Introduction to artificial language learning experiments

Annie Holtz, Marc Meisezahl & Shira Tal

Centre for Language Evolution
University of Edinburgh



@anniesotropic

@mmeisezahl

@tal_shira

Workshop Outline

Morning Session (10:00-13:00): Theoretical introduction

- (Very) brief introduction into the fundamentals of experimental design
- Artificial language learning experiments: Paradigms
- On Schleyer's roots: designing an artificial language
- The procedure: designing and running an artificial language learning experiment

Afternoon Session (14:00-18:00): Practical introduction

- From lab to net: introducing the fundamentals of online experimenting
- Getting your experiment on the screen: introduction to jsPysch

(Very) brief introduction into the fundamentals of experimental design

Why run experiments?

- Real life data vs. controlled experiments
- Studying questions about language change in the lab
  - **Complementing** the scarce typological and historical data
  - **Directly testing** assumptions about cognitive and social mechanisms
  - **Highlighting** important (previously unnoticed) nuances of linguistic theories

- The name of the game: **decomplexify**

Basic skeleton of an experiment

- A set of procedures to systematically test cause-effect relationships, **by collecting evidence to demonstrate the effect of one variable on another**
- Holding all things constant (experimental control) except for the independent variable (experimental manipulation)

Basic skeleton of an experiment

Manipulate the **independent variable** → Measure the effect on **dependent variable** → Infer about target **construct**

Basic skeleton of an experiment

Manipulate the **independent variable**

Noun frequency

→

Measure the effect on **dependent variable**

Accuracy in production

→

Infer about target **construct**

Learning novel nouns

Basic skeleton of an experiment

Manipulate the **independent variable**

Noun frequency

→

Measure the effect on **dependent variable**

Accuracy in a forced-choice task

→

Infer about target **construct**

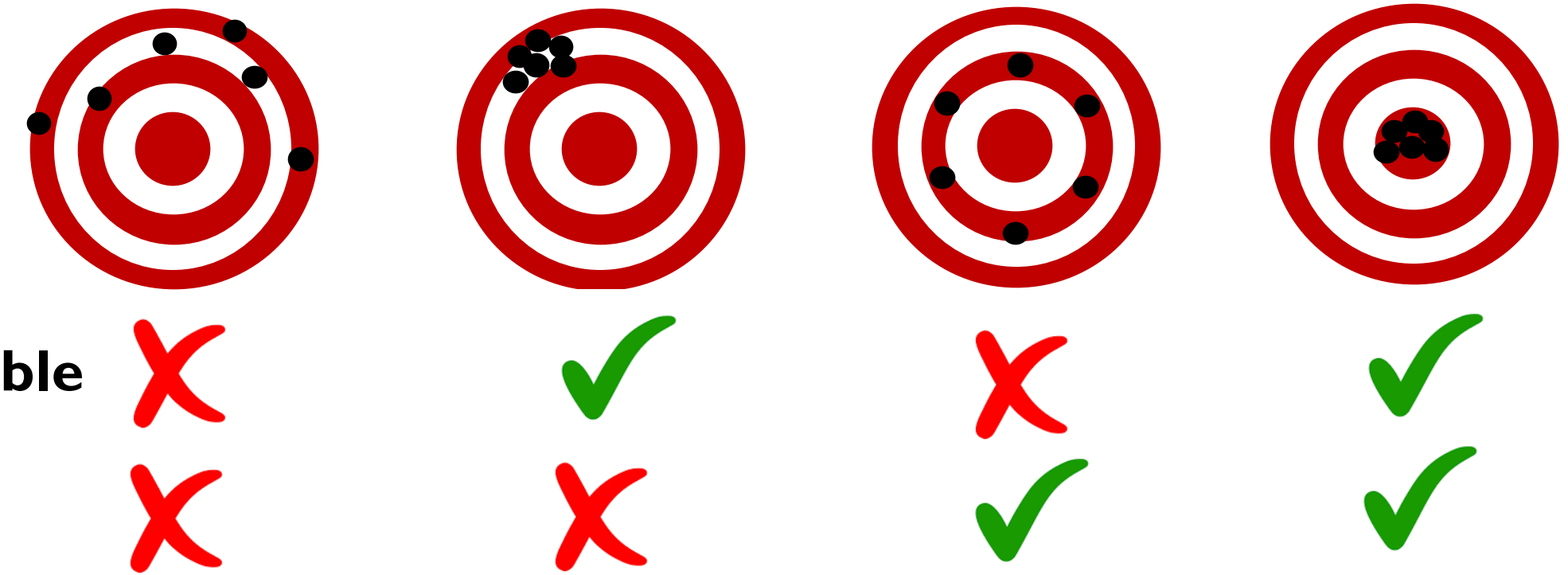
Learning novel nouns

Important concepts

- Reliability: the consistency or stability of an experimental effect
- Validity: whether an experiment measures what it claims to measure

Important concepts

- Reliability: the consistency or stability of an experimental effect
- Validity: whether an experiment measures what it claims to measure



Important concepts

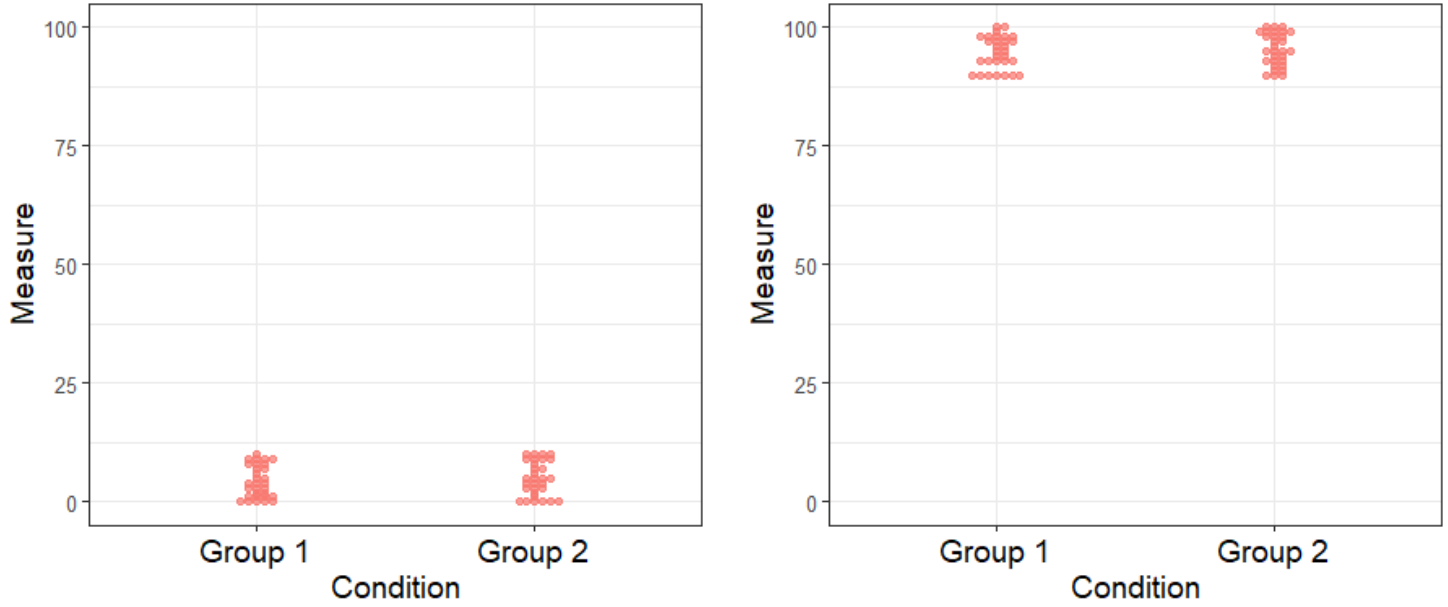
Experimental designs:

- Within subjects: the independent variable is manipulated **within** subjects (all subjects go through the same experiment)

- Between subjects: the independent variable is manipulated **between** groups of subjects (subjects are **randomly** assigned to different experimental conditions)

Important concepts

- Confounds: any extraneous variable that can cause an unintended effect on the results
  - Ways to deal with that: randomization, counter-balancing, controls
- Floor and ceiling effects



Important concepts

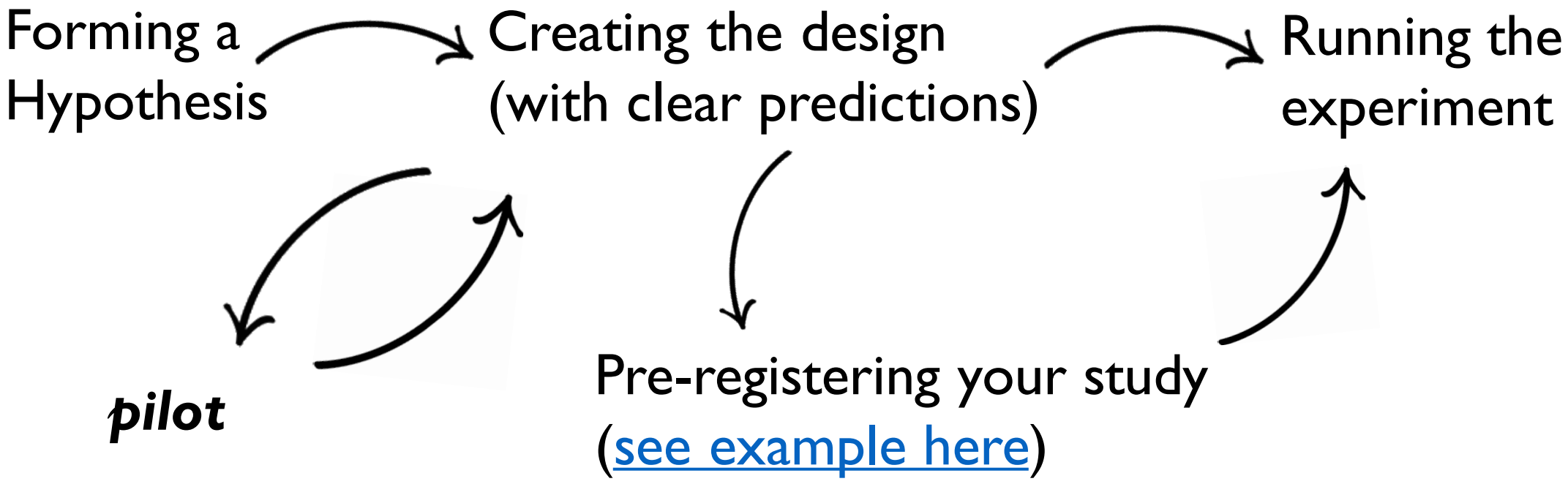
- Inclusion criteria: a set of predefined characteristics used to identify subjects who will be included in a research study

  For example: Certain native language, age, demographic, etc.

- Exclusion criteria: a set of **predefined** reasons for which participants are to be excluded from the study sample
  - Could be defined based on participants' behavior as long as:
    - Relevant to the behaviour measured
    - Distinct from the hypothesized behaviour

  For example: success in attention tests, learning basic building blocks

Timeline of an experimental study



Any questions?

Artificial language learning (ALL) experiments: Paradigms

What are ALL experiments?

- Miniature novel linguistic system (with or without meaning)
- A linguistic petri dish where factors of interest can be isolated and studied
- Important questions and challenges:
  - What type of learning does it represent? (L1/L2)
  - The influence of participants' native languages
  - The influence of **having languages**
- Other options: silent gesture experiments, communication games, computational modelling

ALL Paradigms

1. Ease of learning
2. Regularisation
3. Extrapolation
4. Communication
5. Iterated learning
6. Iterated learning + communication

Ease of learning

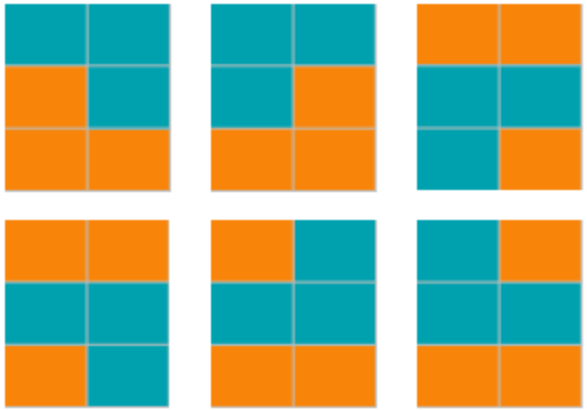
- Learners are trained on patterns of interest
- Speed or accuracy of learning is compared across patterns
- Example: Syncretism patterns (Saldana et al., 2022)

	NATURAL PATTERN		L-TYPE PATTERN		X-TYPE PATTERN	
	Dutch		Hindi		Kapau	
	come PRS		*be* FUT.F		*ford water* PST	
	SG	PL	SG	PL	SG	PL
1	kom	komen	hūṃgī	hoṃgī	qäkamanga	qäkamango
2	komt	komen	hogī	hogī	qäkamangn	qäkamanga
3	komt	komen	hogī	hoṃgī	qäkama	qäkamanga

Natural paradigm:

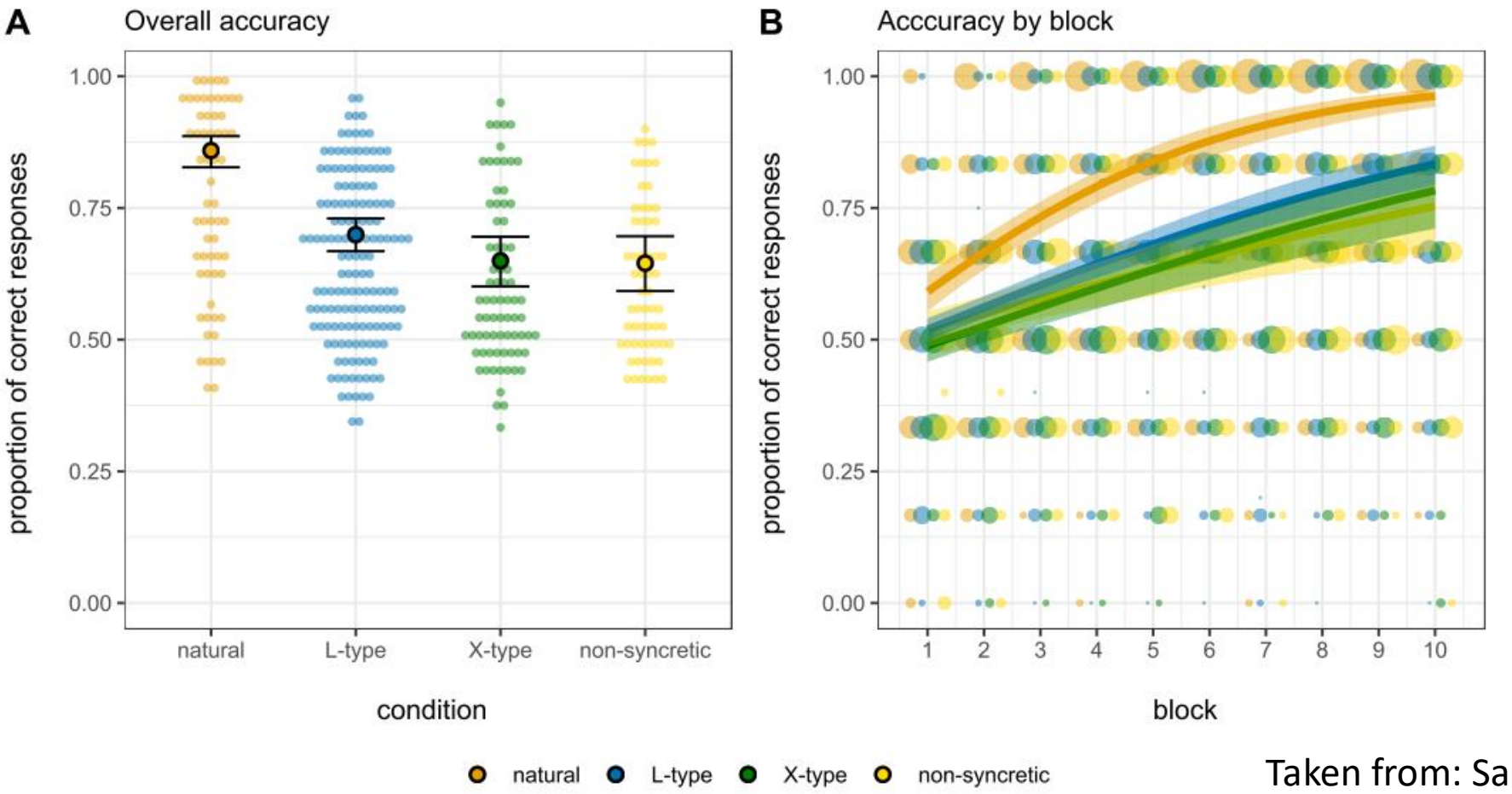
		Number	
		SG	PL
Person	1	1SG	1PL
	2	2SG	2PL
	3	3SG	3PL

L-type paradigms:



X-type paradigms:

Taken from: Saldana, Carmen; Herce, Borja; Bickel, Balthasar (2022). A Naturalness Gradient Shapes the Learnability and Cross-Linguistic Distribution of Morphological Paradigms. Cogsci Proceedings: 787-794.



