Productively Unifying Exception and Variation

Christine Yue, Charles Yang, and Kathryn Schuler

University of Pennsylvania

NELS Special Session: Idiosyncrasy & Grammar

17 October 2025

Idiosyncrasy and grammar

- How do language learners acquire productive rules?
  - In variable contexts?
  - In the face of exceptions?
  - With incomplete input?

- What is the learning mechanism to distinguish categorical rules with exceptions vs. variable rules?

Today: A step in reconciling exception and variation

- The upshot: a learner's **productive generalizations** inform the learning of categorical and variable rules
  - Demonstrated in the lab setting with an artificial language learning task

- The status of optionality:
  - Conditioning factors affect the probabilistic distributions of variable forms, but don't preclude the grammaticality of alternate forms (e.g. Gries, 2003, Bresnan & Nikitina, 2003)
  - The acquisition trajectory often includes inconsistent generalizations
    - e.g. "Mommy go-ed to the gym." (MacWhinney, 2000; Marcus et al, 1999)
    - e.g. Acquisition of Mexican Spanish variable DOM (Callen & Miller, 2022)

Road map

- **Background**
- Proposal
- Experiments
  - The Hidden Wug Task
  - Experiment 1
  - Experiment 2
- Discussion

Learning exception and variation

Regularity and exceptions: **categorical productive generalizations** (e.g. Berko, 1958; Pinker, 1999; Albright & Hayes, 2003; Yang, 2016)

- *English past tense*

(1) She played in the park.

(2) a. He ate.

b. * He eated.

Learning exception and variation

Variation and optionality: **probabilistic mechanisms** (e.g. Kroch, 1989; Yang, 2002; Boersma, 1997; Hayes & Wilson, 2008; Hayes et al, 2009; Bresnan & Nikitina, 2009)

- *Particle verbs in English*
  (e.g. Wasow 1995; Gries, 2003)

  (3) a. She looked the book **up**
  b. She looked **up** the book

- *(-t, -d) deletion*
  (e.g. Labov 1989; Roberts, 1996; 1997)

  (4) / bɛst fɹɛnd /
  [bɛs fɹɛn]
  best friend

Prior approaches

- **Regularization** as default behavior (for children)
  - Driven by input frequency (e.g. Shin & Miller, 2022)
  - How is variation learned?

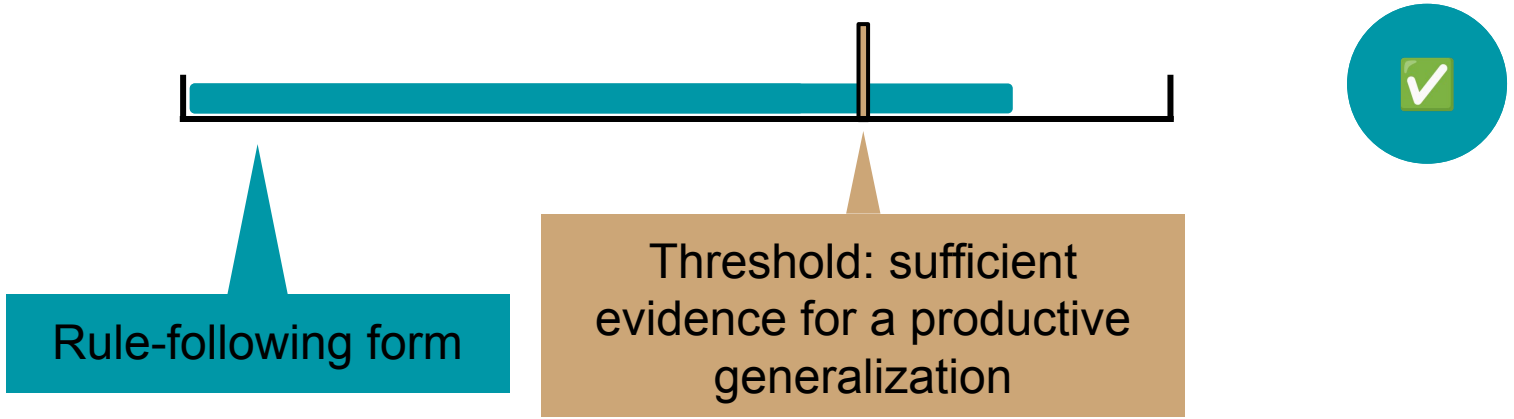
- **Variable patterns** acquired as default
  - Regularization as exceptional behavior caused by learners' limitations (e.g. Hudson Kam & Newport 2005, 2009; Schwab, Lew-Williams, Goldberg, 2018; Austin et al., 2022; Keogh, Kirby, & Culbertson, 2024)
  - Probability-learning mechanisms (e.g. Albright & Hayes, 2003; Ernestus & Baayen, 2003; Hudson Kam & Newport, 2009; Montag, 2021)
  - Why do (the same) learners sometimes regularize their input? (e.g. Schuler, Yang, & Newport 2016)

Road map

- Background
- **Proposal**
- Experiments
  - The Hidden Wug Task
  - Experiment 1
  - Experiment 2
- Discussion

