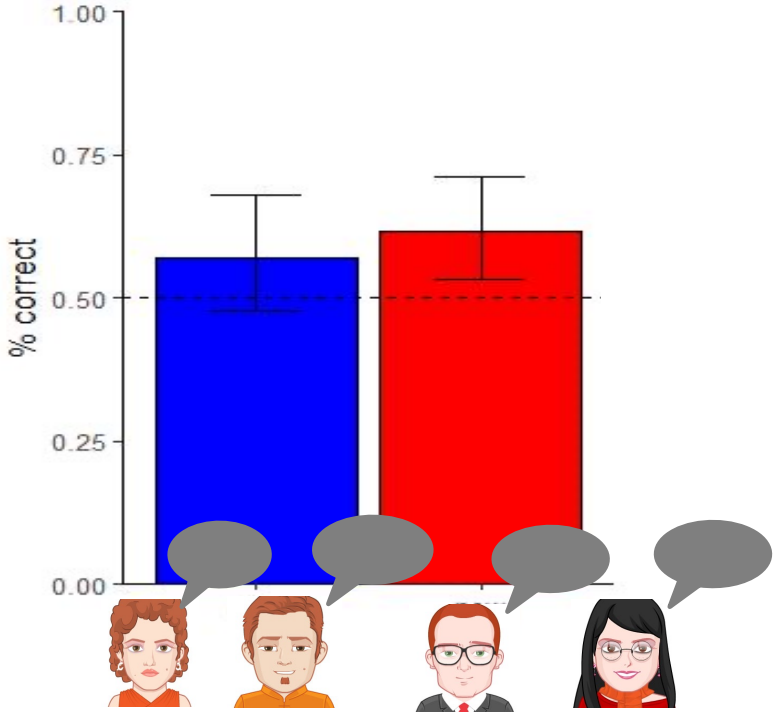
2-day study with six-year-olds (n =19)



UNIVERSITY of GREENWICH

Experiment 4, results





Not evidence of learning in production but above 2afc chance across speakers

BRINGING IT ALL TOGETHER

- 5- 6-year-olds (and adults) can pick up that variant particle usage depends on speaker identity (Exp1) and speaker gender (Exp4)
- Ability emerges through incidental tracking of statistical-based cues in input
- Learning of appropriate conditioning is also seen given probabilistic input (as in natural languages), though, it is harder (Exp2)
- Evidence of regularization (exps 1-3) which involves either boosting the frequency of one particle over the other (children > adults), or conditioning of particle use on lexical items (adults > children)
- Taken together, experiments establish (semi)artificial language learning methods with children to address questions re: sociolinguistic development

UNIVERSITY of GREENWICH

Future directions

- Paradigm well-suited to investigate children's ability to learn different types of conditioning cues as well as different types of variation (e.g. lexical vs. phonological)
  - Learning most likely affected by the complexity and salience of both cues and variations (which can be manipulated in our experiments)
  - In what order are sociolinguistic constraints learnt? Social constraints first (e.g. Labov, 1989) or linguistic constraints required first (e.g. Smith et al., 2013)?
  - Role of context, above and beyond speaker-based conditioning (small BA grant submitted)
- Adapting methods for use with younger children?

Elizabeth Wonnacott
University College London

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Kenny Smith
Edinburgh University

Cheryl Chew

Helen Brown
Nottingham Trent University

Chantal Miller

Thank you for your attention!

a.samara@greenwich.ac.uk