Taken from: Saldana, Carmen; Herce, Borja; Bickel, Balthasar (2022). A Naturalness Gradient Shapes the Learnability and Cross-Linguistic Distribution of Morphological Paradigms. Cogsci Proceedings: 787-794.

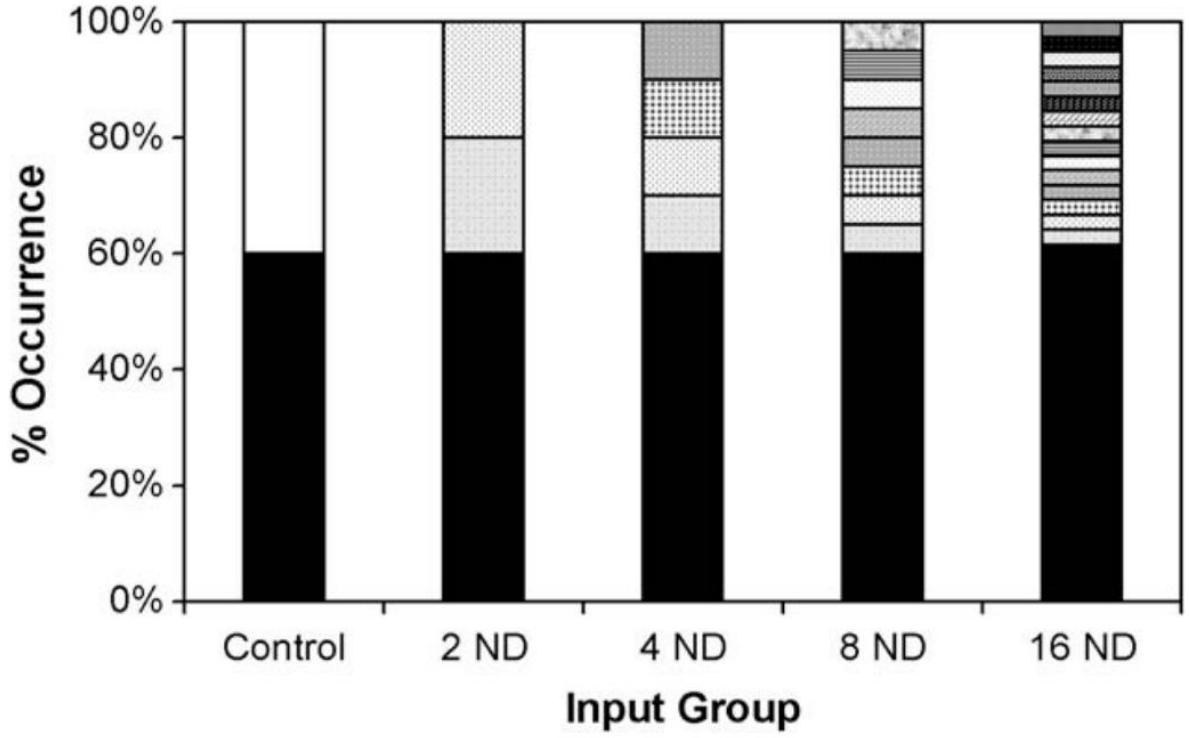
Any questions?

Regularsiation

- Training participants on a language system with variability in it
- Compares amount/type of variability in training and testing
- Change in variability indicates preference/dispreference

Hudson-Kam and Newport, 2009 (Experiment 1)

- Trained on (comparatively large) artificial language
- Manipulation in number and distribution of determiners



- Majority determiners 60% of the time
- Increasing number of minority determiners

Main results

C.L. Hudson Kam, E.L. Newport / Cognitive Psychology 59 (2009) 30–66

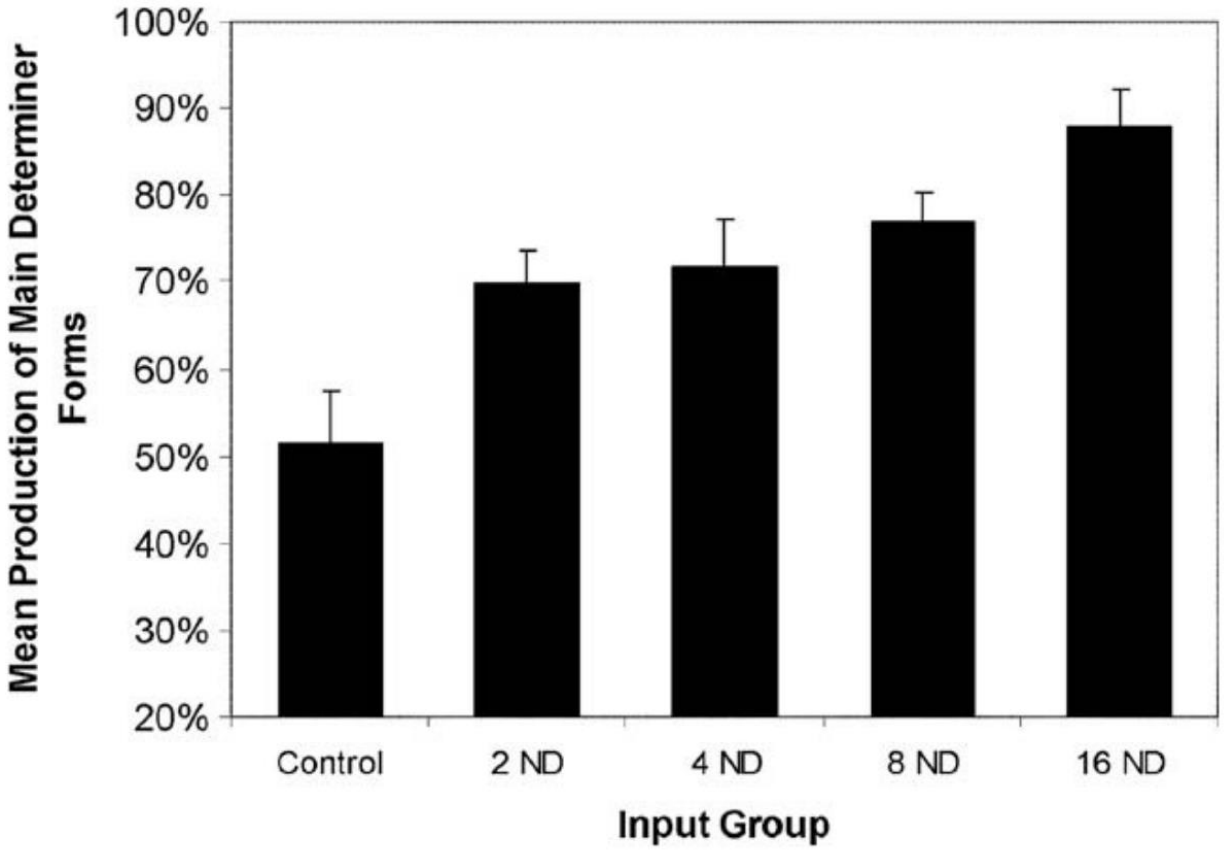


Fig. 4. Mean production of main determiner forms by input group.

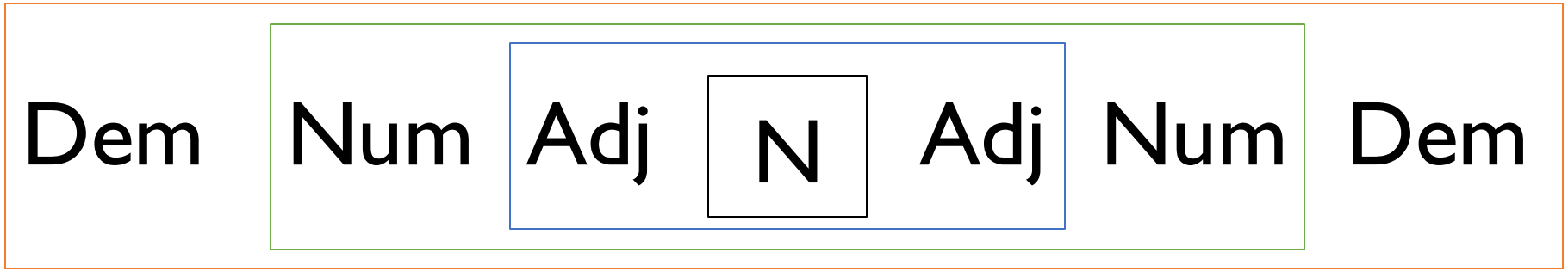
Any questions?

Extrapolation

- Sometimes called "poverty of the stimulus method"
- Training participants on part of the artificial language
- Structure of interest:
  - Withhold
  - Compatible with several hypotheses
- Tested on disambiguating tasks
- What do participants do in absence of evidence?

Martin et al., 2020 (Experiment 2)

- Relative ordering of modifiers in the noun phrase
- Artificial language consisting of nouns, adjectives, numerals, and demonstratives
- Do participants assume a certain relative order between these modifiers?

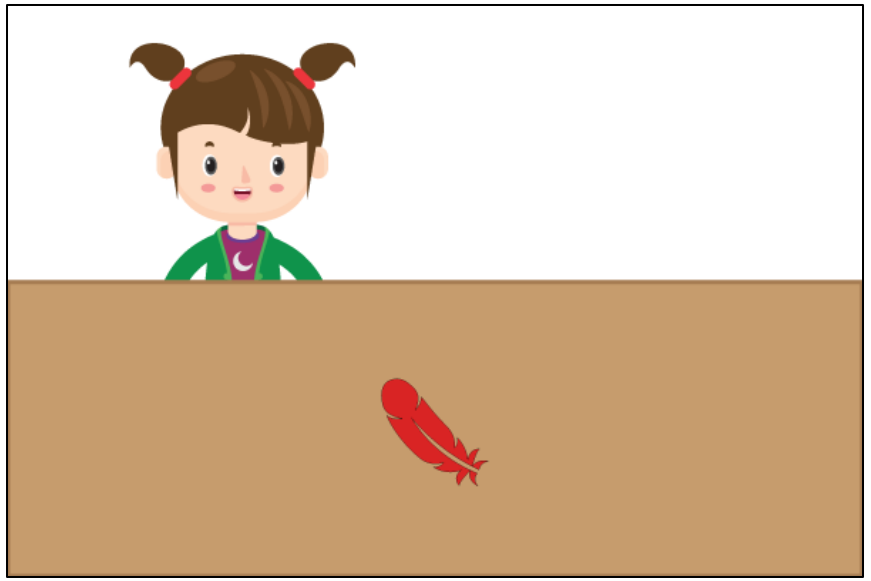


Training: Adj-Dem condition

Demonstrative training

puku hono

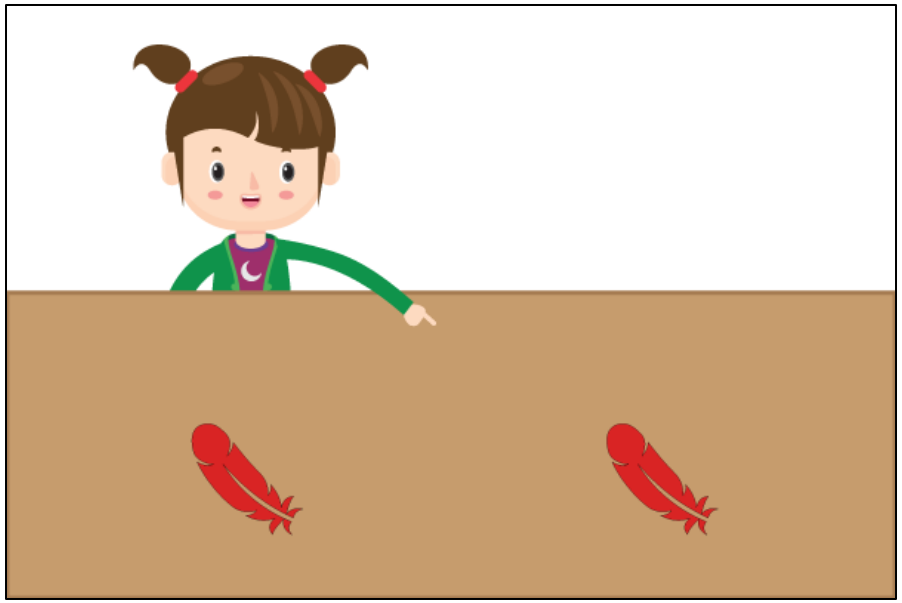
Adjective training



puku taka

Testing: Adj-Dem condition

Testing trial



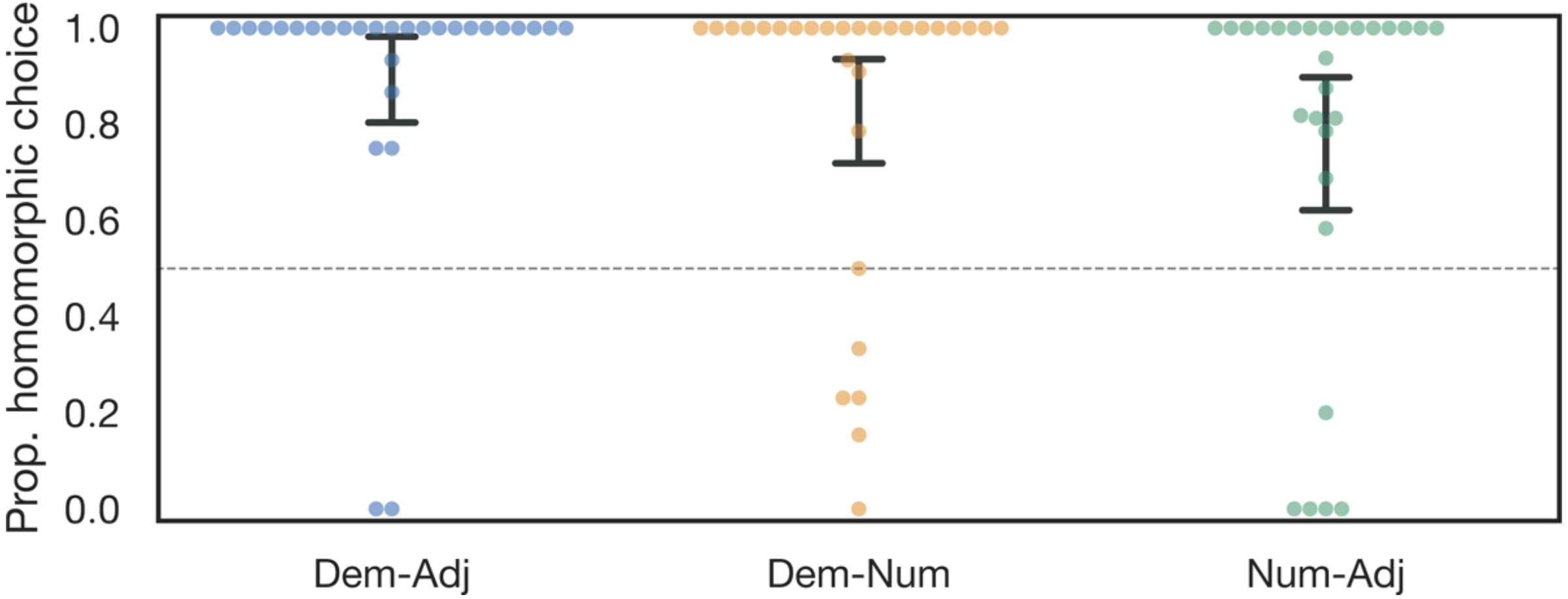
- What relative ordering do they produce?

puku hono taka

or

puku taka hono

Results


Prop. homomorphic choice
1.0
0.8
0.6
0.4
0.2
0.0
Dem-Adj
Dem-Num
Num-Adj

Any questions?