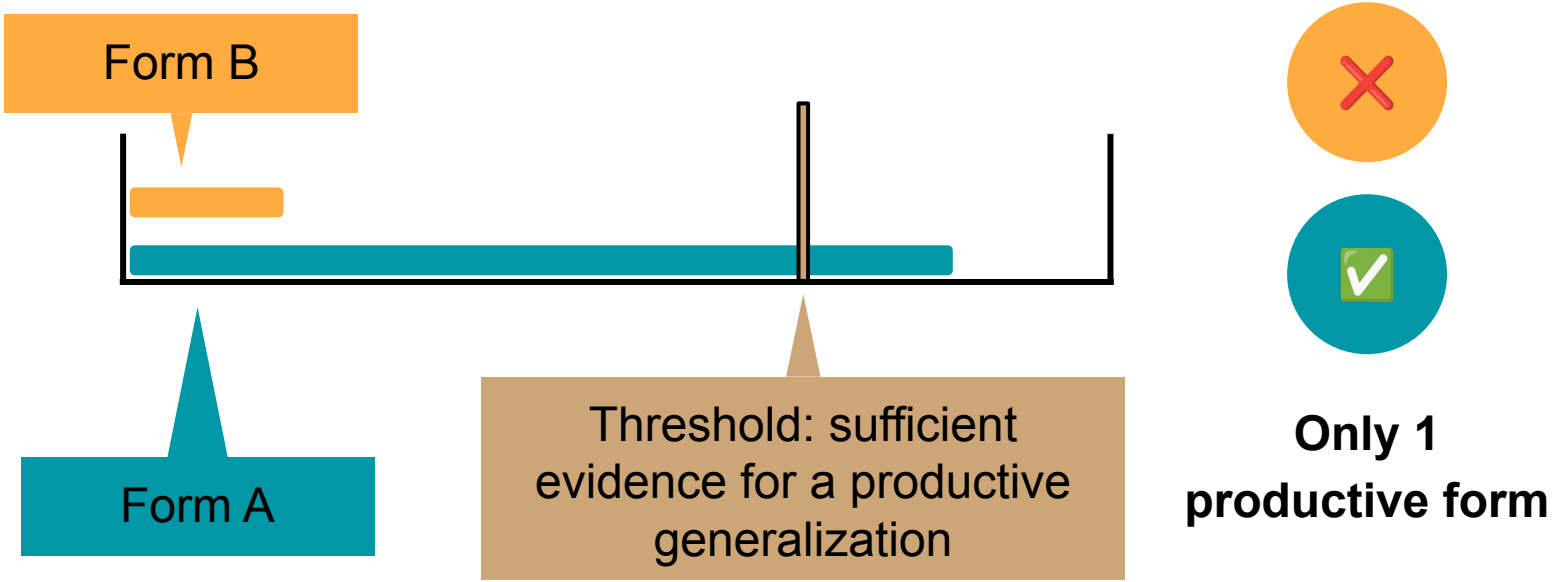
The proposal

Productive generalization underlies both regularization and systematic variation



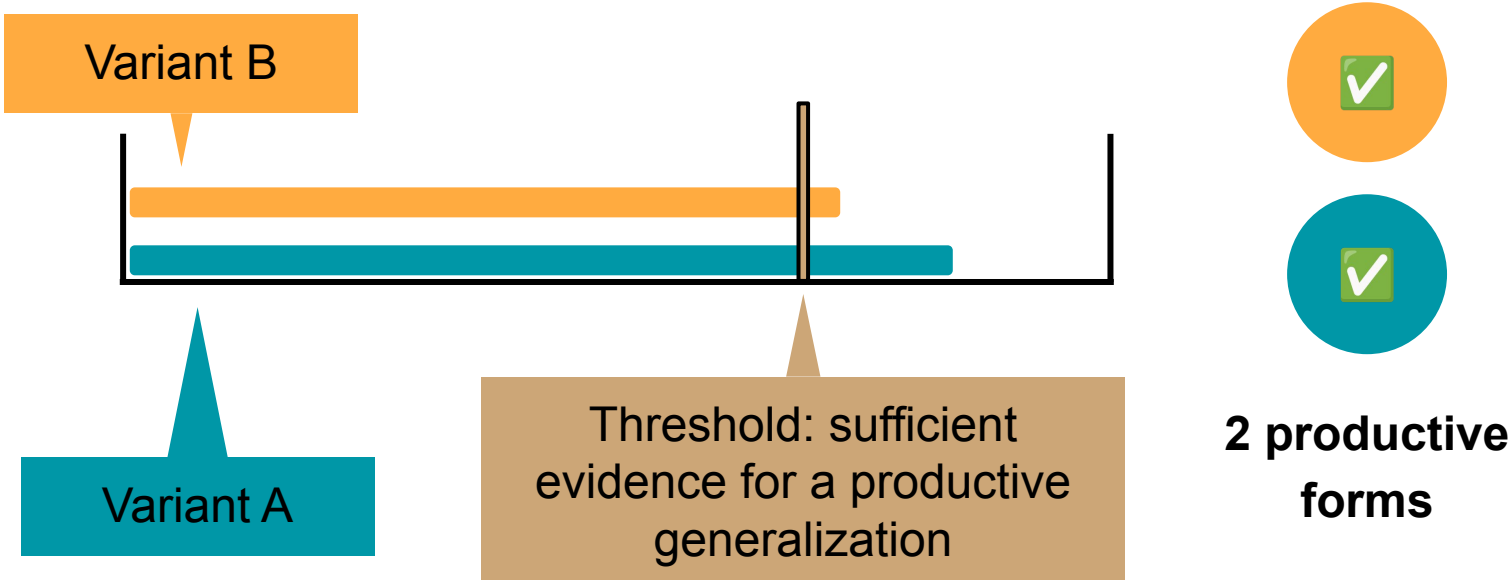
The proposal

Productive generalization underlies both **regularization** and systematic variation



The proposal

Productive generalization underlies both regularization and **systematic variation**



The proposal: A concrete, testable mechanism

- Concretizes the mechanism via which learners acquire rules – categorical or variable
  - Observed in the literature that though children often start with a single variant everywhere, they sometimes learn the variants simultaneously (e.g. Shin & Miller, 2022)
  - Makes predictions that can be tested experimentally (a step in this direction today)

The proposal: Types vs. tokens

- Type distribution
  - Determines the productive generalization of a rule over forms (e.g. Aranoff, 1976; Kiparsky, 1973; Rumelhart & McClelland, 1987; Marchman & Plunkett, 1993; Bybee, 1995; Yang, 2016)
- Token distribution
  - Affects learning of types (more frequent forms learned first) (Pierrehumbert, 2001; Singleton & Newport, 2004)
  - Informs learning of the distribution of the variants (probability learning as a general mechanism), but crucially after the variants are established via generalization

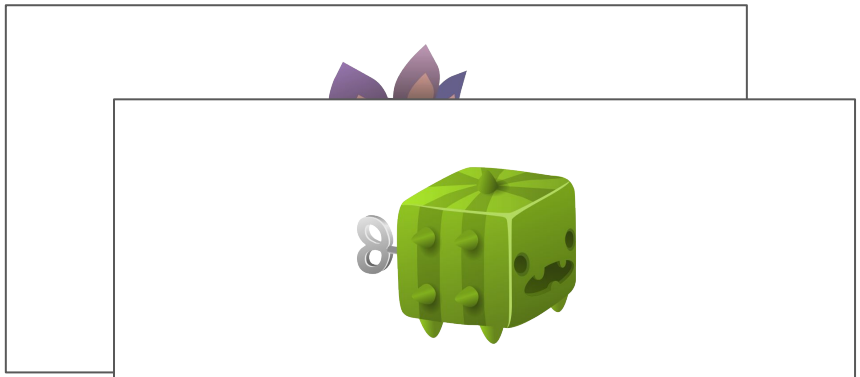
Road map

- Background
- Proposal
- **Experiments**
  - The Hidden Wug Task
  - Experiment 1
  - Experiment 2
- Discussion

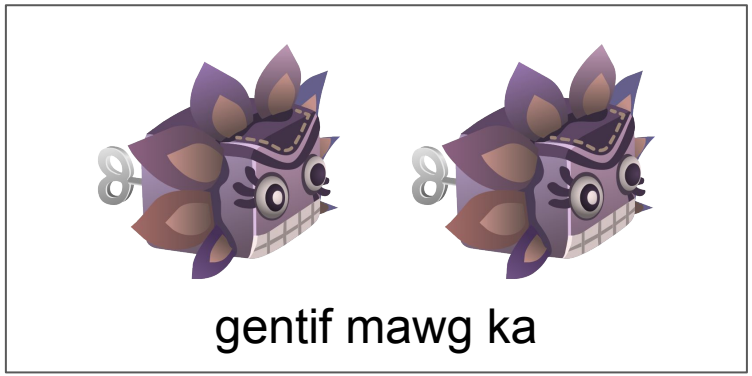
Artificial language learning: Hidden Wug task

- Precisely control the evidence in the learners' input to support productive processes while capturing features of natural language
- Plural marker task (Schuler, Yang, & Newport, 2016; Schumacher & Pierrehumbert, 2021)
- Assess underlying productive generalizations with “hidden” items, which learners have not seen the plural form in exposure, using a wug test (Berko, 1958)

