Integrative approaches in child language processing: examining word meaning disambiguation through language corpora, behavioural and computational experiments

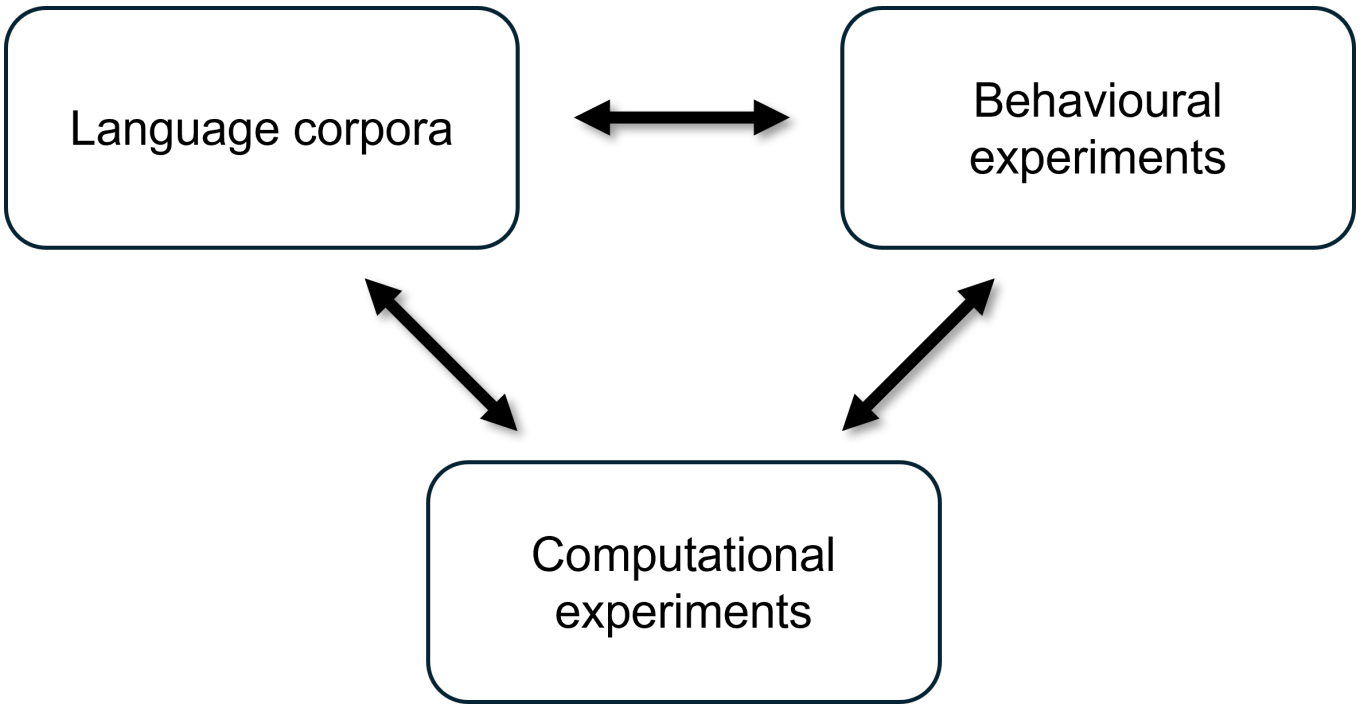
Francesco Cabiddu UCL
francesco.cabiddu@ucl.ac.uk

Supervisors: Chiara Gambi, Lewis Bott, Gary Jones

PSICOSTAT - 15/03/2024 | Presentation OSF

Combining corpus, behavioural, and computational evidence

- Understanding language in its natural habitat
- Testing processing mechanisms



Monaghan & Rowland, 2017

Child lexical disambiguation

- Lexical disambiguation depends on sentence parsing
- "*Elmo watched a movie with a castle and a princess. That was a funny* /**nʌɪt**/"
- Night or Knight?