Communication

- Paradigm where participants interact at some or all stages of the experiment

Communication

Principal procedure

1. (Participants are individually trained on AL language)
2. (Outcome of training is measured)
3. Communication
4. Outcome of communication is measured

Communication

Communication set-up

- Language
  - Language provided to participants (Atkinson, Smith & Kirby 2018)
  - Participants create own language (Selten & Warglien 2007, Raviv, Meyer & Lev-Ari 2019)
  - Symbols provided to participants (Bowerman & Smith 2022)
  - Participants create own signs/symbols (Galantucci, Kroos & Rhodes 2010)

Communication

Communication set-up

- Number of communicators
  - Dyad (Winters, Kirby & Smith 2018, Fehér, Wonnacott & Smith 2016)
  - Multiple participants, e.g. 4,8 (Raviv et al. 2019)
- Nature of interlocutor
  - Human (Winters et al. 2018)
  - Computer – disclosed (e.g. alien) (Fehér et al. 2016, Tal et al. 2022)
  - Computer – undisclosed (Fehér et al. 2016)
- Typically silent communication through connected computers

Communication

Communication game: director-matcher task

1. Participant A sees image/video/other type of stimuli and describes what they see
2. Description is passed on to participant B who selects image/video from set of provided answers.
3. Participant A and B receive feedback on their performance
4. Participant A and B switch roles

   →Fixed sender & receiver assignment possible (Winters et al. 2018)

Communication

Demonstration of a communication experiment

(Taken from Kenny Smith, based on Kanwal et al. 2017)

Fehér, Ritt & Smith, 2019

- Question: Can asymmetric accommodation during interaction explain constraints on linguistic variation?
- Accommodation: speakers adapt their language use to interlocutor
  - Asymmetric accommodation: speakers who can use feature variably accommodate to speakers who use feature invariably
- Example: Development of optionally used demonstratives into obligatorily used articles in history of English

Fehér, Ritt & Smith, 2019

(1) a Eadmund clypode ænne bisceop […] þa forhtode **se bisceop**
then Edmund summoned a bishop […] then was frightened the bishop
'Then Edmund summoned a bishop […] the bishop was frightened.'

(2) Stonc ða æfter **Ø stane** stearcheort onfand **Ø feondes** fotlast.
jump then behind Ø rock stouthearted, found Ø enemy's footprint
'He jumped behind the/a rock, courageously, and discovered the enemy's footprint'
(Fehér et al. 2019: 3)

Fehér, Ritt & Smith, 2019

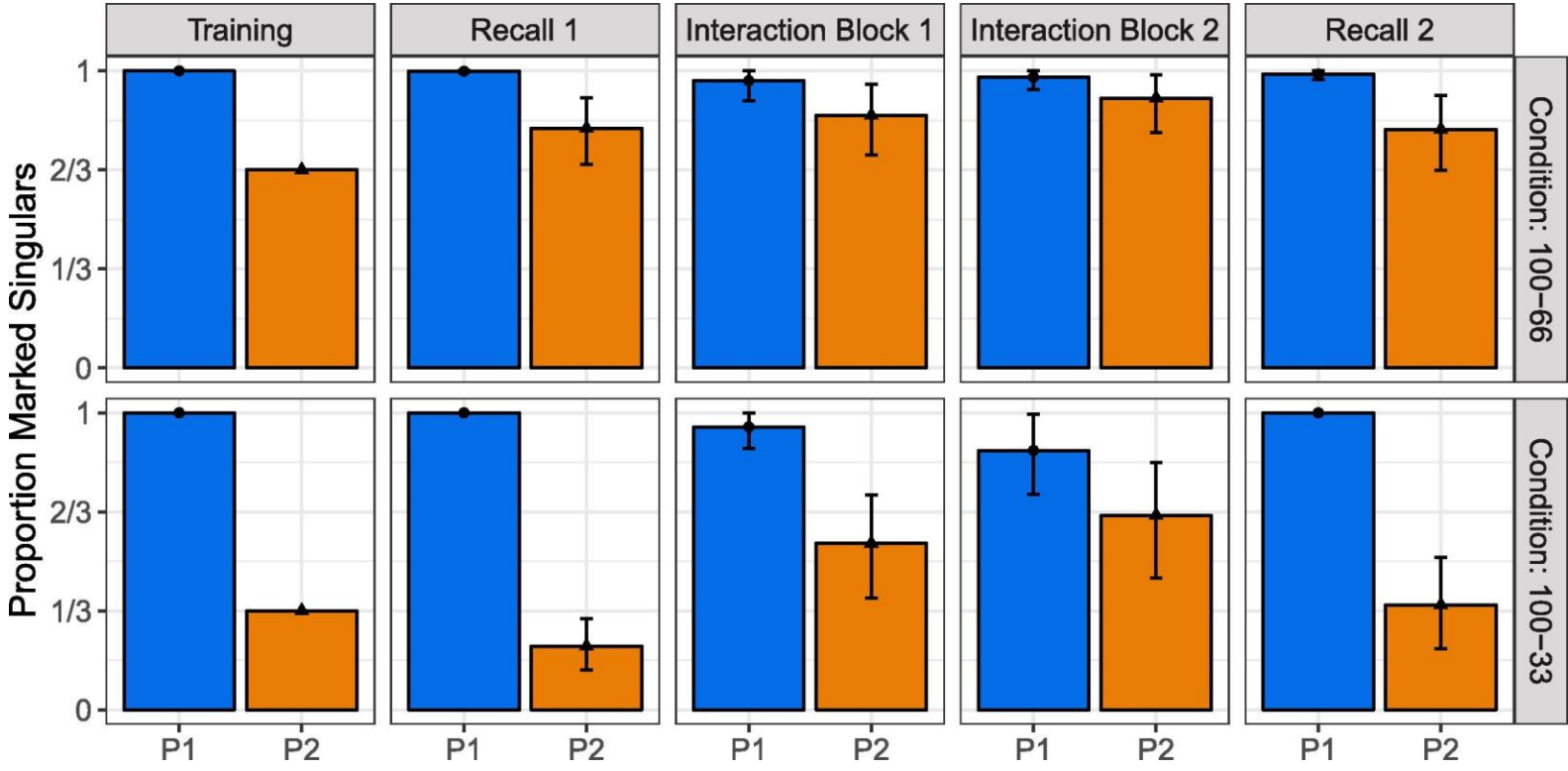
- Experiment 2: Communication between variable and categorical users
- Language:
  - Variable language: random number marking (singular), not conditioned on any factors (frequency: 33%, 66%)
  - Invariable language: consistent number marking (100%)

Fehér, Ritt & Smith, 2019

- General procedure:
  - Noun training
  - Noun testing
  - Sentence training
  - Recall test 1 (production task: describe scenes on screen)
  - Interaction: director-matcher game
  - Recall test 2 (production task: describe scenes on screen)

Fehér, Ritt & Smith, 2019

- Categorical users (P1) remain consistent users
- On average, variable users (P2) accommodate categorical users
- Some evidence for lasting effect



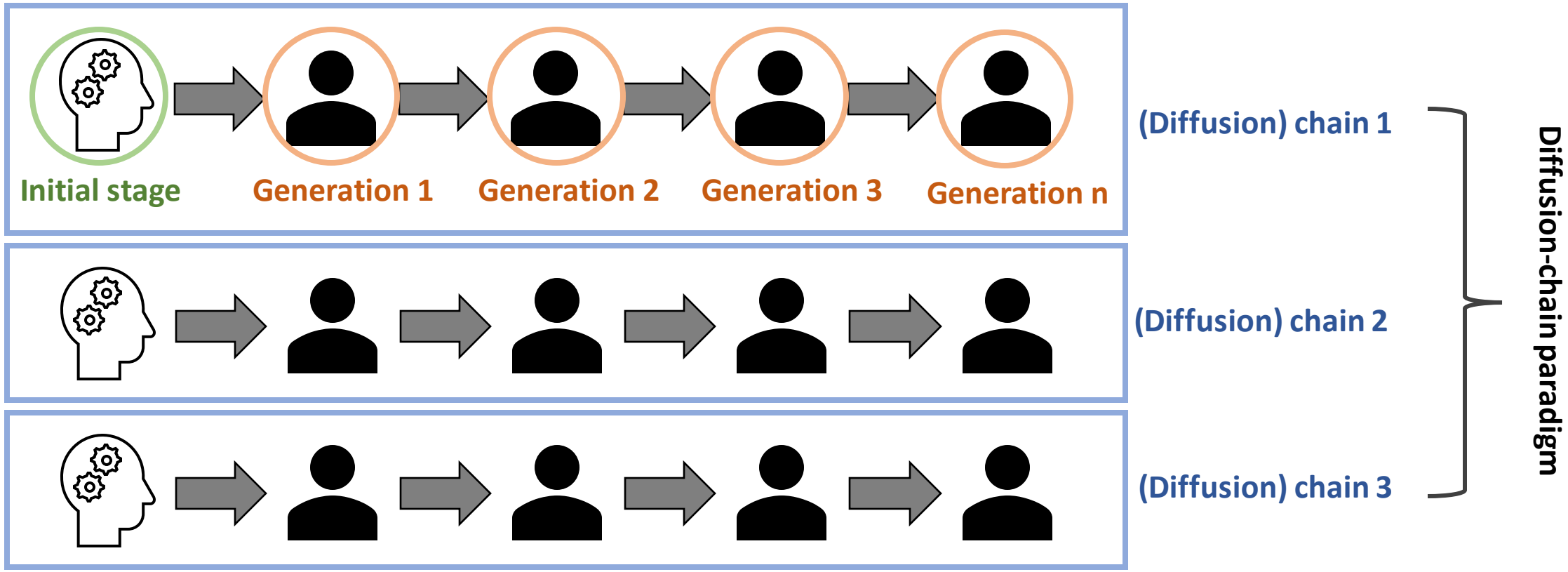
Any questions?

Iterated learning

- Experimental version of Chinese whispers (telephone game)
- Output of one generation forms input of subsequent generation: cultural transmission
  (Kirby, Griffiths & Smith 2014, Smith 2022)
  - Amplification of weak biases through transmission
    (Culbertson & Kirby 2016)

Iterated learning

Terminology



Iterated learning

- Introduced by Bartlett (1932): picture reproduction task
- Used across different modalities & species
  - Horner, Whiten, Flynn & de Waal (2006): tool use for foraging purposes in chimpanzees and human children
- Kirby (2001) first to use paradigm with language to model language evolution
- First experimental study involving human language: Kirby, Cornish & Smith (2008)

Iterated learning

Phases of an iterated learning experiment

1. Initial stage (first generation only)
2. Training phase
3. Testing phase
4. Transmission phase

Iterated learning

Initial stage

- Randomly generated strings → emergence of structure
  - Beckner, Pierrehumbert & Hay (2017): σ → {t,k,s,v} {i,a,o,u} {n,l,∅}; three syllables in total
  - Saldana, Kirby, Truswell & Smith (2019): 1-8 random CV syllables, randomly divided into 1-8 chunks
- Predefined strings → development of existing structure
  - Smith & Wonnacott (2010): fixed lexicon, probabilistic plural marker
  - Roberts & Fedzechkina (2018): fixed lexicon, word order, case marker
  - Berdicevskis & Semenuks (2022): 15 different languages

Iterated learning

Training phase

- Amount of training adjustable:
  - Number of repetitions: Berdicevskis & Semenuks (2022) operationalise imperfect learning by fewer training trials
  - Creation of data bottleneck (poverty of the stimulus) by withholding picture-string pairs (Kirby et al. 2008)
  - Cornish (2010): data bottleneck may not be needed → memory as bottleneck, similar results as Kirby et al. (2008)
- Input from different sources (varieties): more or less socially prestigious group (Roberts & Fedzechkina 2018)

Iterated learning

Testing phase

- Elicitation of image/video descriptions in AL
  - Whole meaning space (Kirby et al. 2008)
  - Unfamiliar meanings (Roberts & Fedzechkina 2018)
- Use of same string to multiple meanings can be blocked to ensure expressivity (Carr, Smith, Cornish & Kirby 2017, Saldana et al. 2019)
- Participants are typically unaware that their productions will be used for next generation

Iterated learning

Transmission phase

- Output (i.e. productions) of generation used to generate input for next generation
- What should be transferred to next gen.? Filtering according to hypothesis!
  - Exclusion criteria not met (e.g. sufficient variability)? → rerun generation if not (Beckner et al. 2017, Roberts & Fedzechkina 2018)
  - Transfer of full system (full transmission) (Kirby et al. 2008)
  - Transfer of relevant properties only
    - Smith & Wonnacott (2010): proportion of different plural markers in participants' productions
    - Roberts & Fedzechkina (2018): proportion of different word orders/patterns

Iterated learning

Transmission phase

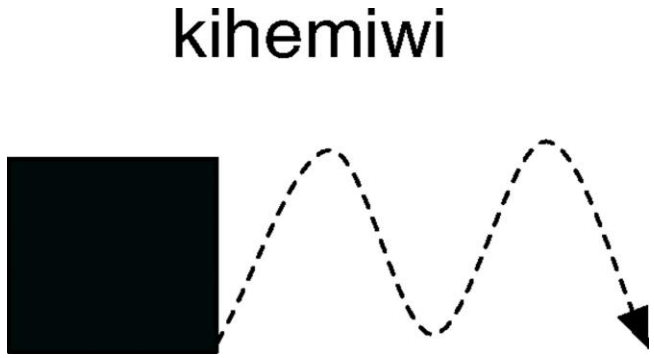
- Number of generations: typically 5-10

Analysis

- One chain = one data point (⚠ statistical power, probability to correctly reject H_0 when H_1 is true)
- See Winter & Wieling (2016) for tutorial on analysing data from iterated learning experiments (GAMS, GCA)

Kirby, Cornish & Smith, 2008

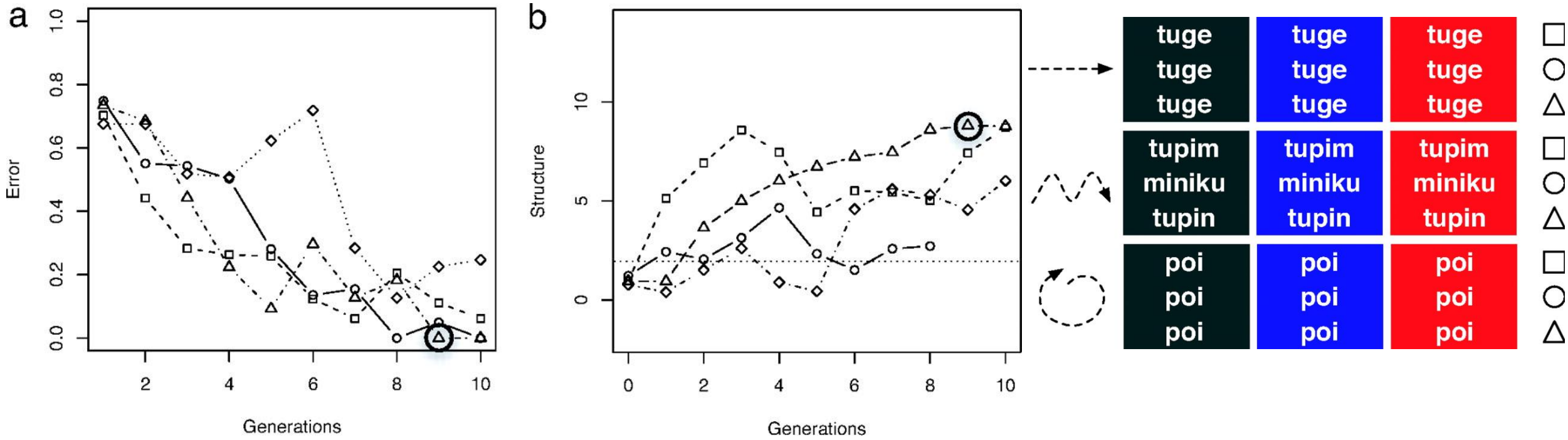
- "Design without designer": can cultural transmission lead to emergence of structure?
- Stimuli:
  - Semantic space: SHAPE, COLOUR, MOTION
  - Random pairing of strings & images
  - SEEN & UNSEEN set



- Procedure:
  - 3 training rounds (2x SEEN set per round)
  - 3 testing rounds (final testing round: elicitation of descriptions for all SHAPE x COLOUR x MOTION combinations)
  - Answers from last testing round = input for next generation