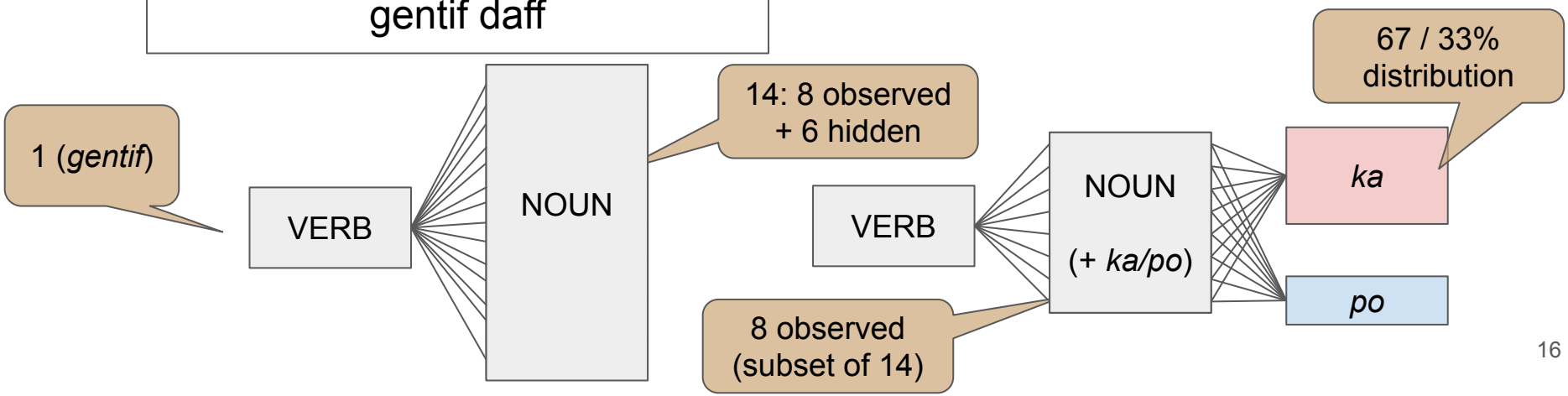
Artificial grammar



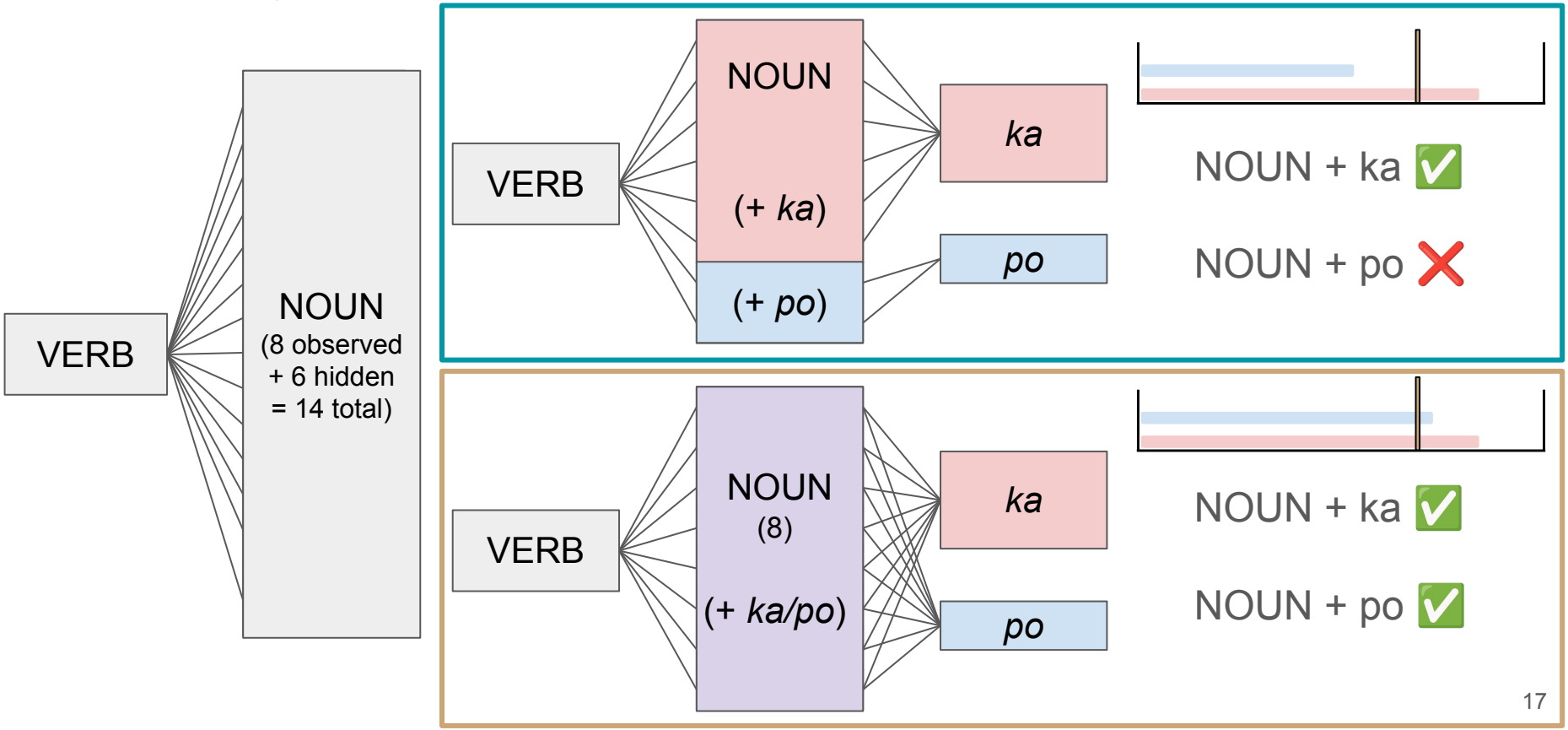
gentif daff



gentif mawg ka



Artificial grammar manipulation



Procedure

Exposure: 228 learning trials

Test: 54 wug test trials

gentif zup ka

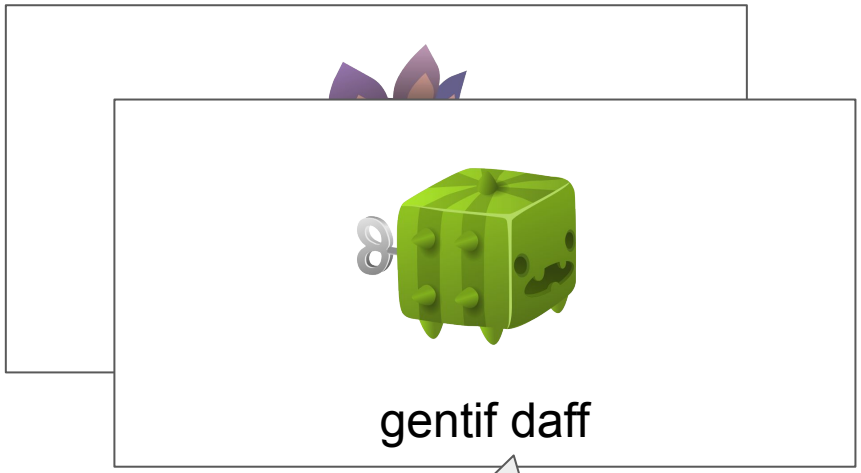
If the sentence for this image is:

gentif zup

How do you say this?

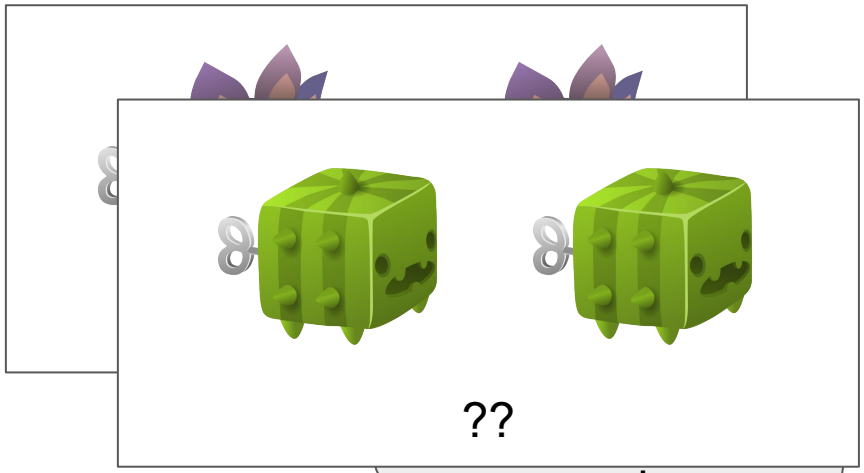
_____

Artificial grammar:



gentif daff

Familiar from encountering in singular form in exposure



??

mawq ka

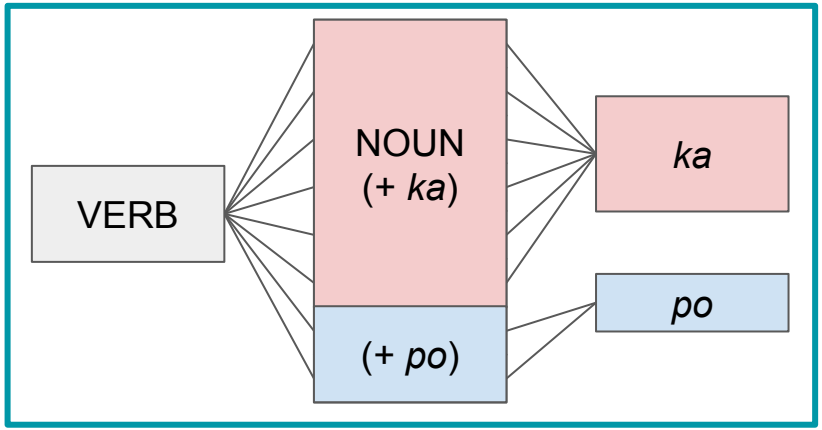
🤔: Add “ka” to make plural?