Bottom-up account

Low-level cues: "castle" and "princess" frequently co-occur with **knight**

Informativity account

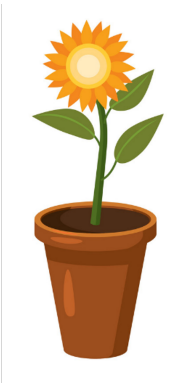
High-level cues: People typically watch movies at **night**

Rabagliati et al., 2013; Snedeker & Yuan, 2008; Trueswell & Gleitman, 2007

The role of verb-event structures

- Verbs facilitate young children's word and syntactic processing (Mani et al., 2016; Andreu et al., 2013):

Pushing the…



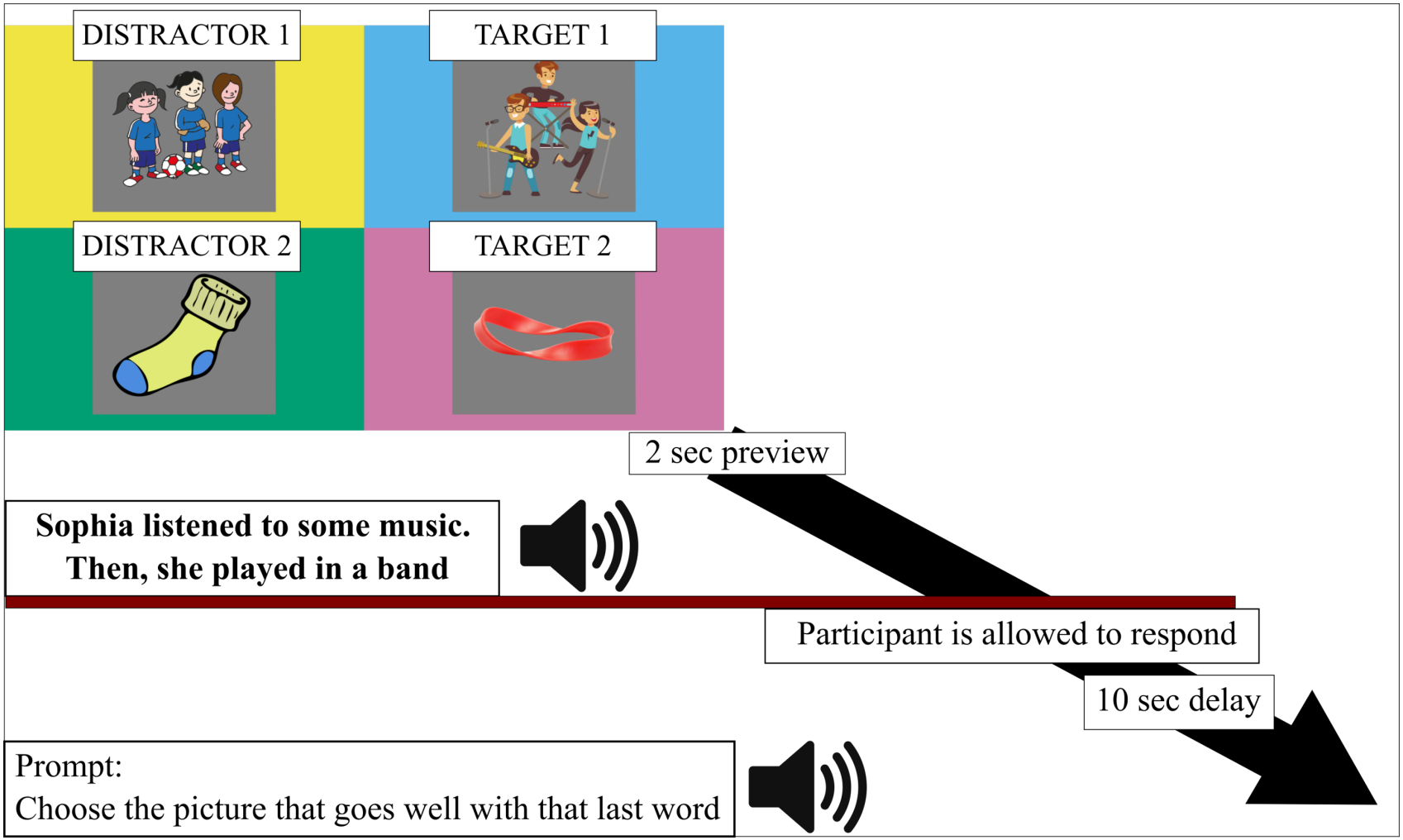
- In lexical disambiguation, not clear whether effects are driven by verb-object associations or verb-event generalisation: "*She **met** the **star***" (Hahn et al., 2015)
  - (bottom up) the child might have heard the expression "meeting a (famous) star" in the past
  - (top down) the child might infer that one can only meet an animate entity, not an inanimate astronomical object