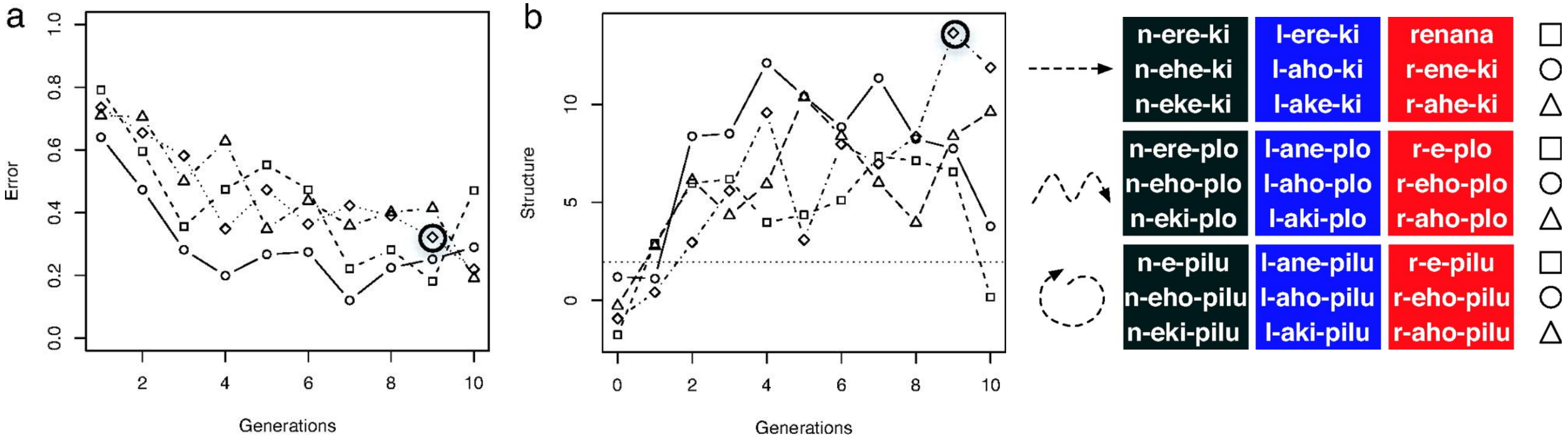
Kirby, Cornish & Smith, 2008

Experiment 1



Kirby, Cornish & Smith, 2008

Experiment 2: removal of strings with multiple meanings from input



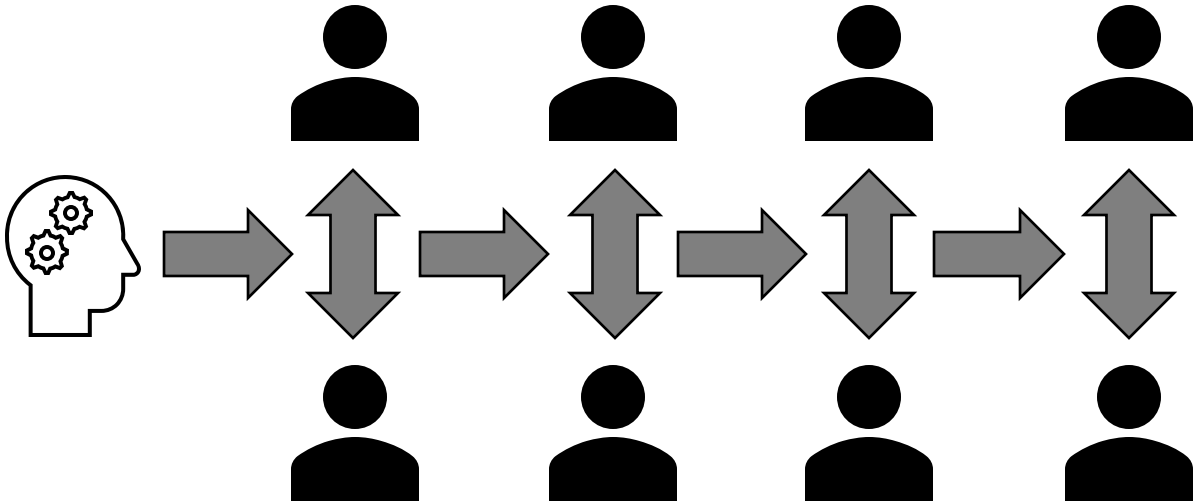
Kirby, Cornish & Smith, 2008

- Experiment 1
  - Decrease of transmission error & increase of structure
- Experiment 2
  - Expressivity constraint implemented
  - Also decrease of transmission error, increase of structure & expressivity maintained
- Findings replicated by Beckner et al. (2017) with larger sample size in online study

Any questions?

Iterated learning + communication

- Combination of two experimental paradigms: iterated learning + communication paradigm
- Allows natural induction of pressure for expressivity (Kirby, Tamariz, Cornish & Smith 2015, Carr et al. 2017, Saldana et al. 2019)



Iterated learning + communication

Procedure

- Training phase: Participants are individually trained on language
- Communication phase: director-matcher game (cf. communication)
- Transmission phase:
  - Transfer of productions of only one participant (Saldana et al. 2019, Silvey, Kirby & Smith 2019, Ota, San José & Smith 2021)
  - Mixing data from multiple participants for input of next generation slowed down rise of regularity/simplification in simple iterated learning experiments (Smith et al. 2017, Atkinson et al. 2018)
  - Transfer of language from successful trials (Berdicevskis 2012)

Kirby, Tamariz, Cornish & Smith, 2015

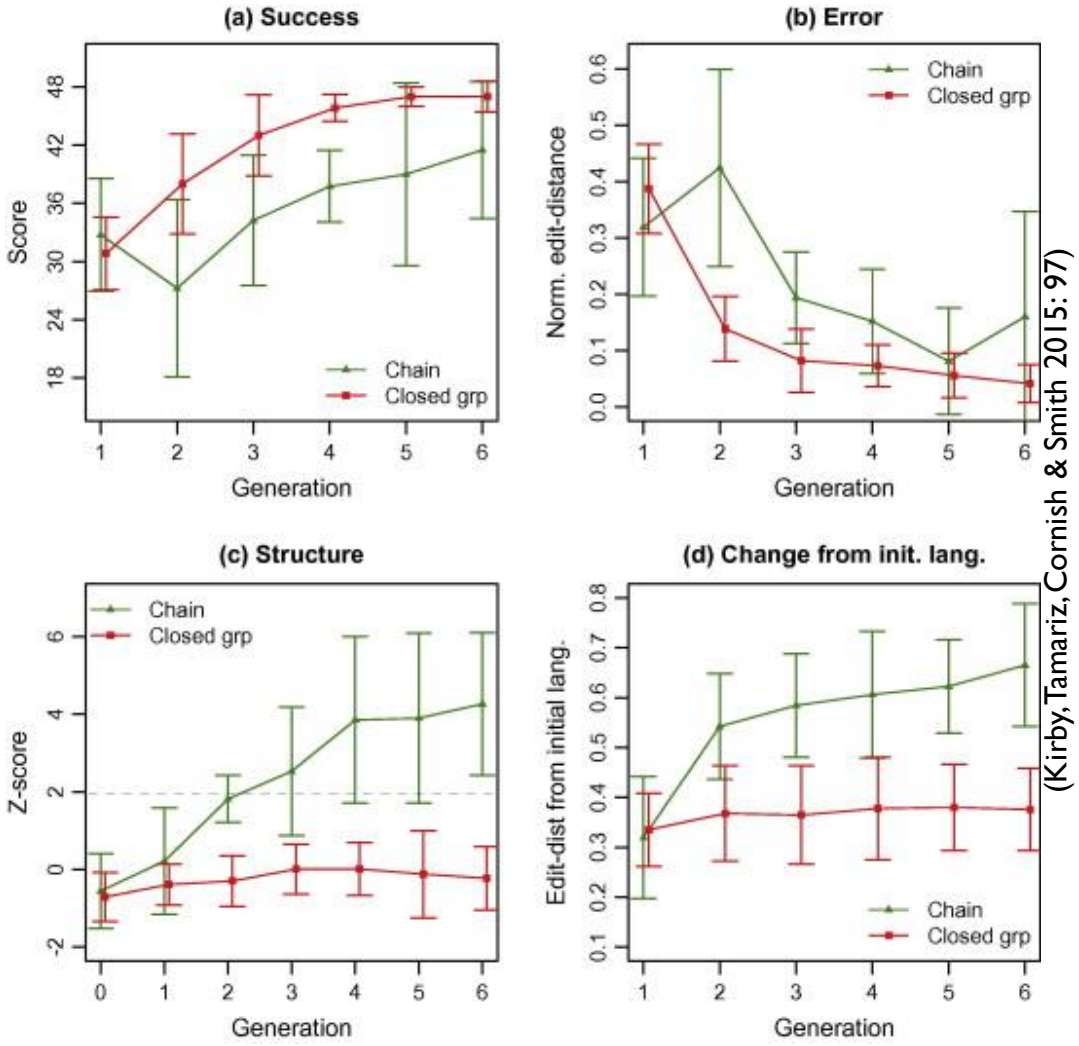
- Follow-up study to Kirby et al. (2008)
- Question: Does the combined pressure of communication and cultural transmission give rise to expressive, yet structured language?
- Stimuli:
  - Semantic space: SHAPE, TEXTURE
  - Unique appendage for each of 12 SHAPE X TEXTURE combinations
  - Random strings assigned to image

Kirby, Tamariz, Cornish & Smith, 2015

- Procedure:
  - Conditions:
    - *Chain*: iterated learning
    - *Closed group*: participants repeatedly retrained on own communication syst.
  - Dyads but trained individually on same input (6x repetitions for each string-image pair)
  - Communication with director-matcher task: A provides label for B who needs to identify correct meaning from array of 6 images
  - Randomly selected participant as source for transmission/retraining

Kirby, Tamariz, Cornish & Smith, 2015

- Results:
  - Communicative success and transmission error improve in both conditions
  - Success rate significantly higher in closed group condition
  - Structure increase only in chain condition



(Kirby, Tamariz, Cornish & Smith 2015: 97)

Kirby, Tamariz, Cornish & Smith, 2015

- Results:
  - Communicative success and transmission error improve in both conditions
  - Success rate significantly higher in closed group condition
  - Structure increase only in chain condition

Closed group condition

pihino	kapa	newhomo
nemone	gakho	kamone
piga	wuwele	gaku
kawake	nepi	hokako

Chain condition

ege-wawu	mega	gamene-wawu
ege-wawa	mega-wawa	gamene-wawa
ege-wuwu	mega-wuwu	gamene-wuwu
ege	wulagi	gamane

(Kirby, Tamariz, Cornish & Smith 2015: 95)

Any questions?

On Schleyer's roots: designing an artificial language

Konstanz and AL: the beginning

- Johann Martin Schleyer (1831-1912)
- Pastor in Litzelstetten
- Inventor of constructed language Volapük

(3) Lif ela Schleyer äbinon vemo nitedik

'Schleyer's life was interesting'

AL recipe

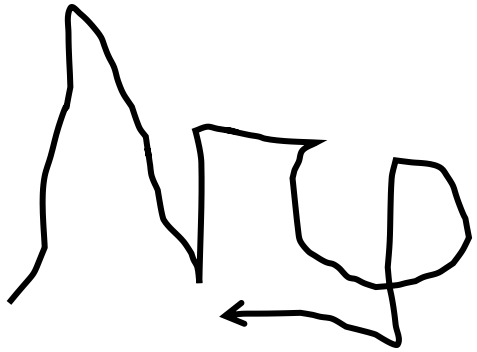
Take…

- Lexical items
- Phonology
- Morphemes
- Syntax

and mix them together.

Semantics

- Existing referents & actions
- Novel referents & actions
  - Objects/shapes & actions participants are unfamiliar with
  - Advantage: no L1 terms readily available
  - Novel Object and Unusual Name (NOUN) Database (Horst & Hout 2016)

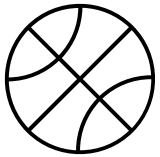


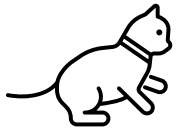
Semantics

- No referents
  - Statistical learning: stream of continuous syllables (Saffran, Aslin & Newport 1996)
  - MER HOX JES LUM TAF KER follows grammar, just no associated meaning (Thompson & Newport 2007)

Lexicon

Lexical items taken from natural languages (e.g. Culbertson & Adger 2014)


ball

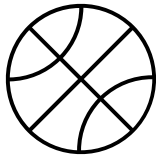

cat

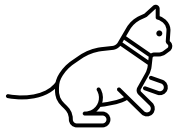

trumpet


kick

→alien elements in syntax/morphology

Semi-artificial lexical items (e.g. Smith & Wonnacott 2010, Atkinson et al. 2018)


bil

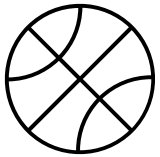

kit

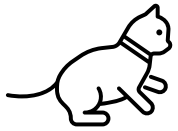

bup


wafa

→iconic lexical items: woof 'dog'

Fully-artificial lexical items (e.g. Tabullo et al. 2012, Getz 2018)


kako


stapi


meko


apu

Issues & practical considerations

Lexicon

- Which degree of artificiality is appropriate?
- Well… it depends →*What is relevant for the research question?*
  - Semi-artificial/natural language:
    - eases pressure of learning lexicon
    - focus on higher level orders/patterns
    - more variability in training pattern (decrease chances of lexical effects)
  - Experiments with natural language stimuli and artificial language stimuli show similar results (English: Culbertson & Adger 2014, AL: Martin, Holtz, Abels, Adger & Culbertson 2020)

Issues & practical considerations

Lexicon

- What do you think should the size of the lexicon for an artificial language be?
- Go to **menti.com** and enter **2140 5759**

Issues & practical considerations

Target audience

- Children vs. adults
  - Significantly reduced complexity necessary for ALL with children
  - Hudson Kam & Newport (2005):
    - Adults: 36 N, 7 intr.V, 5 tr.V, 1 Neg, 2 Det → two noun classes
    - Children: 12 N, 2 intr.V, 2 tr.V, 1 Det → one noun class
  - Culbertson & Newport (2017):
    - Adults: 10 N, 5 A, 5 Num
    - Children:: 4 N, 3 A, 3 Num

Issues & practical considerations

Target audience

- Online vs. offline testing
  - Not only usual student population that participates in online experiments
- Cultural background of participants
  - Colours & symbols have culture-specific meanings, e.g. red cross (cf. Bowerman & Smith 2022)
- One vs. multiple populations
  - Avoiding structures that exist in one of the tested languages but not in others (Martin & Culbertson 2020)

Issues & practical considerations

Influence of participants' L1 on ALL

- Tabula rasa assumption does not hold: ALL (by children & adults) influenced by knowledge of other languages (Siegelman, Bogaerts, Elazar, Arciuli & Frost 2018)
  - Finn & Hudson Kam (2008): identification of novel words from speech stream impaired when initial consonant clusters violate phonotactics of L1
  - Elazar et al. (2022): better identification of words form speech stream when syllables have higher co-occurrence in native language
  - Tang & Baer-Henney (2021): L1 lexicon and AL lexicon contribute to wordlikeness ratings of seen and novel items
  - Onnis & Thiessen (2013): dominant word order of native language (English-SVO, Korean-SOV) affects parsing of ambiguous syllable sequences

Issues & practical considerations

Can we dissociate L1 and AL at all?

- Ensure that feature of interests not attested in L1
- Same AL but different populations (i.e. different L1s)
  - Culbertson, Franck, Braquet, Navarro & Arnon (2020) & Martin, Holtz, Abels, Adger & Culbertson (2020)
- Silent gesture →different modality! (Goldin-Meadow, So, Özyürek & Mylander 2008, Schouwstra & de Swart 2014, Motamedi, Schouwstra, Smith, Culbertson & Kirby 2019)

Any questions?

The procedure: designing and running an artificial language learning experiment

Procedure of an ALL

- Divide and Conquer approach: divide language learning into separate phases
- Traditional approach
  - Noun learning
  - Noun testing
  - Sentence learning
  - Sentence testing

(Though sometimes you might want to use a different order, see for example Arnon & Ramscar, 2012)

Testing participants' knowledge

- In principle, every psycholinguistic measure can be used
- Judgement
  - Forced-choice task (typically 2 options)
  - Binary judgement →Likert scale dispreferred
- Production
  - Oral productions
  - Typed productions
  - Bag-of-words (puzzle piece response)

Two examples

I. **Learning a V2 language**

non artificial lexicon, online study, adult participants

II. **Testing whether redundant morphology benefits learning**

Semi-artificial lexicon, in person, child and adult participants

Learning V2 in the lab

Question

How do changes to the distribution of preverbal elements affect the learning of a V2 language? → Loss of V2 e.g. in Engl., Fr.