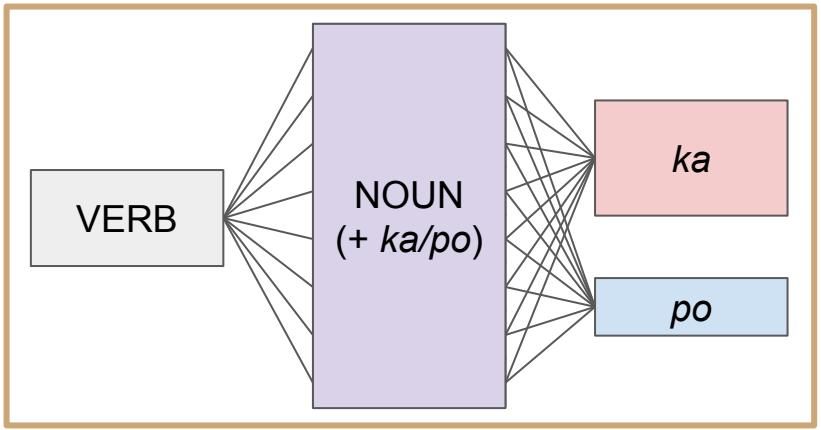
Expt 1: Perfectly consistent

- Consistent language
  - Simplified miniature language resembling *English past tense*
  - One generalization supported

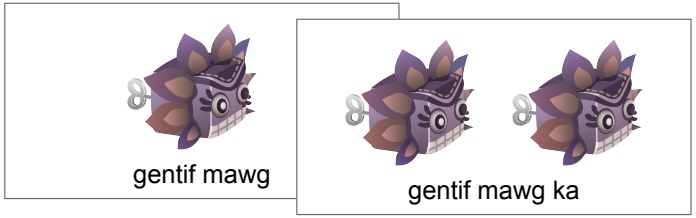
- Inconsistent language
  - Resembles *English particle verbs*
  - Two generalizations supported

- Run on Prolific: n = 60 (30 adults in each language condition) after exclusions

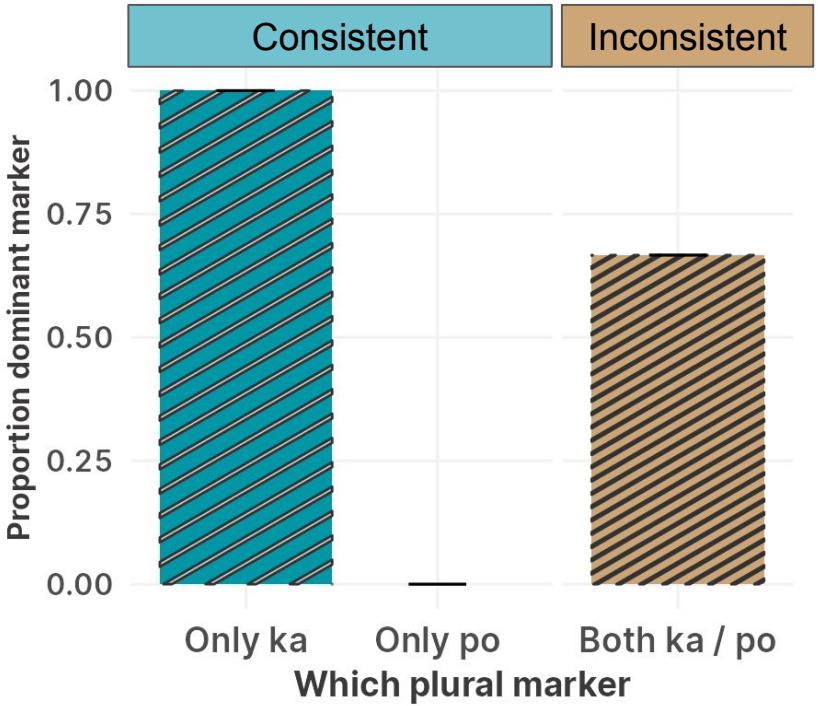




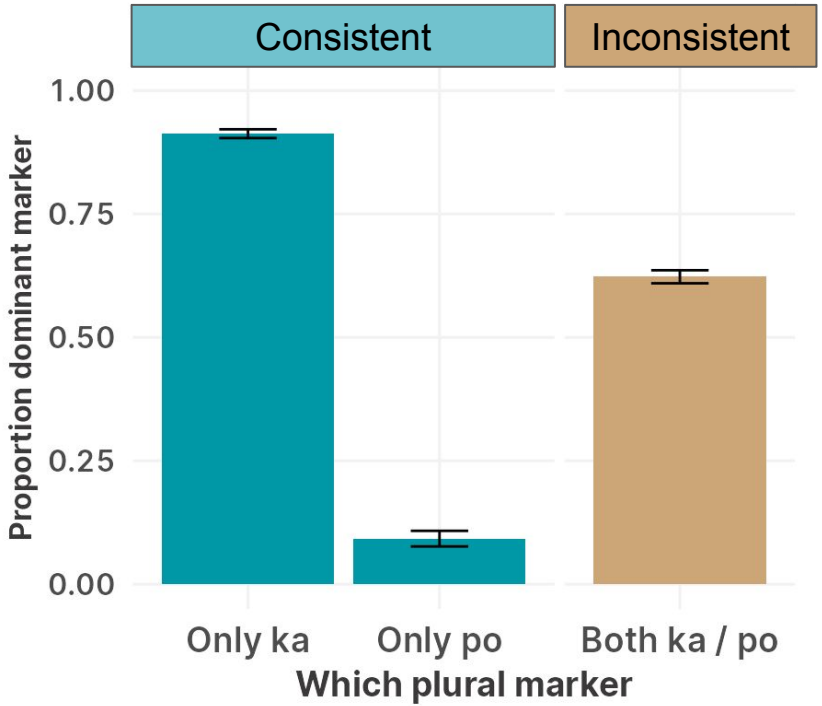
Did participants learn the input?



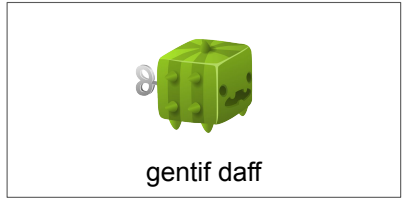
Prediction: Observed distribution



Human Performance

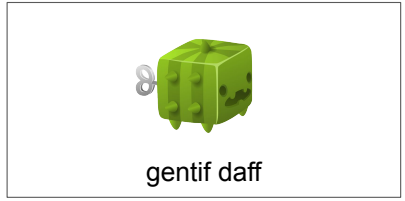


Do participants' generalizations differ?



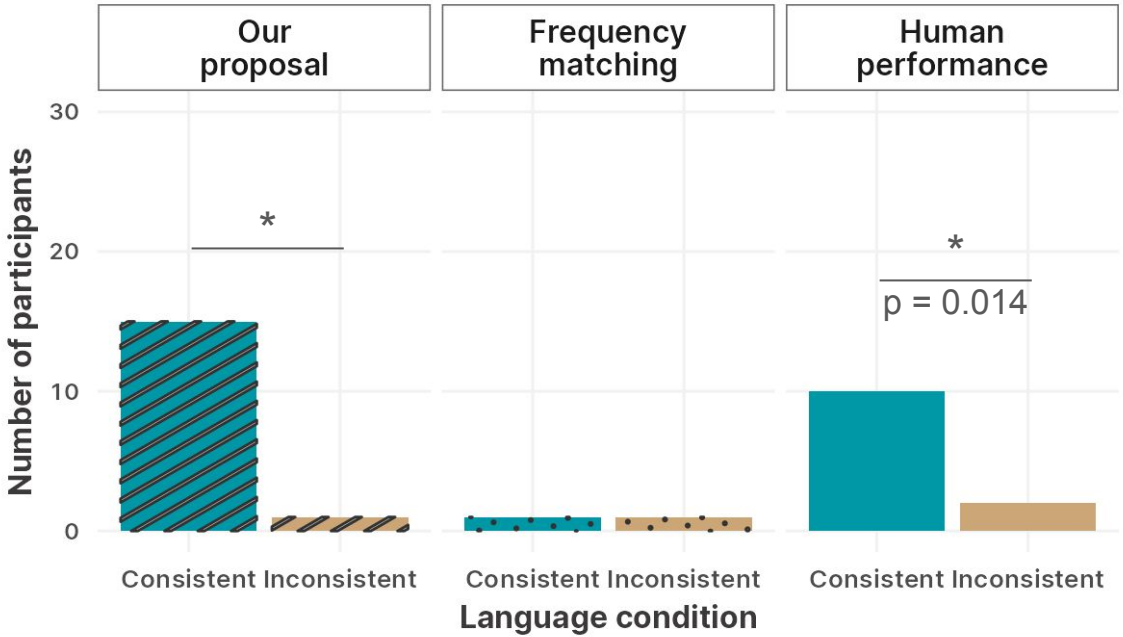
- **Hidden items**: Participants saw only the singular form in exposure
- 6 items x 1 opportunity at test
- Measure: exclusive usage of the dominant marker for all hidden items
  - Strict measurement (for adult participants) due to task effects

Yes, participants' generalizations differ!