Using a sense-annotated corpus to disantagle the effects of verb-object and verb-event cues

- *ChiSense-12* (Cabiddu et al., 2022): a new open-access sense-tagged corpus of English child-directed speech
- All English child-directed sentences up to child age 4 years, available in the CHILDES database (MacWhinney, 2000)
- 15,581 sentences. 12 ambiguous nouns used in dominant (e.g., elastic band) and subordinate sense (e.g., music band)
- The sentences were tagged for noun meaning and instances of verb-noun use

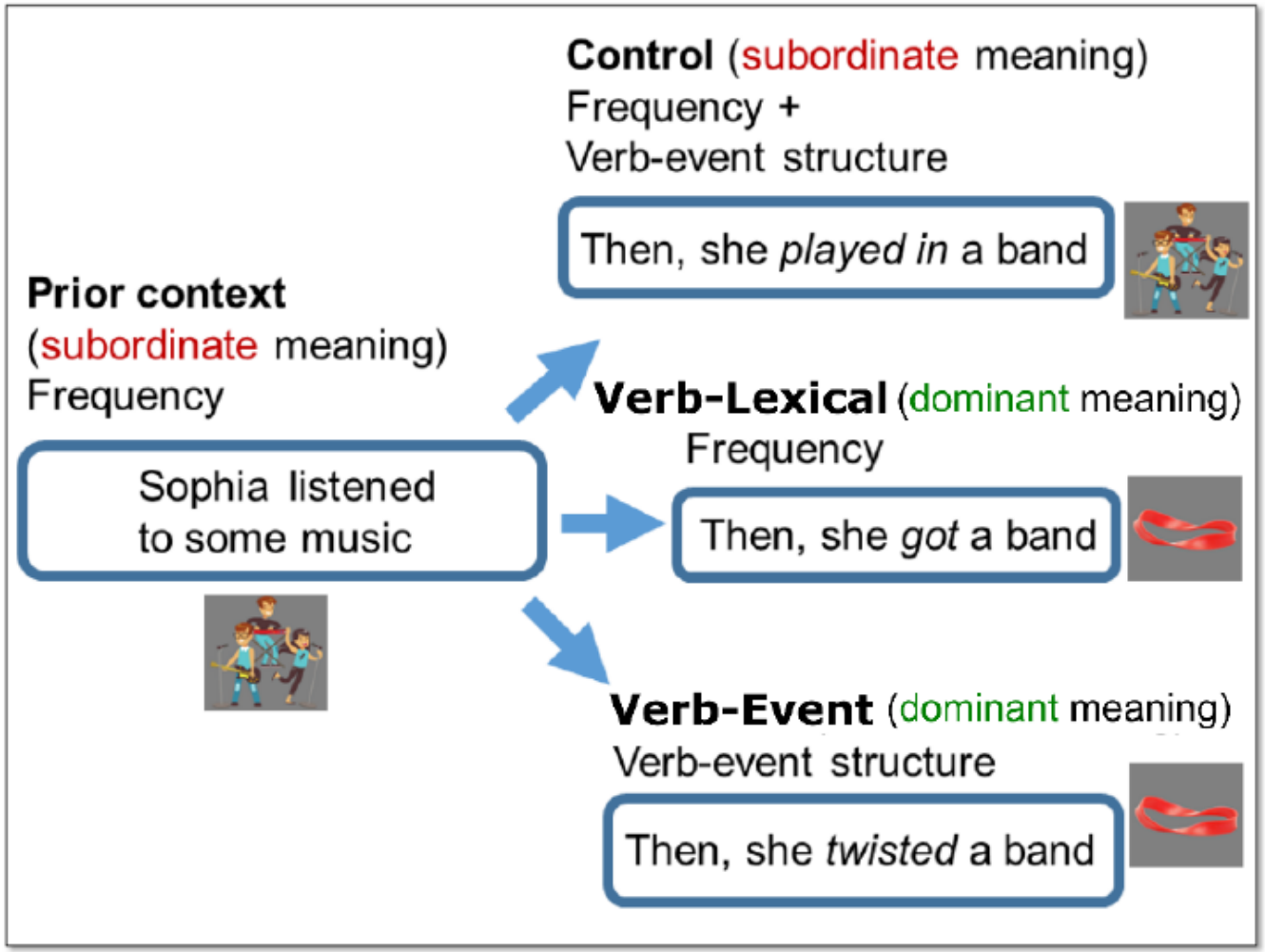
ID	Gloss	Target	Sense	Verb
311504	who put the rubber band on there	band	object	put on
326153	are you in a marching band	band	music group	be in
326190	oh a clown's in the band	band	music group	be in
326293	remember Child when did we see a band	band	music group	see