Hypothesis

A V2 language in which the evidence for V2 is maximal be easier to learn

→Evidence for V2 = non-subject-initial sentences

→Learning V2 = extrapolating flexibility of initial position to novel types

Learning V2 in the lab

General set up

- Three conditions:
  - Uniform: Subjects, objects, adjuncts occur equally frequent preverbally
  - Object-dominant: 60% object-initial, 20% subject-initial, 20% adjunct-initial
  - Adjunct-dominant: 60% adjunct-initial, 20% subject-initial, 20% object-initial
- 3 phases:
  - Training phase
  - Testing phase
  - Post-test questionnaire

Learning V2 in the lab

Materials

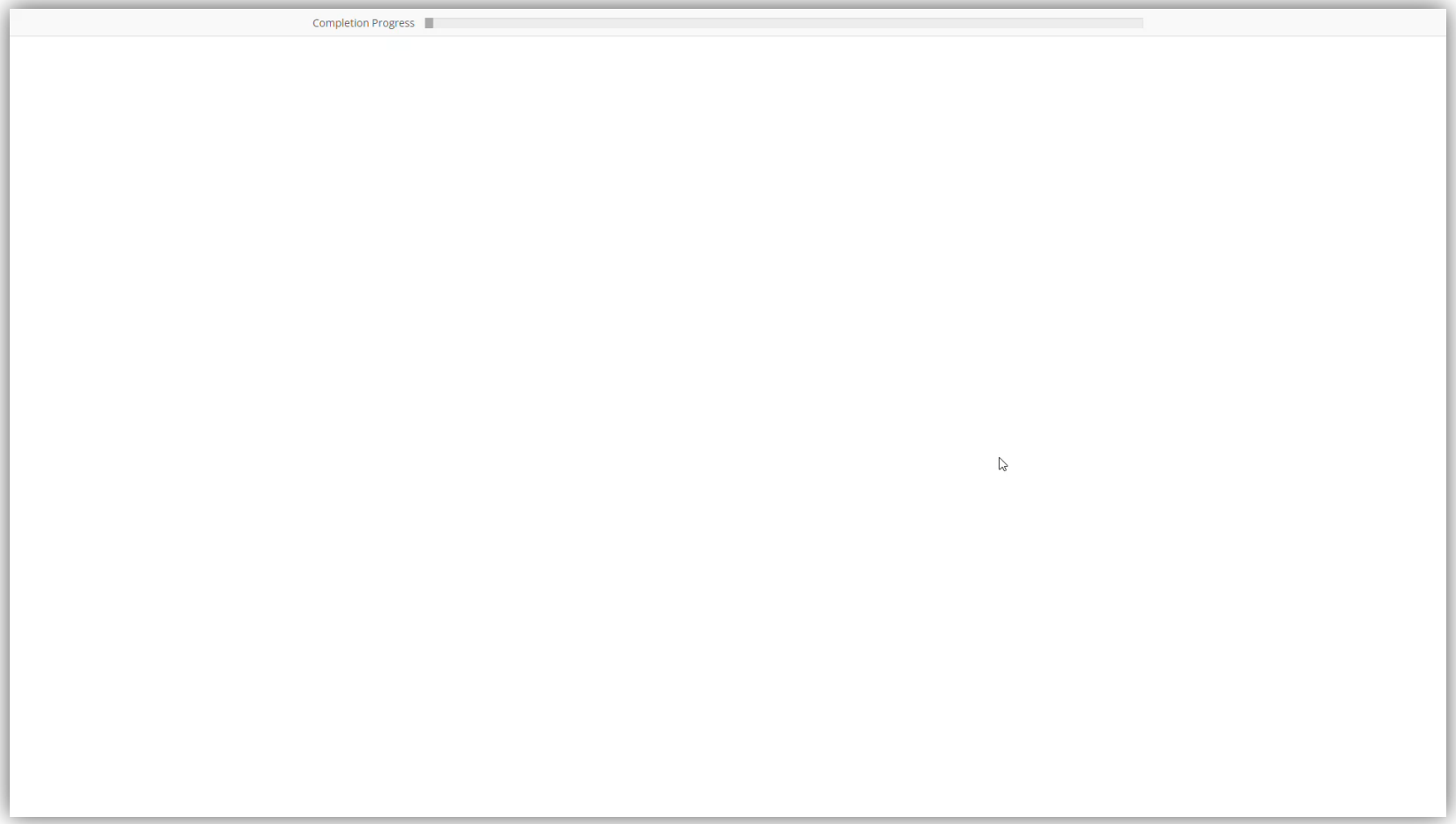
- Semi-artificial language: English lex. + V2 syntax
- 90 V2 sentences constructed from 30 {S, O, V, A} sets

(4) a. The author **revises** eventually a novel in Boston.

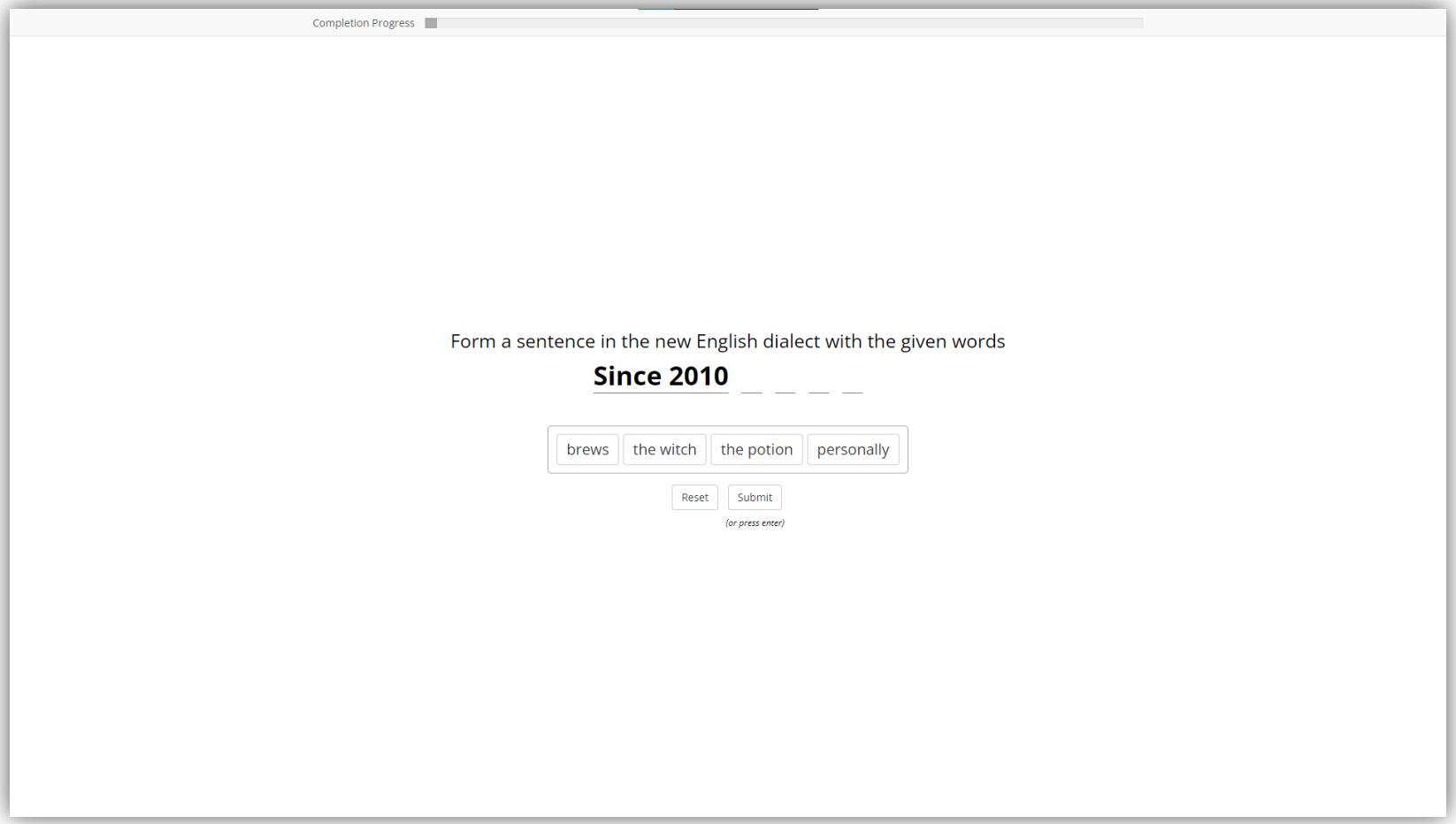
b. A novel **revises** the author eventually in Boston.

c. In Boston **revises** the author eventually a novel.

Learning V2 in the lab



Learning V2 in the lab



Testing phase

Production task

- Participants are provided with scrambled English words and must form sentence in artificial language
- Seen constituent types (4 trials):
  - S, O, A (e.g. *Sophia*, *a carol*, *on Christmas*)
- Novel constituent types (4 trials each):
  - indirect objects (e.g. *to the prosecutor*)
  - complex adjuncts (e.g. *during the conflict*)

(5) {the waiter, awkwardly, to the guest, **passes**, the saltshaker}

Learning V2 in the lab

Judgement task

- Participants see V2 & V3 sentences and need to judge grammaticality of it (binary choice)
- Seen constituent types in initial position (4 trials each):
  - Direct objects
  - Simple adjuncts
- Novel constituent types in initial position (4 trials each):
  - Indirect objects
  - Complex adjuncts

Learning V2 in the lab

(6) To the congregation **shows** the priest silently the candle.

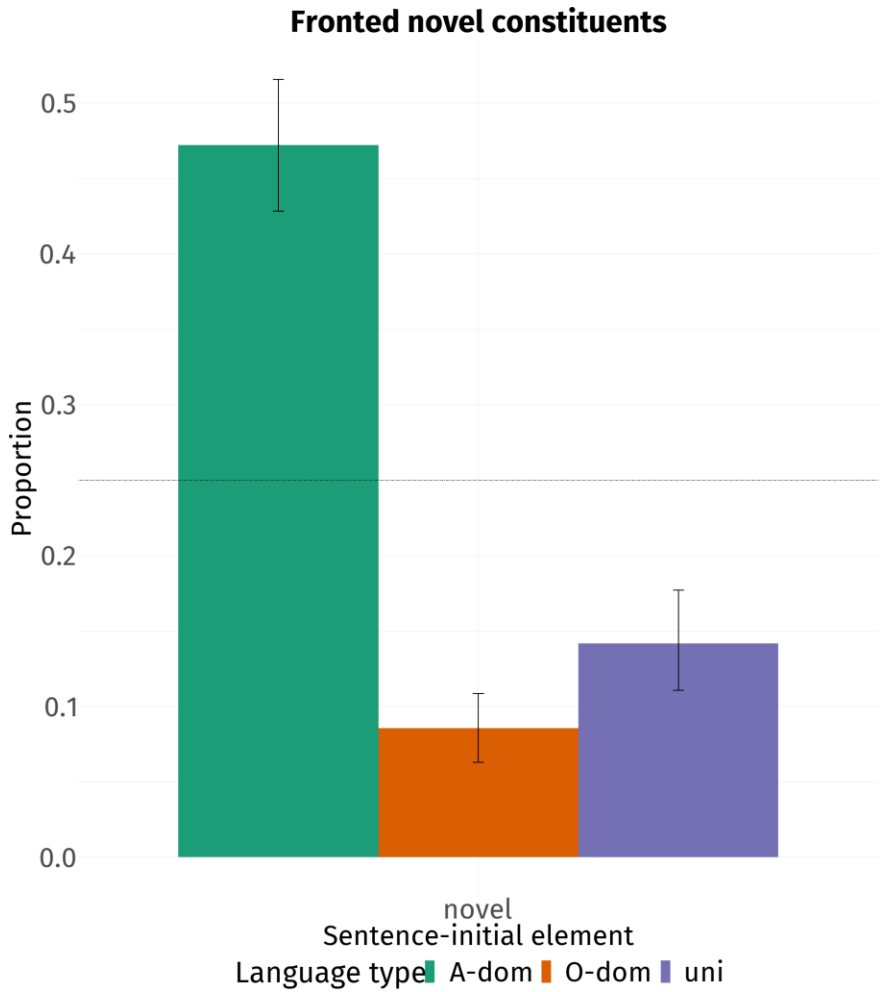
(7) In late April **regrets** the politician openly his misconduct.

(8) To the doctor the patient **describes** precisely the pain.

(9) At the moment the referee **verifies** briefly the decision.

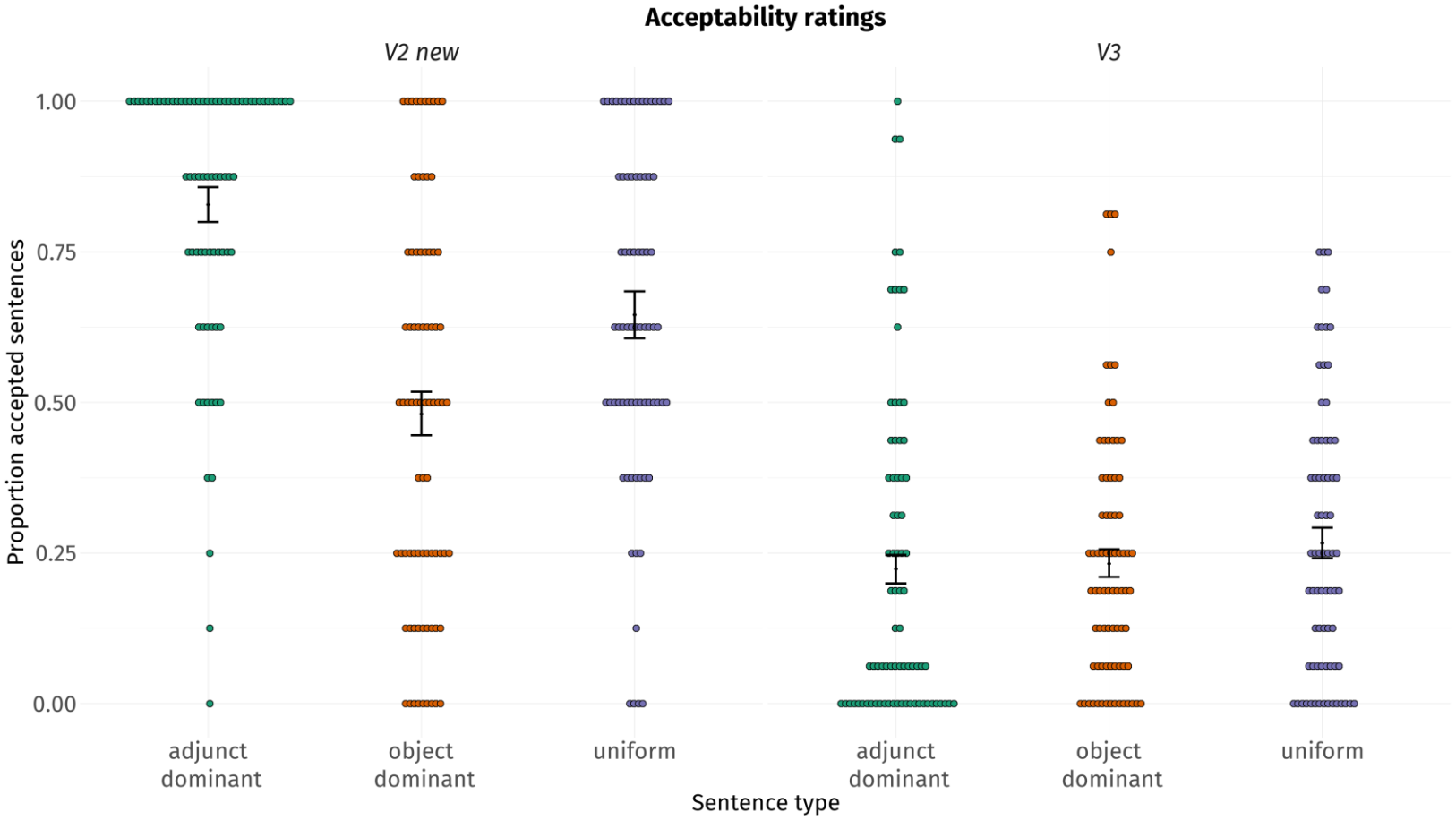
Learning V2 in the lab

- **Prediction**: fewer novel constituents fronted in skewed condition
  - Confirmed for O-dom. but not for A-dom.
  - Apparent advantage for learners in A-dominant condition



Learning V2 in the lab

- **Prediction**: Higher ratings for *V2 new* in uni. condition
  - V2-new: A-dom. > Uni > O-dom.
- **Prediction**: Better discrimination btw. *V2 new* & *V3* in uni. condition
  - Discrimination: A-dom. > Uni = O-dom.



Learning V2 in the lab

Significance

- Learning of complex word order pattern in relatively short time (approx. 30min)
- Complements results from historical records
- Demonstration that language change can be studied using ALL in the lab

Morphological redundancy

Tal & Arnon, Cognition, 2022

Background

Morphological redundancy (e.g., *she talk**s***) is prevalent across languages, despite:

- being dispreferred in language use (e.g. Frank & Jaeger, 2008)
- added complexity (Leufkens, 2020; Lupyan & Dale, 2010)

Why?