Competing predictions between our proposal and frequency matching

- One generalization vs. two generalizations

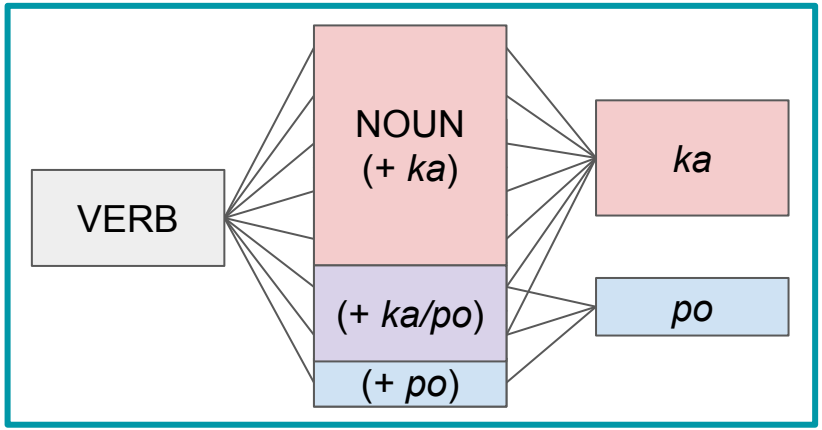


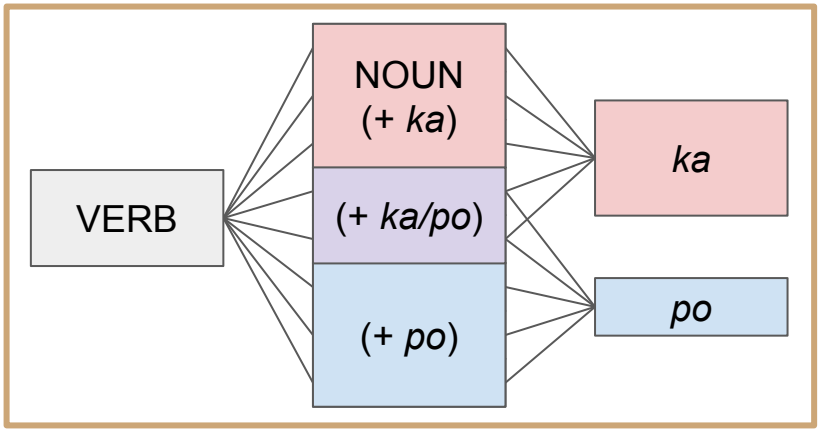
Expt 2: Partially consistent

- One-generalized language
  - Simplified miniature language
    English past tense + doublets

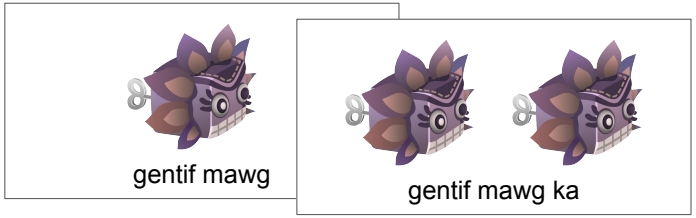
(5) a. She dreamt
    b. She dreamed

- Both-generalized language
  - Resembles *(-t, -d) deletion*
- Not all items eligible for a rule are attested in the input (Yang 2013)
- Run on Prolific: n = 60

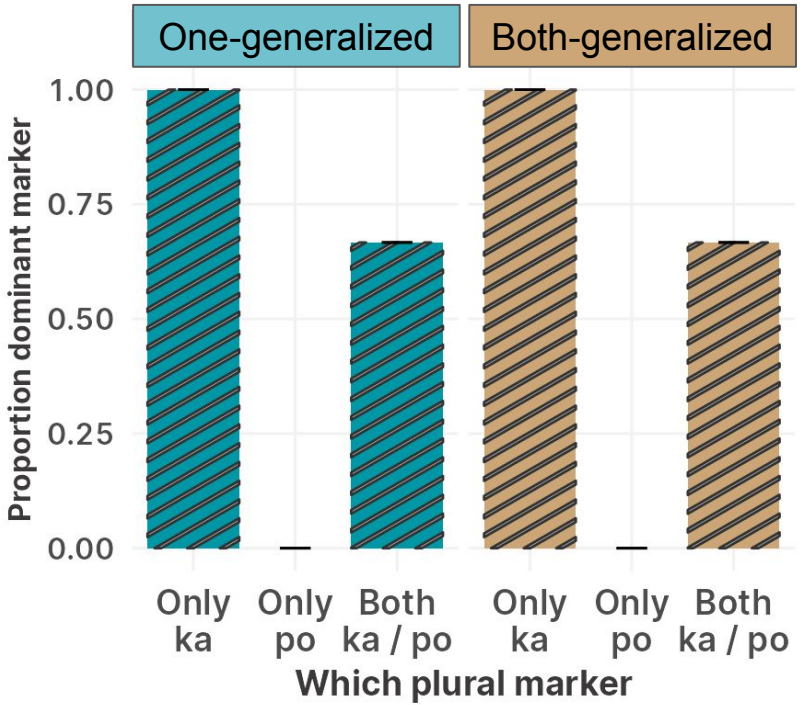




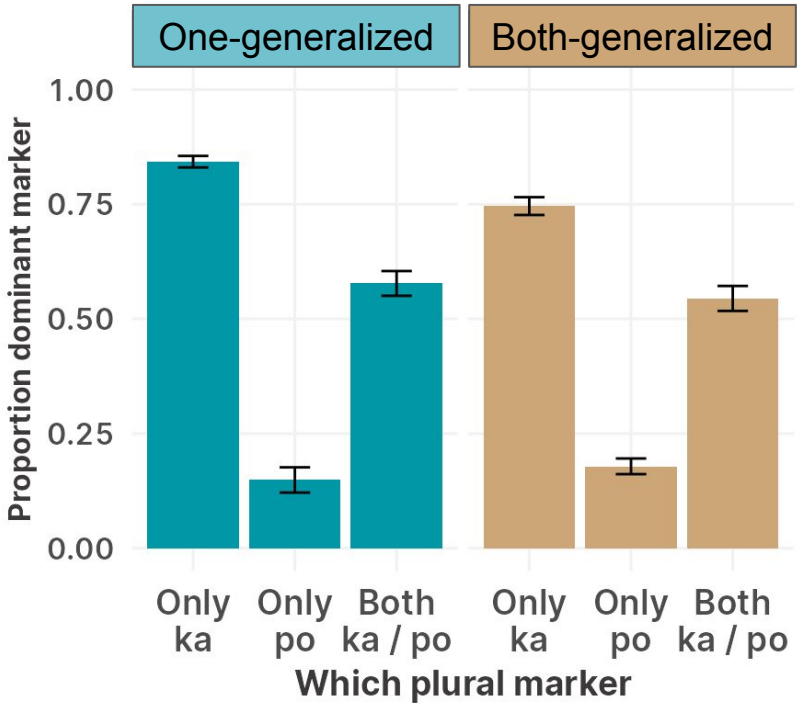
Did participants learn the input?