Hypothesis: Redundancy can benefit learning of linguistic properties

Morphological redundancy

Case study: learning thematic assignment (who-did-what-to-whom)

Prediction: having two cues (fixed word order + case marking) will lead to better learning compared to having one cue (fixed word order)

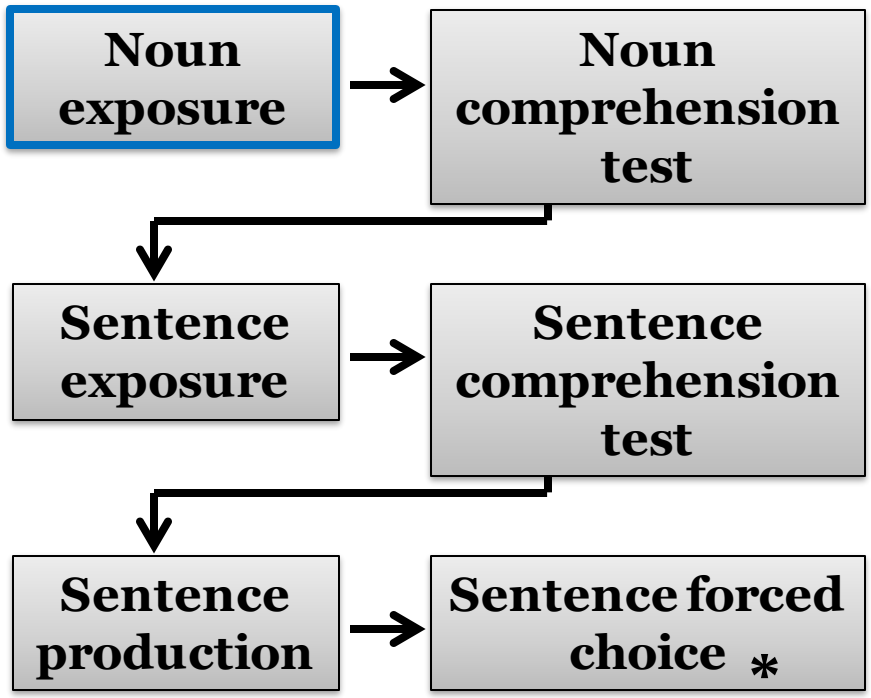
- Crosslinguistic studies: children seem to rely on multiple cues (Chan, Lieven, & Tomasello, 2009 ; Dittmar et al., 2008; Ibbotson & Tomasello, 2009; Matsuo, Kita, Shinya, Wood, & Naigles, 2012; O'Shannessy, 2010)
- BUT the redundant form is usually more frequent in child-directed speech (Dittmar et al., 2008; Ibbotson & Tomasello, 2009)
- Tease these two factors apart by conducting an artificial language learning experiment

Paradigm: Ease of learning

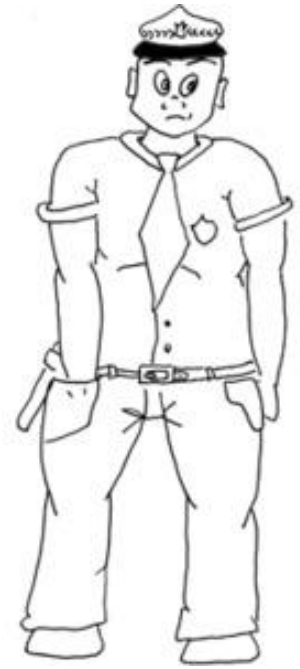
Design

- 60 Children (mean age: 7.10y), 56 adults – all Hebrew speakers
- The language:
  - Semi-artificial lexicon: Hebrew nouns with artificial suffixes (6 nouns, 2 verbs)
  - Fixed OSV word order: Not Hebrew-like
- Two conditions:
  - *Redundant language*: additional case marking on objects (100%)
  - *Control language*: no case marking

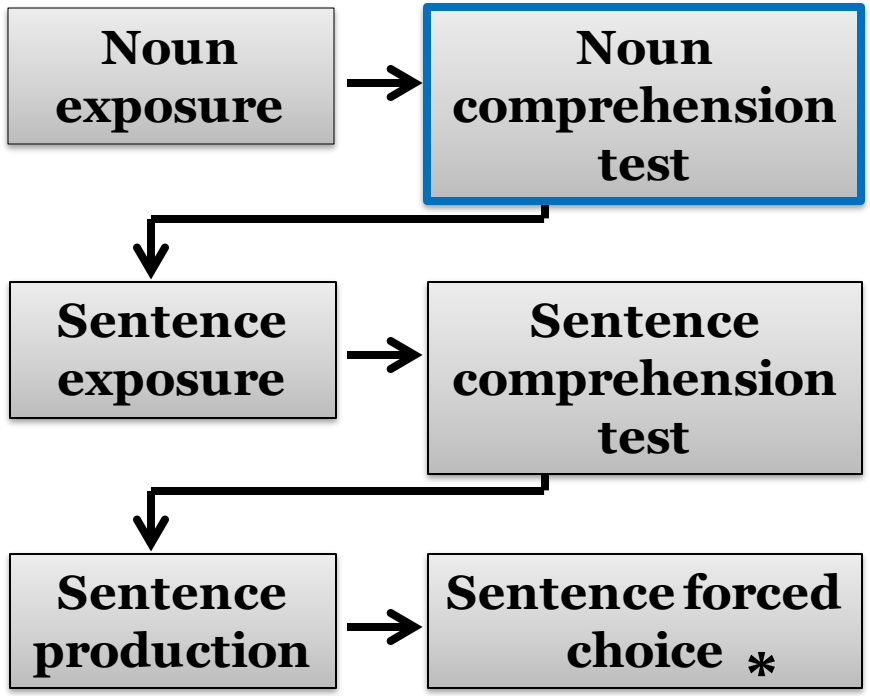
Procedure



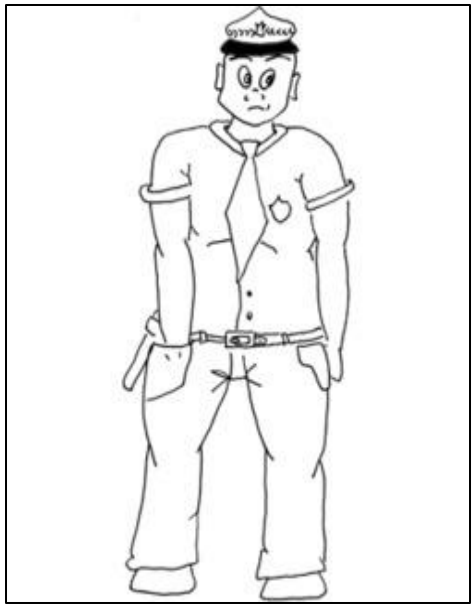
shoterig



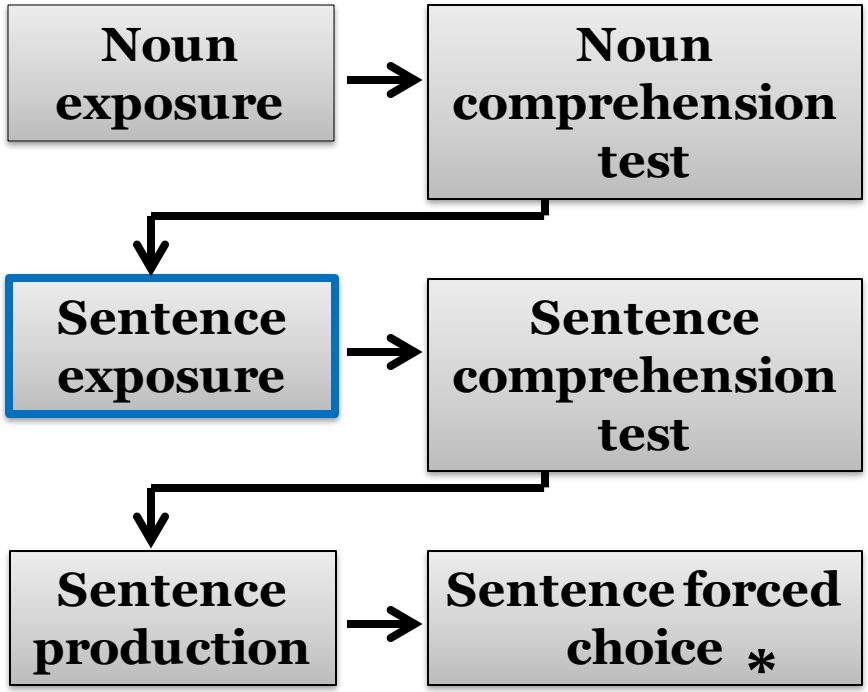
Procedure



shoterig



Procedure



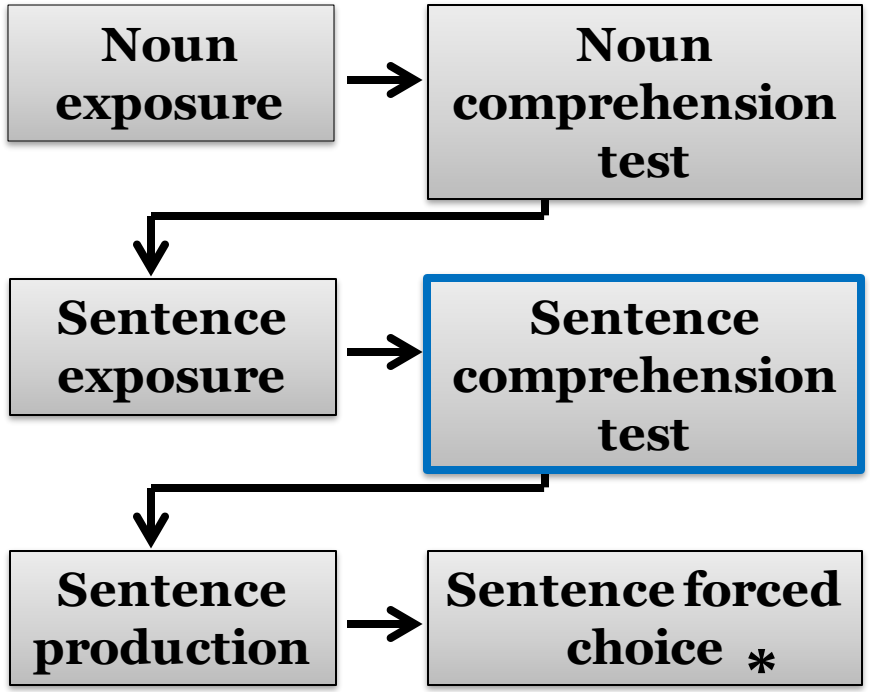
Redundant language

tabaxig**pats** ganavig ba'at

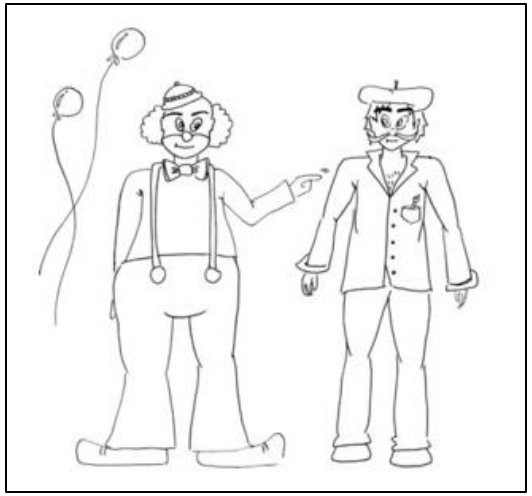
Control language

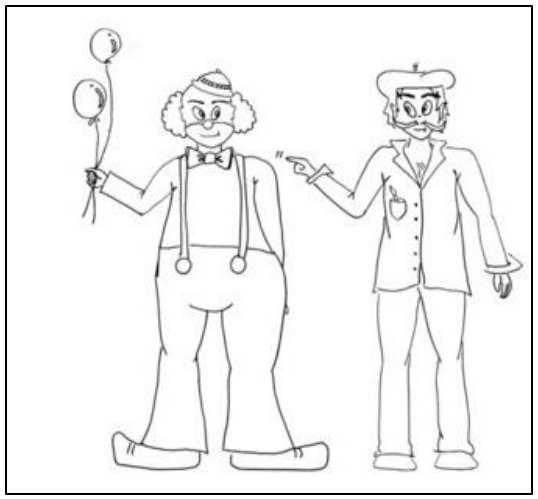
tabaxig ganavig ba'at

Procedure

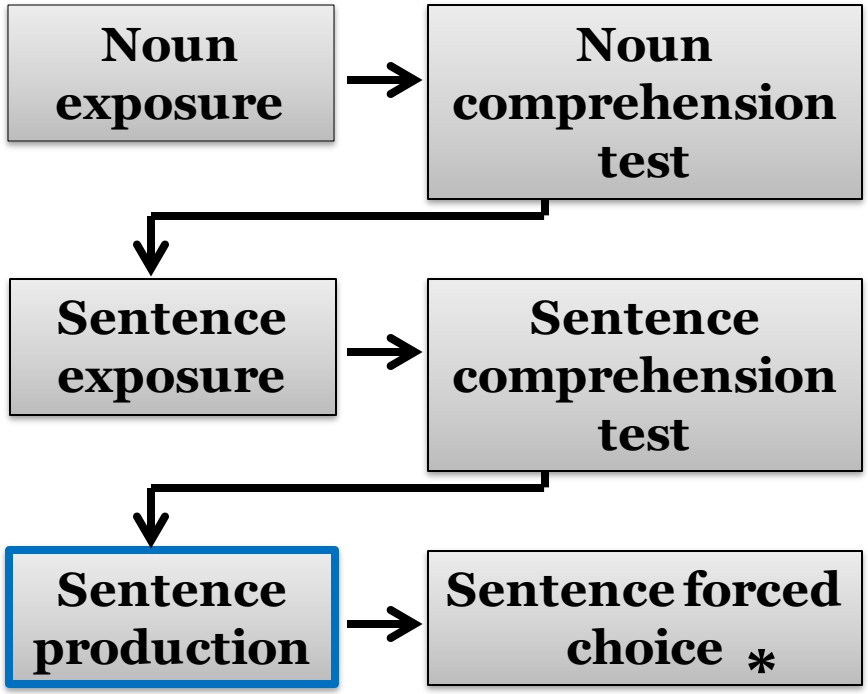


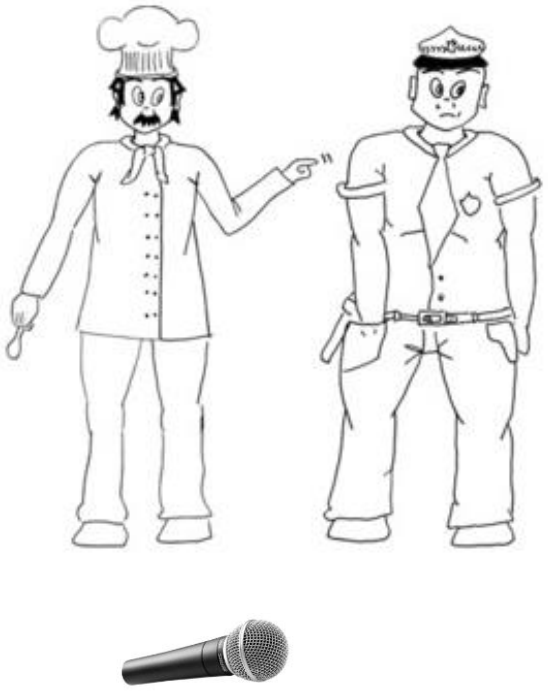
zayarig(pats) leizanig naga



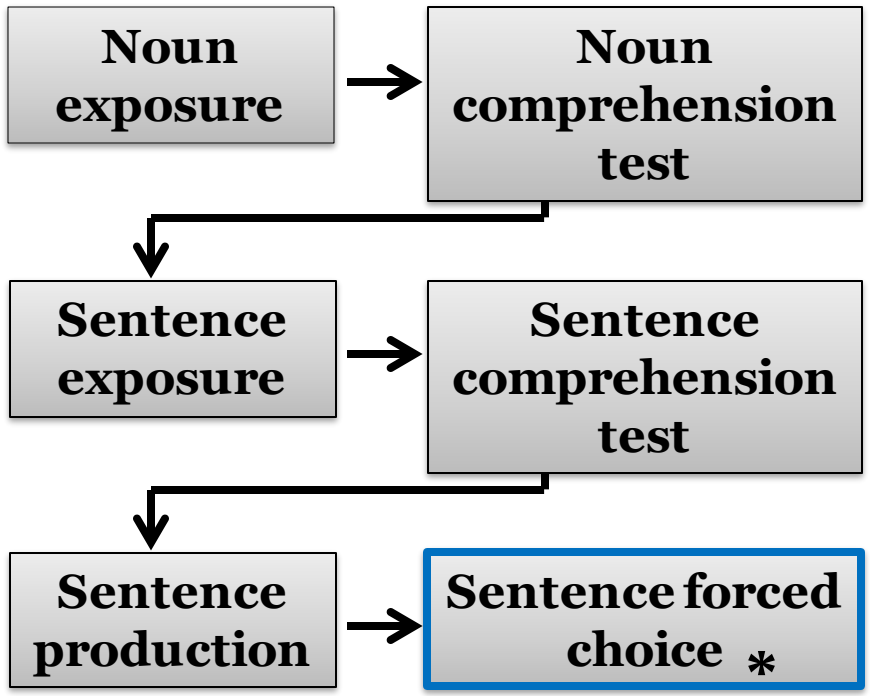


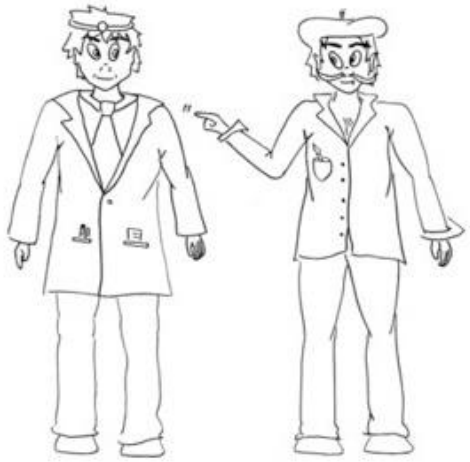
Procedure


Noun exposure
Noun comprehension test
Sentence exposure
Sentence comprehension test
Sentence production
Sentence forced choice *