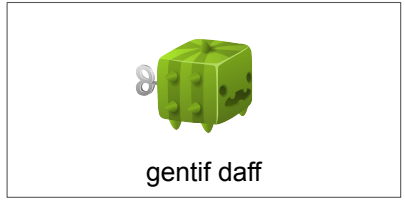
Prediction: Observed distribution



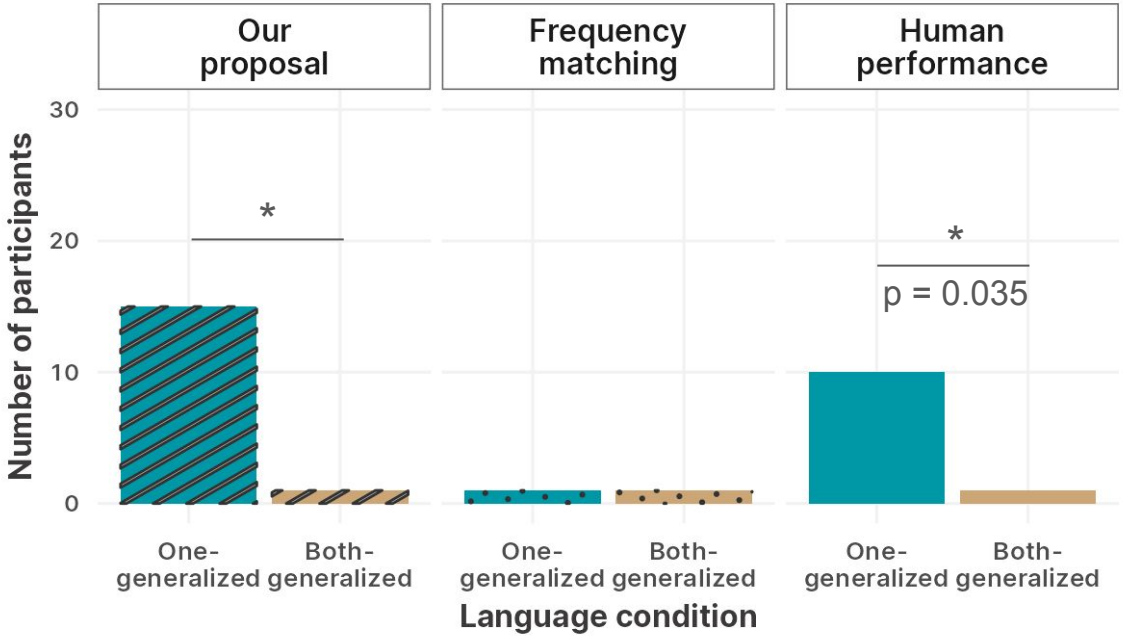
Human Performance



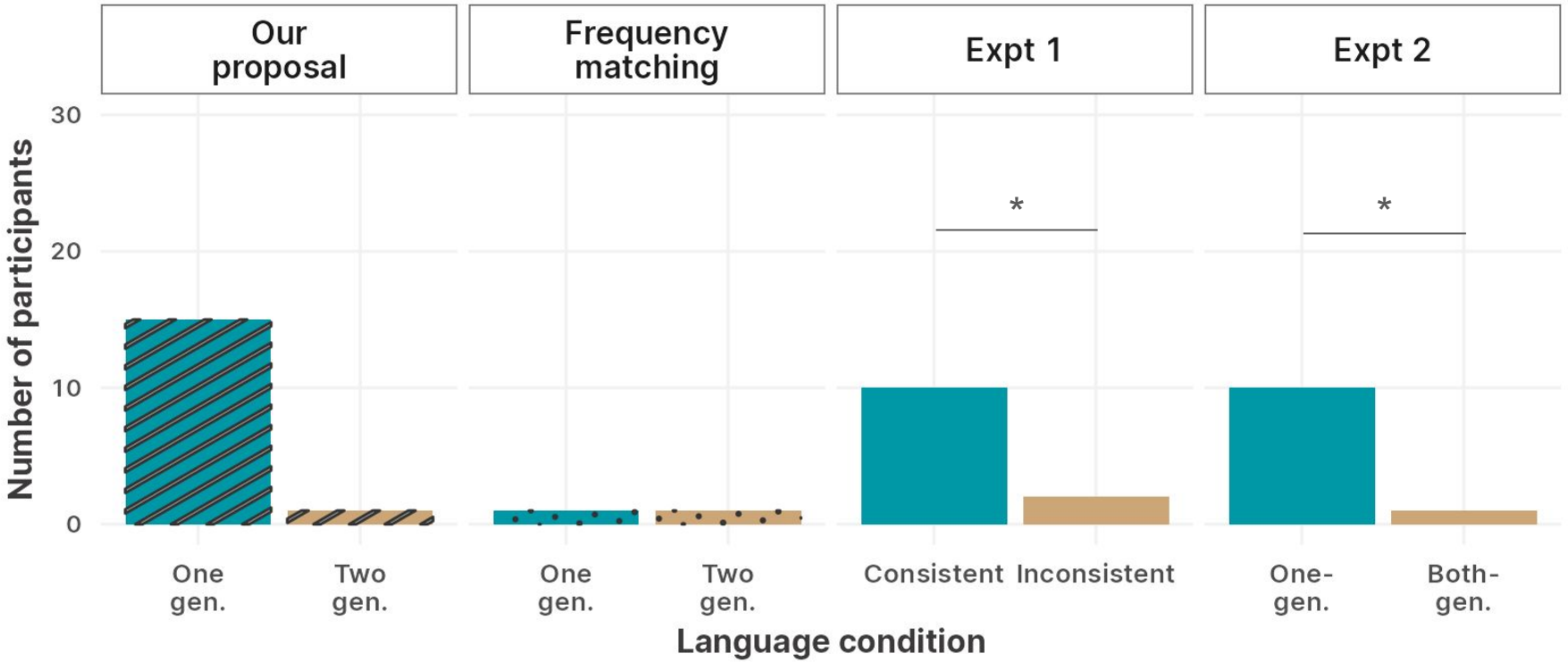
Participants' generalizations differ!



- Hidden items
- Examining exclusive usage of the dominant marker
- Same competing predictions as in Expt 1
- One generalization vs. two generalizations



Combined results: Hidden items



Road map

- Background
- Proposal
- Experiments
  - The Hidden Wug Task
  - Experiment 1
  - Experiment 2
- **Discussion**

Discussion

- Adult learners learn the input veridically across various inputs and can generalize across both consistent and inconsistent patterns in the input
- For productive generalizations, type distributions matter
  - Token distributions held constant across conditions
- Importance of testing generalizations with “hidden” items
  - Methodological innovation to ameliorate some task effects for adults

Future work

- Role of token frequency?
  - How does it impact the learning of forms?
  - How is it incorporated by probabilistic learning mechanisms?
- Developmental differences?
  - Typically, children are said to regularize while adults probability match, but on inconsistent variation (e.g. Hudson Kam & Newport 2005; Austin et al 2022)
  - We suspect that token frequency may play a role in explaining this difference

Idiosyncrasy and grammar

- Theoretical implications
  - A key step in understanding idiosyncrasy, grammar, and variation is to uncover the generalizations
  - Implementations should consider this as a first step
- Experiments reveal (and learning models should capture)
  - A striking difference in qualitative behavior (grammars) based on quantitative manipulation of distributions (input)
  - Distinction of exception and variant as a result of a single distributional learning mechanism

Thank you!

Thanks to the Child Language Lab and the Penn Syntax Lab for discussion

Thanks also for our funding sources: National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1845298, University of Pennsylvania

Appendix

Status of optionality: Particle verbs

Corpus work from Gries, 2003 indicates that constraints may affect the probabilistic distributions of various constructions, but there is no deterministic set of constraints

Table 6:1 Observed distribution of constructions relative to COMPLEX

	Simple ← *Bare NPs (0)*	*Intermediate NP (1)*	*Complex* → *Complex NP (2)*	*Row totals*
Construction$_0$	76	102	16	194
Construction$_1$	186	22	1	209
Column totals	262	124	17	403

Table 6:4 Distribution of constructions relative to LENGTHW^0

	Short ← *1*	*2*	*3*	*4*	*5*	*6*	*7*	*Long* → *8+*	*Row totals*
Construction$_0$	26	51	35	25	15	10	8	24	194
Construction$_1$	103	81	15	5	2	2	1	-	209
Column totals	129	132	50	30	17	12	9	24	403

Table 6:6 Distribution of constructions relative to TYPE

	Pronominal	*Semi-pronominal*	*Lexical*	*Proper name*	*Row totals*
Construction$_0$	–	3	186	5	194
Construction$_1$	77	10	115	7	209
Column totals	77	13	301	12	403

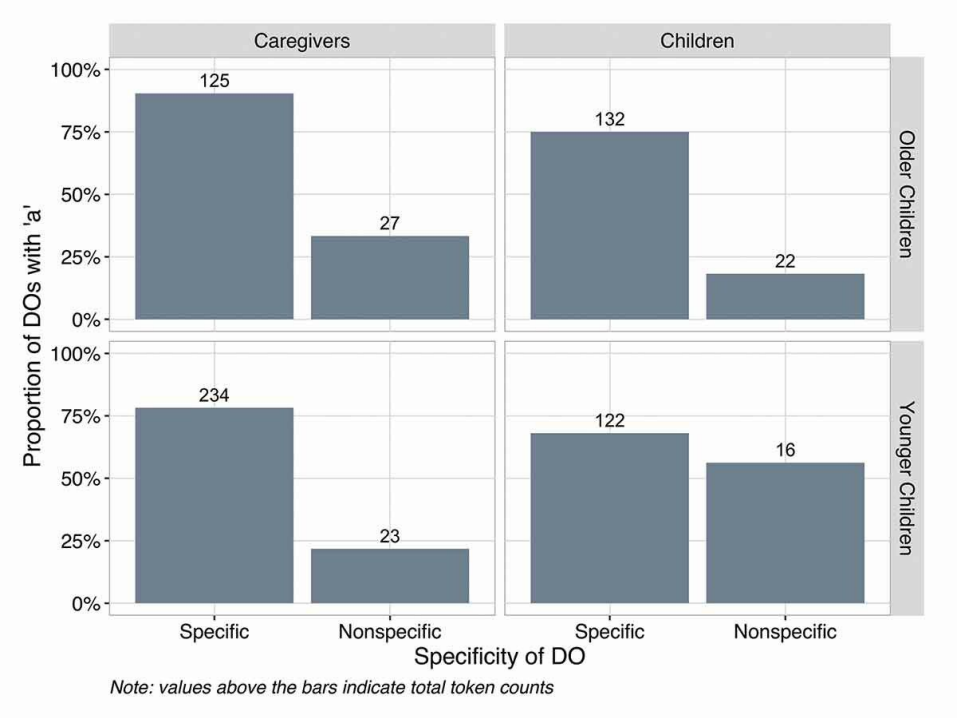
Table 6:5 Distribution of constructions relative to LENGTHS

	Short ← *1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*Long* → *10+*	*Row totals*
Construction$_0$	5	26	26	21	19	14	18	7	7	51	194
Construction$_1$	82	51	35	18	8	8	1	2	4	–	209
Column totals	87	77	61	39	27	22	19	9	11	51	403

Status of optionality: Children's acquisition

Callen & Miller, 2022 find that while older children (>3;0) demonstrated adult-like knowledge of both animacy and specificity constraints on variable DOM, younger children (≤3;0) only demonstrated adult-like knowledge of animacy constraints

(Figure 5)



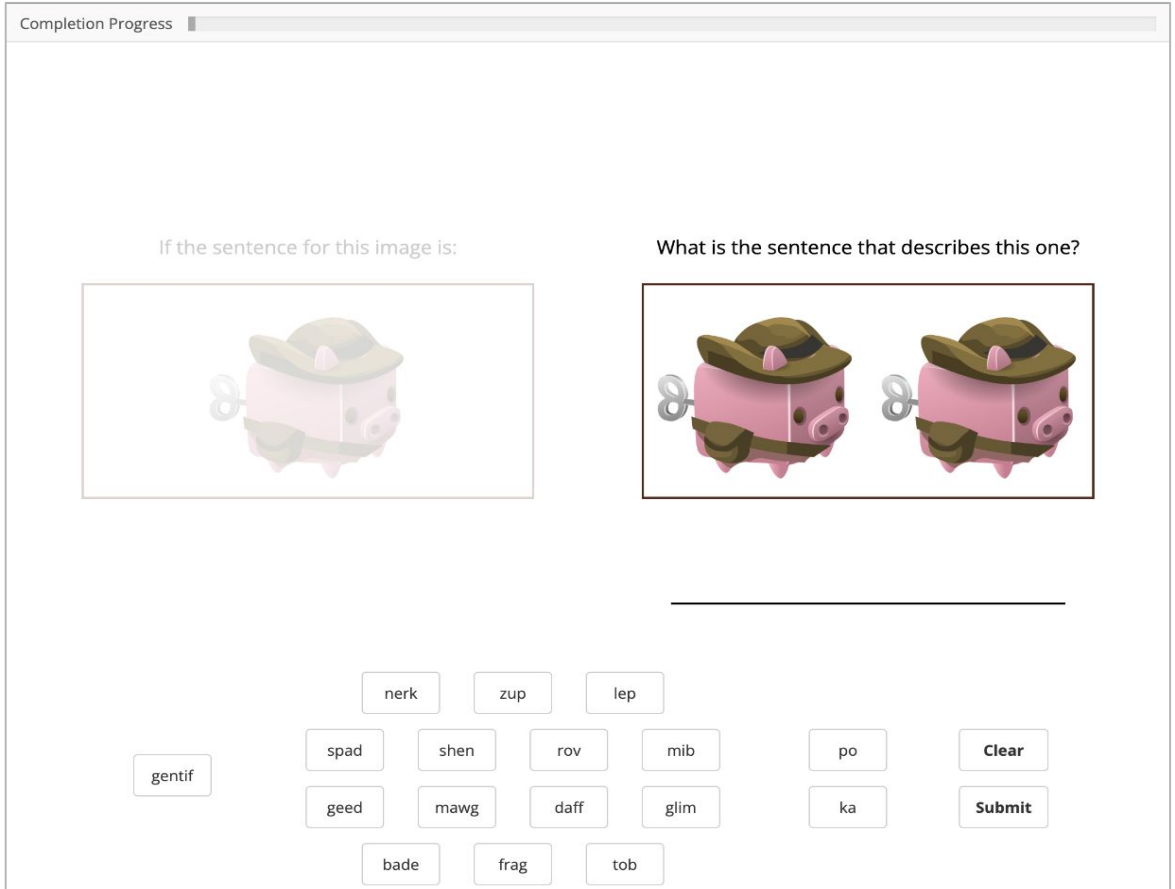
Participants and exclusions

- **Participants**
  - Self-reported monolingual native English speaking adults with normal vision and hearing, with no known language delays, and a high approval rate (>85%) on Prolific
- **Exclusions**
  1. Self-reported attention below 75%
  2. Poor noun learning (< 75% accuracy in the last ⅓ of noun learning task)
  3. Not enough data contributed
     - < 90% productions in test of the form VERB NOUN (PLURAL)