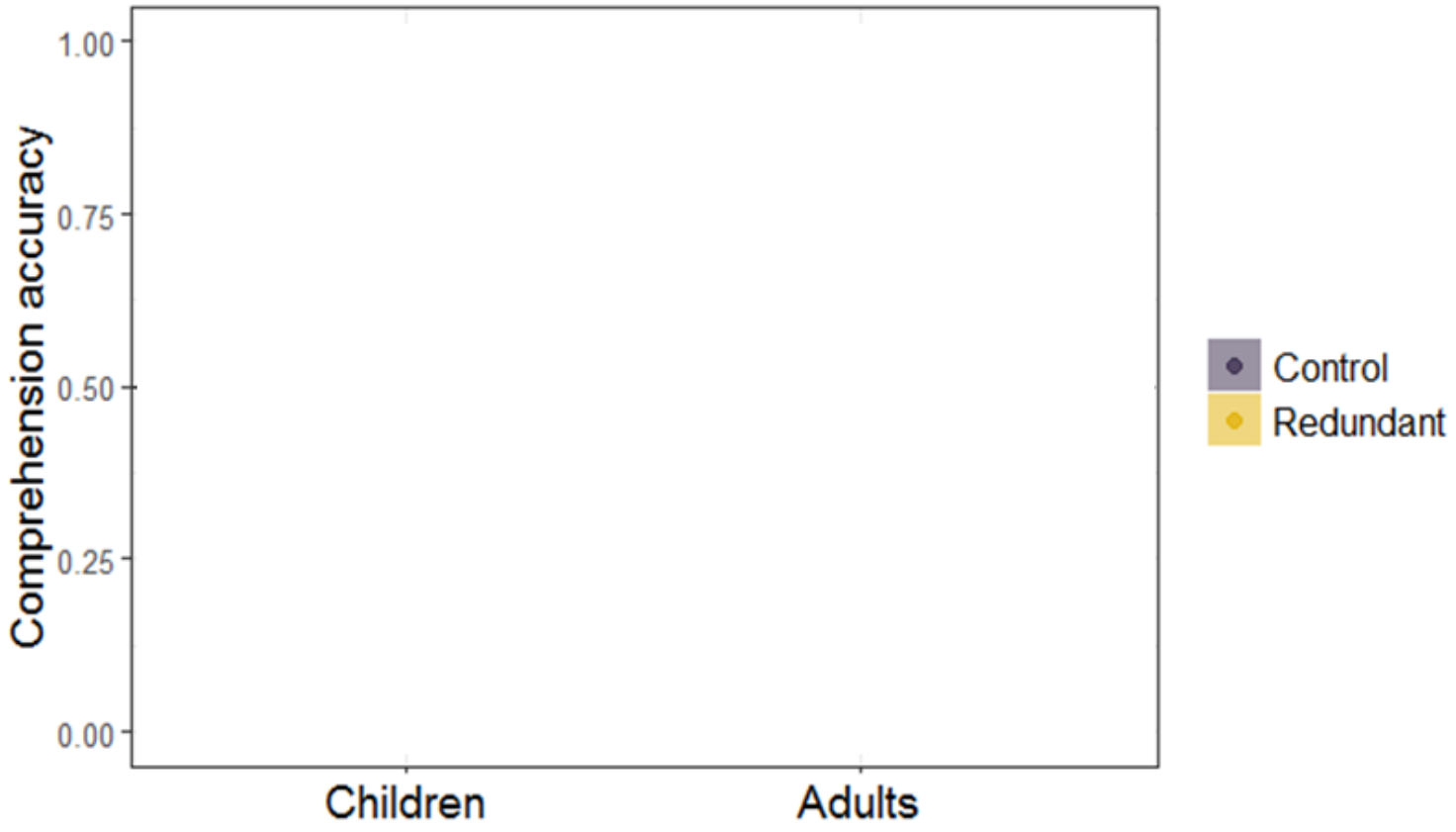
Procedure





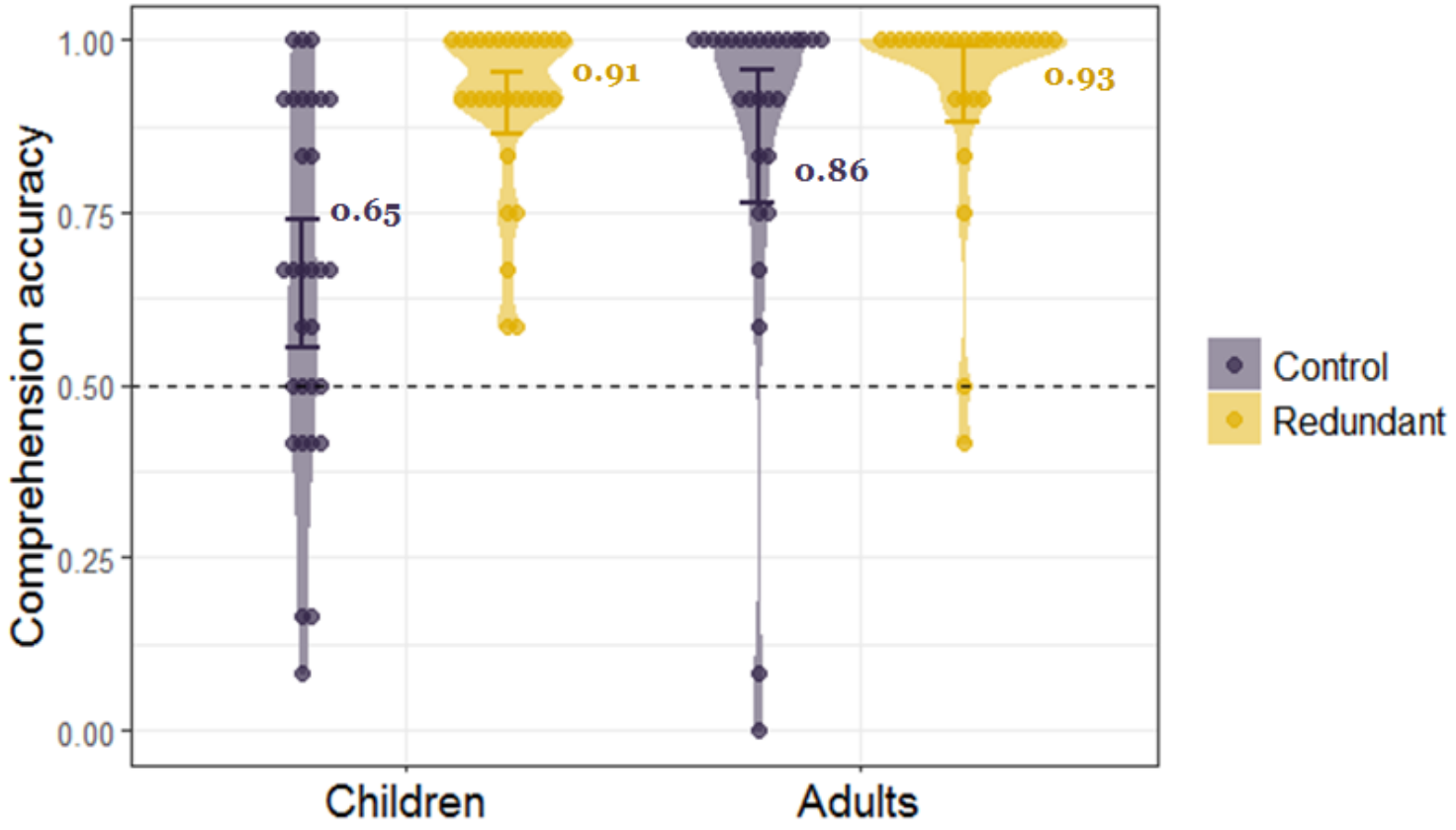
1. rofeig zayarig naga
2. rofeig**pats** zayarig naga

Comprehension results



Comprehension results

- Children in the redundant condition showed better learning
- Adults were at ceiling in both conditions



The utility of using an ALL paradigm

- Directly test a hypothesis about a cognitive mechanism
- Compare the learnability of different language systems
- Compare different types of learners

From lab to net: introducing the fundamentals of online experimenting

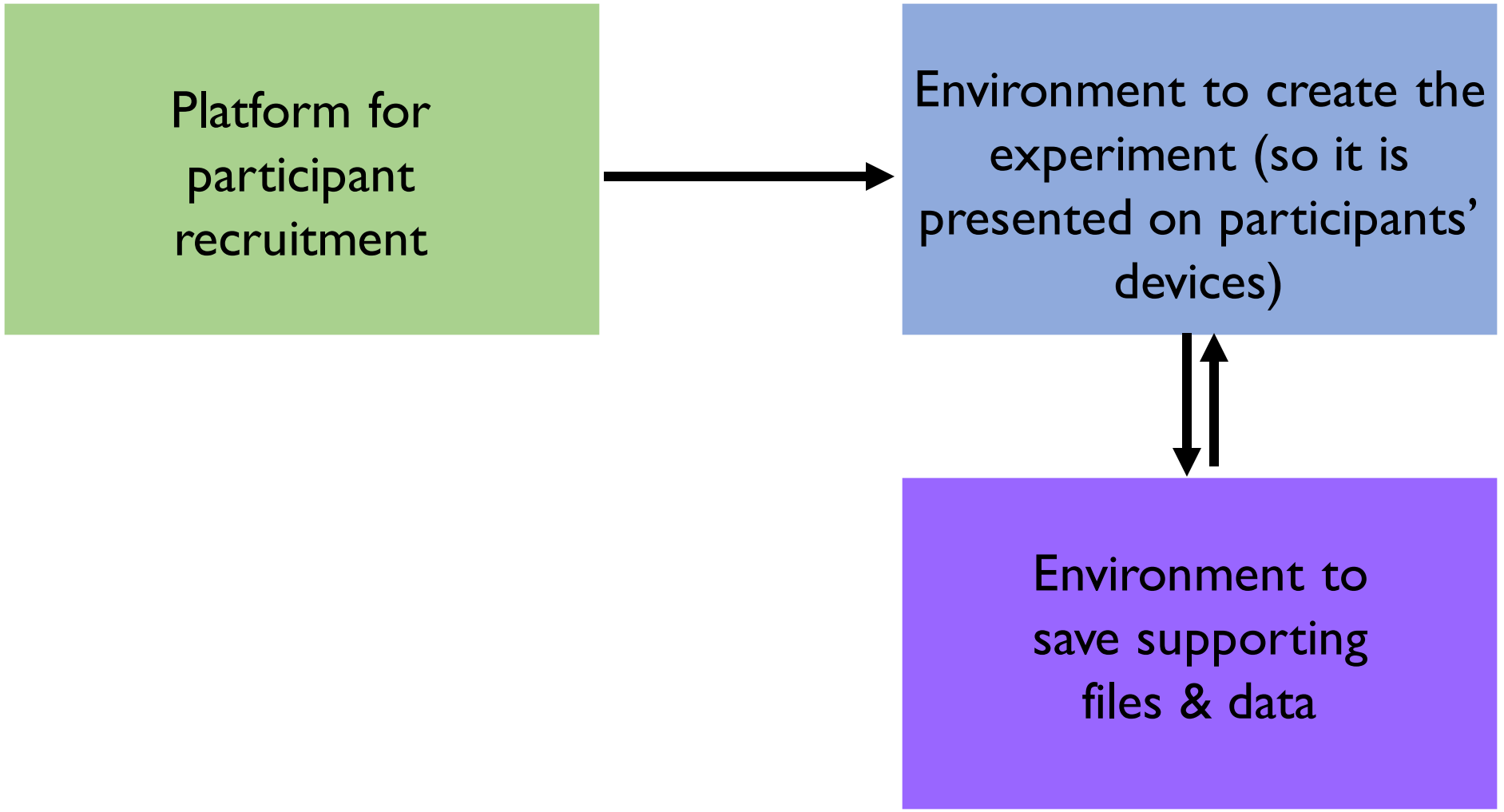
Why run experiments online?

- Pandemics!
- Faster
- Larger samples
- More diverse populations (less WEIRD)
- Access to specific populations

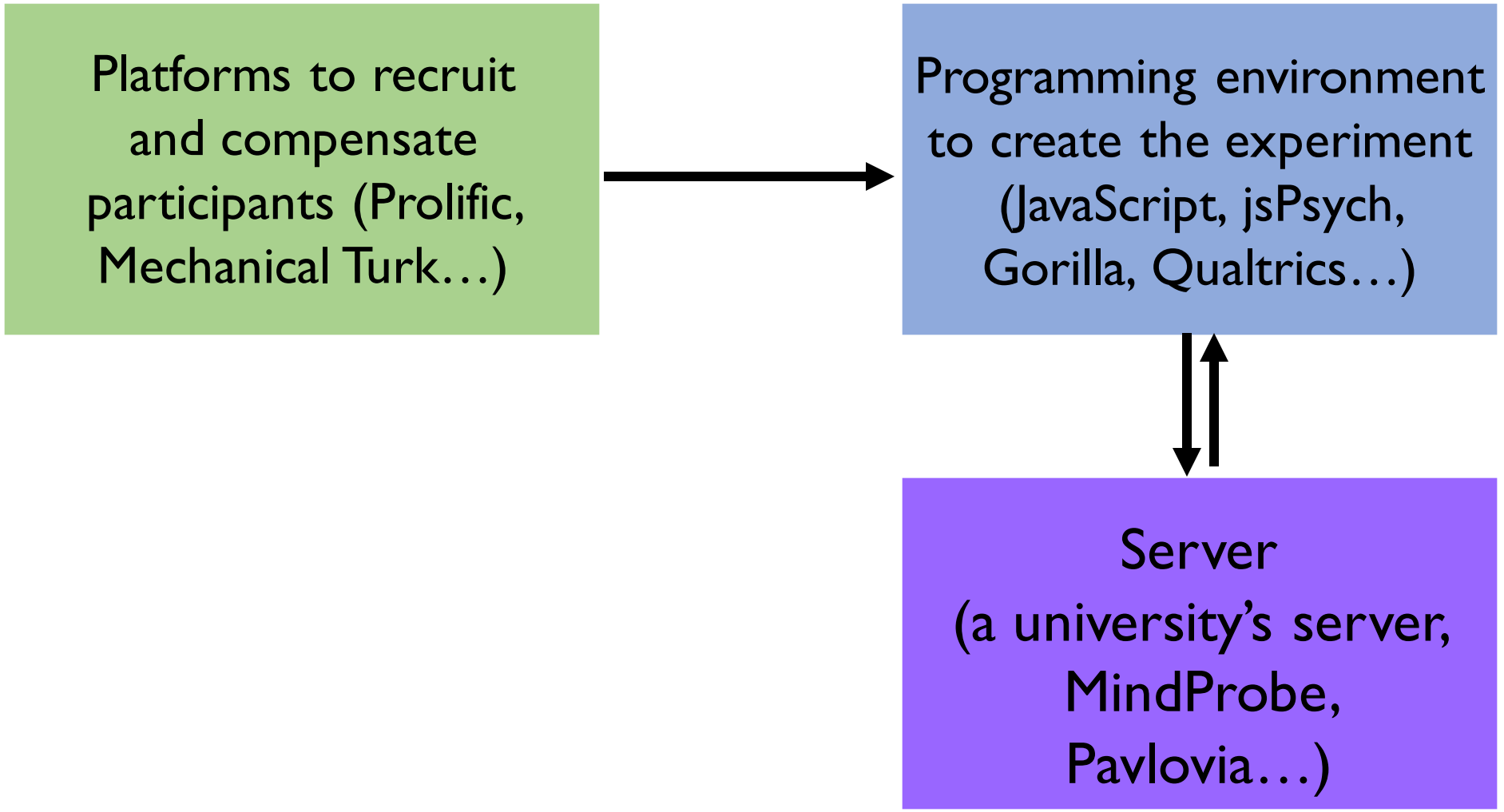
Running experiments online

- Participants are doing the experiment on their **own** devices
- We want them to access it on their browser (rather than asking them to download any particular software)
- We want the data to be saved **outside of their devices**