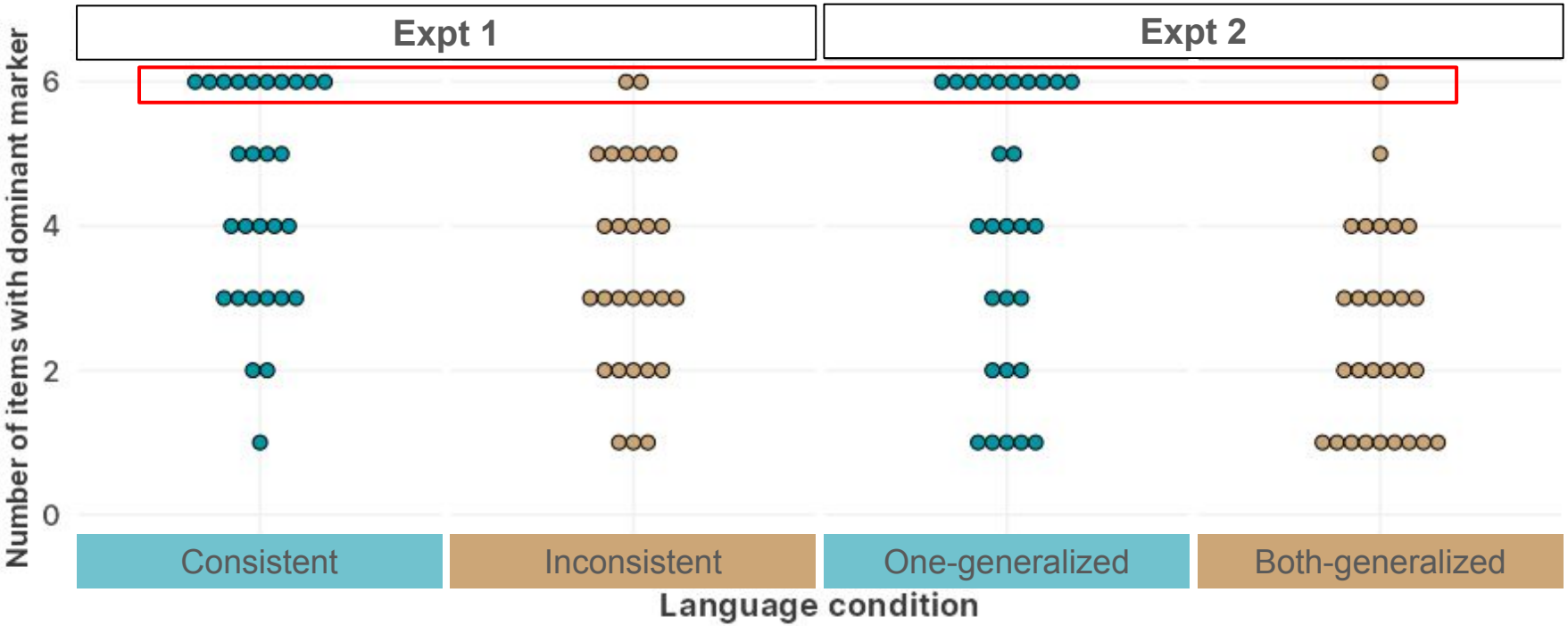
Test: Array

Capturing the stable lexicon

But may contribute more task effects due to having the plural marker options explicitly available



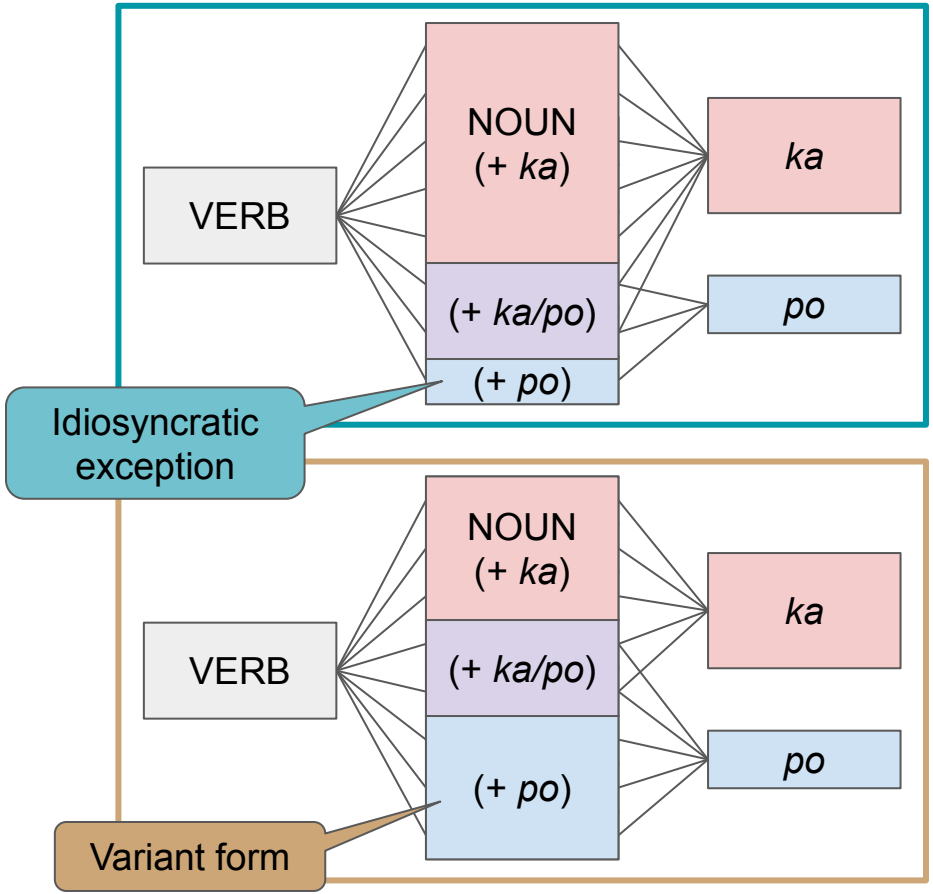
Hidden items



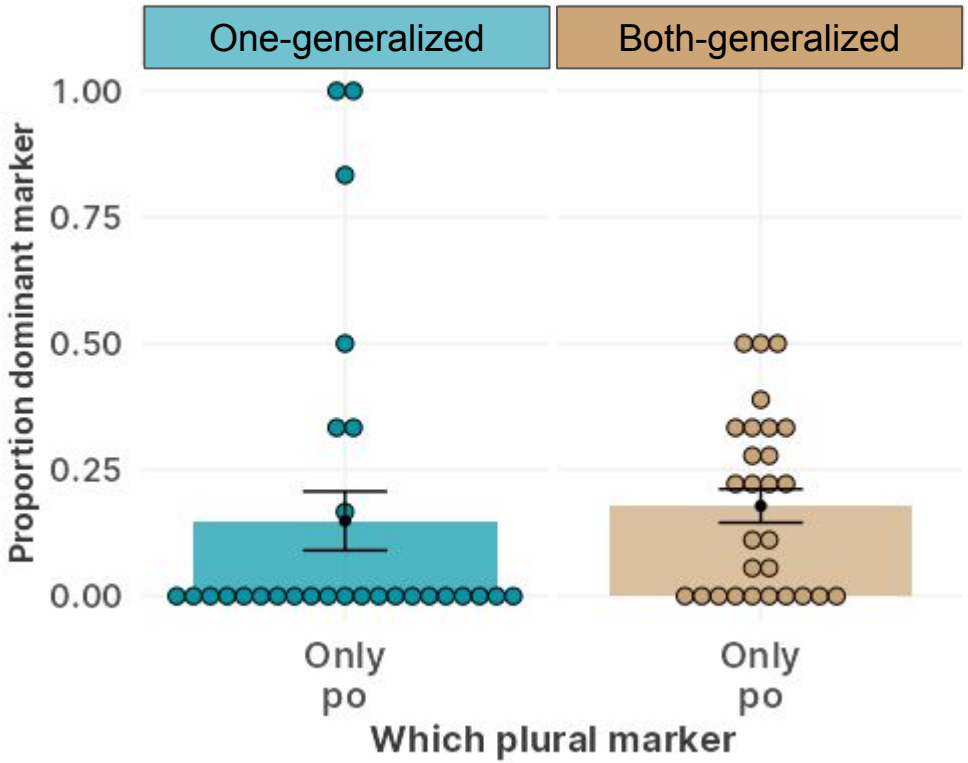
Learning exceptions

Minority-only item(s) are:

- Idiosyncratic exception in the One-generalized condition
- Variant form in the Both-generalized condition



Learning exceptions



- Dots represent individual participants
- Participants treat the minority-only words as po-only words when exceptions than variants (β=1.707, SE = 0.740, p = 0.02)