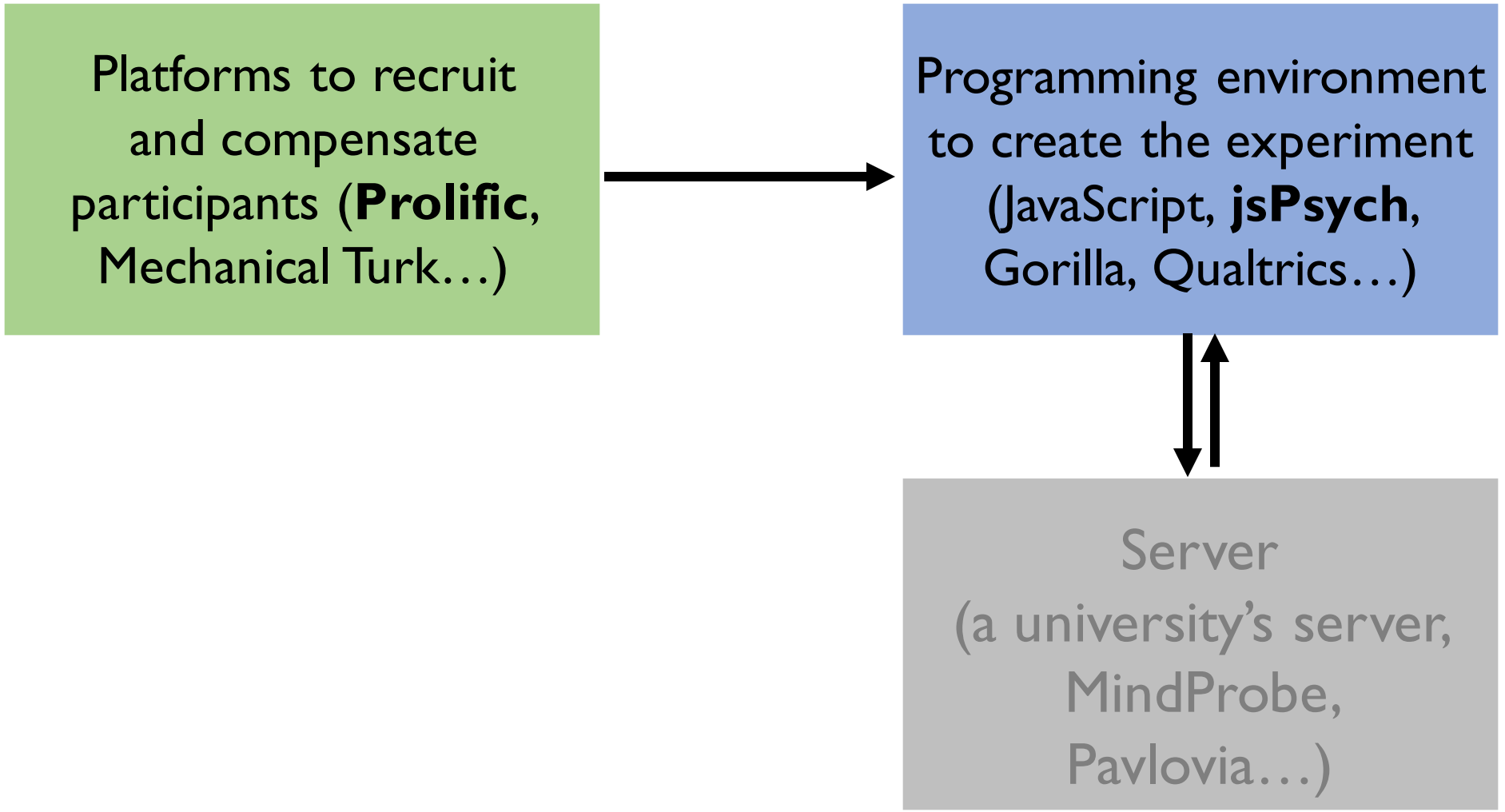
Running experiments online



Running experiments online



Running experiments online



Any questions?

Recruiting participants online

Crowdsourcing

- Once an experiment runs on a browser, it can be potentially sent to anyone with an internet connection
- Crowdsourcing sites

Designed for crowdsourcing anything

Designed for scientific data collection

A quick look at Prolific

Prolific

What do you want to do in Prolific?

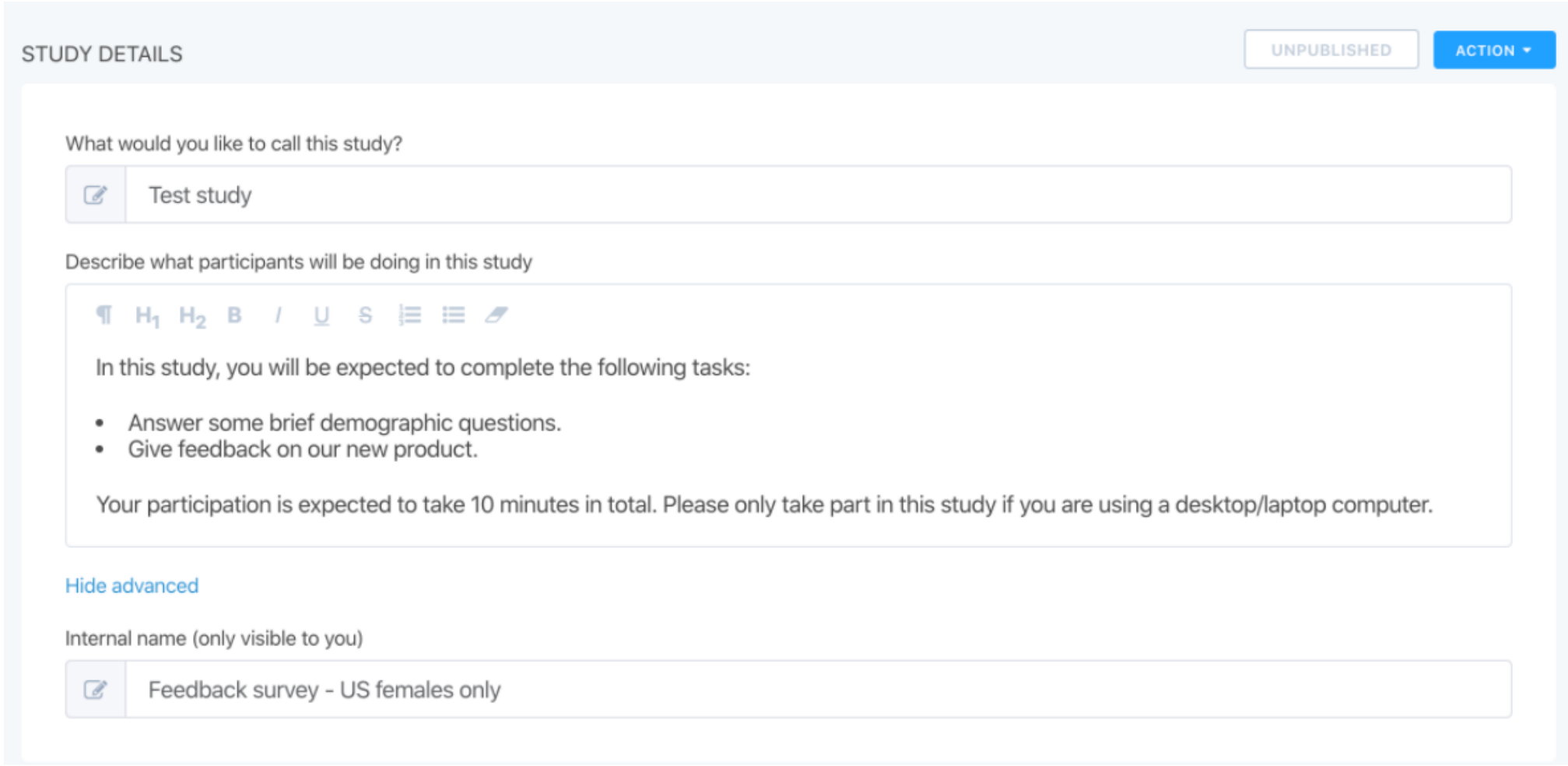
Take studies

Take part in engaging research, earn cash, and help improve human knowledge.

Run research

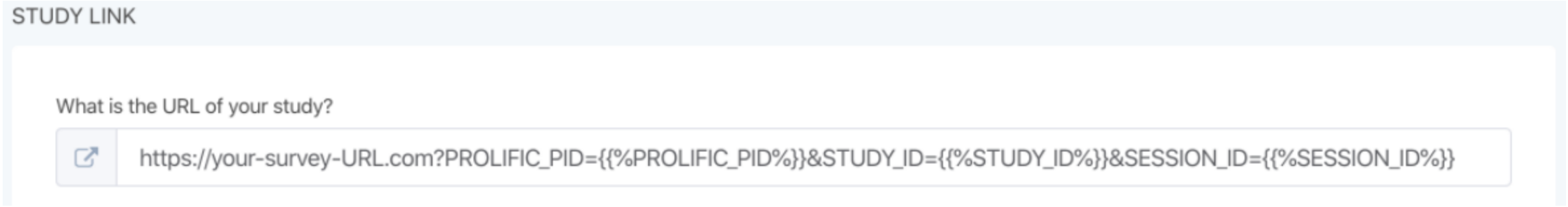
Recruit people around the world and collect high quality responses within minutes.

A quick look at Prolific



A quick look at Prolific

- Adding a link to the actual study

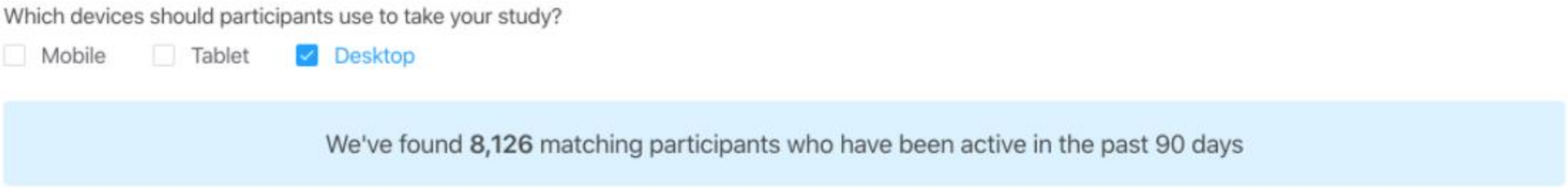


A quick look at Prolific

- Targeting a specific audience

A quick look at Prolific

- Targeting a specific audience



Data quality

- General aim: detecting unmotivated, unfocused or non-real participants
- Think about specific concerns, and design exclusion criteria to address them accordingly (Jenni Rodd, BeOnline2018; BeOnline2020)
  - Measure completion times
  - Repeat key questions in different ways
  - Language tests
  - Attention checks
  - Debriefing
  - Make random clicking annoying for participants
    - Make them repeat a trial during training when they get it wrong
    - Make pauses after wrong answers longer

Data quality

- Make the experiment as short and fun as possible
- **Pilot** before starting

Ethical practices

- Online participants ***should not*** be paid less than lab participants (+ note that Mturk/Prolific charge extra fee)
  - Mturk has no minimum pay rate
  - Prolific has a cheap minimum pay rate (£6/hour)
  - **Pay fairly**, match at least the National Minimum Wage
- Exclusion criteria: applied to data, not payment!
- Treat participants with respect
- **Pilot** before starting

Any questions?

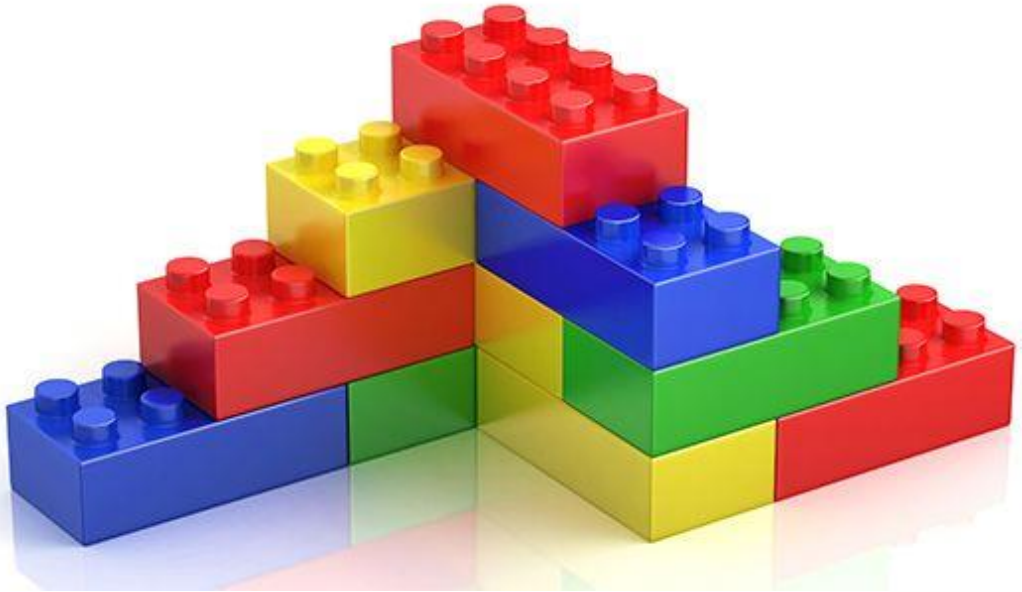
Getting your experiment on the screen: introduction to jsPysch

What is jsPsych

- JavaScript library
- Specifically designed for experiment building

Josh de Leeuw, Vassar College

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behavior Research Methods, 47(1), 1-12.



- JavaScript library
- Collection of plugins

- An experiment!

Quick notes about programming

- Languages: HTML, JavaScript and CSS
- HTML: provides basic structure of sites
- CSS: used to modify and control presentation of content
- JavaScript: controls the behaviour of elements on the site

- Naming conventions:
- PascalCasing, camelCasing or underscore_casing

Before the practical

- Do you all have a laptop?
- Pair up!
- Github repo: https://github.com/annieholtz/ALLWorkshop
- Do you all have a text editor?
- MyExperiments folder download/jsPsych download

Practical

When the right trial doesn't exist

- E.g. image and audio stimuli on the same trial
- Incorporate images in parameters that allow HTML content
- Edit the plugins/build your own plugin
  - https://www.jspsych.org/7.3/developers/plugin-development/

Other things to know

- jsPsych version control and updates
- Preloading stimuli (https://www.jspsych.org/7.3/plugins/preload/)
- Running experiments online (https://www.jspsych.org/7.3/overview/running-experiments/)
- Random assignment to conditions
- Conditional trials/timelines
- Audio and video recording:
  - https://kennysmithed.github.io/oels2021/
  - https://www.jspsych.org/7.3/plugins/html-audio-response/

Further resources

- Good documentation:
  - https://www.jspsych.org/7.3/
- Active community:
  - https://github.com/jspsych/jsPsych/discussions
- Extended courses, tutorials and workshops:
  - Online experiments course by Kenny Smith: https://kennysmithed.github.io/oels2021/
  - jsPsych tutorial for online experiments by Alisdair Tullo: https://softdev.ppls.ed.ac.uk/online_experiments/index.html
  - Extensive Edinburgh virtual workshop on artificial language learning: https://www.youtube.com/playlist?list=PLNRhI4Cc_QmsAnzLddCkCPHqdHtMg4TVO

Questions? Thoughts?

Contact us:

annie.holtz@ed.ac.uk

m.s.meisezahl@sms.ed.ac.uk

stal@ed.ac.uk