Experimental set up



Cabiddu et al., 2022; 2024

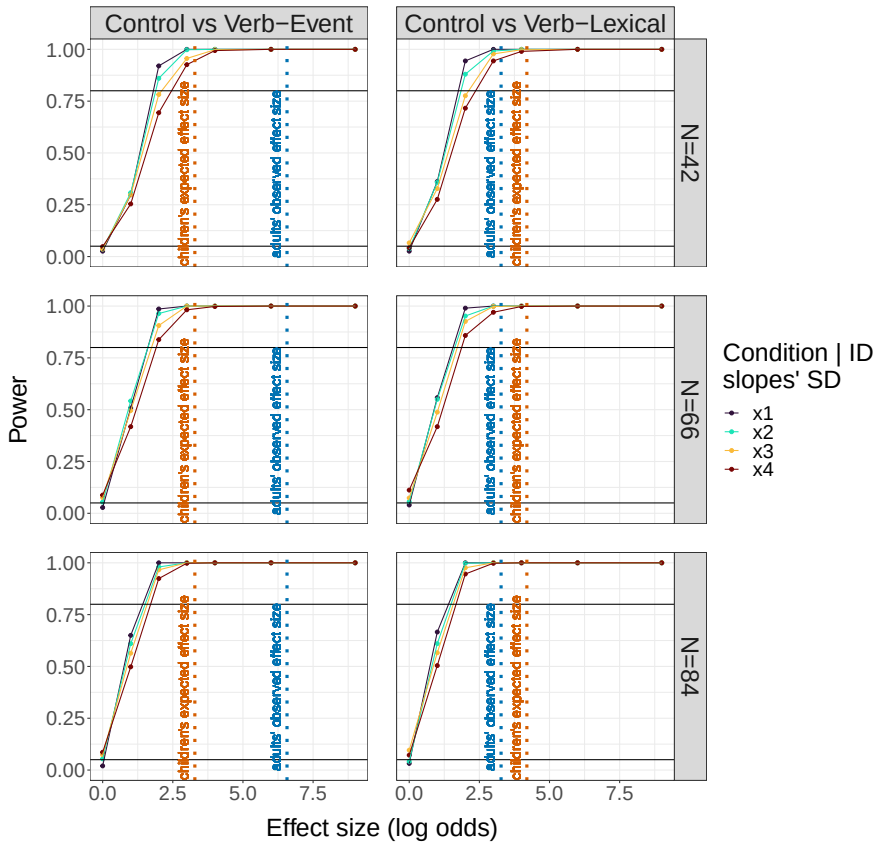
Experimental conditions



Cabiddu et al., 2022; 2024

Data collection and power estimation

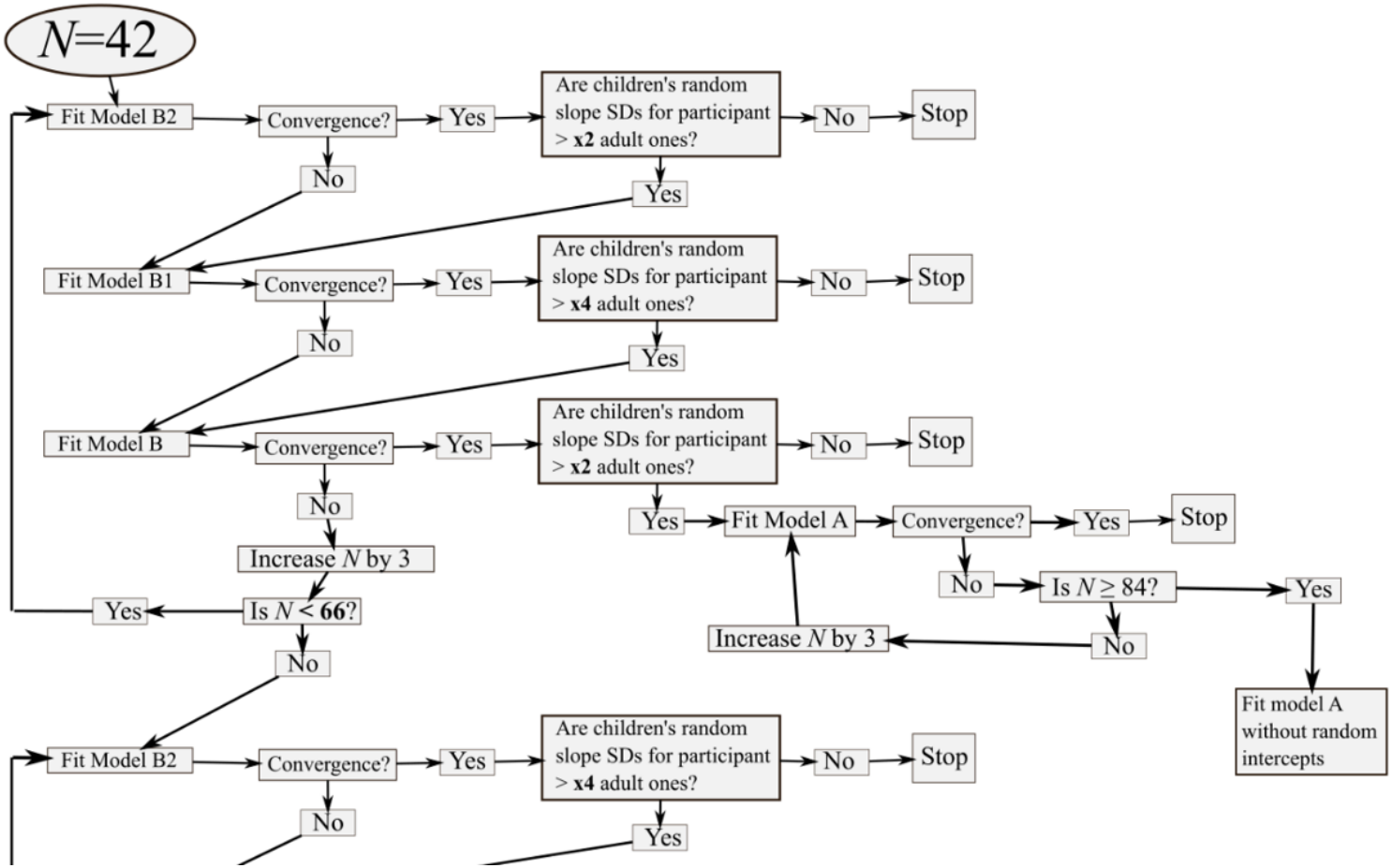
- 83 English-speaking adults recruited from Prolific (*Mean Age* = 23 years, *SD* = 4.5 years; 55 females).
- Pre-registration of hypotheses, power simulations and related data collection stopping rule for child data collection



Cabiddu et al., 2021

- Adult pilot data to estimate power at *N* = 42, 66, and 84
- Power as a function of sample size, effect size, and SD of Condition | ID random slopes
- 4 different models of increasing complexity: + random slopes, + correlations between intercepts and slopes

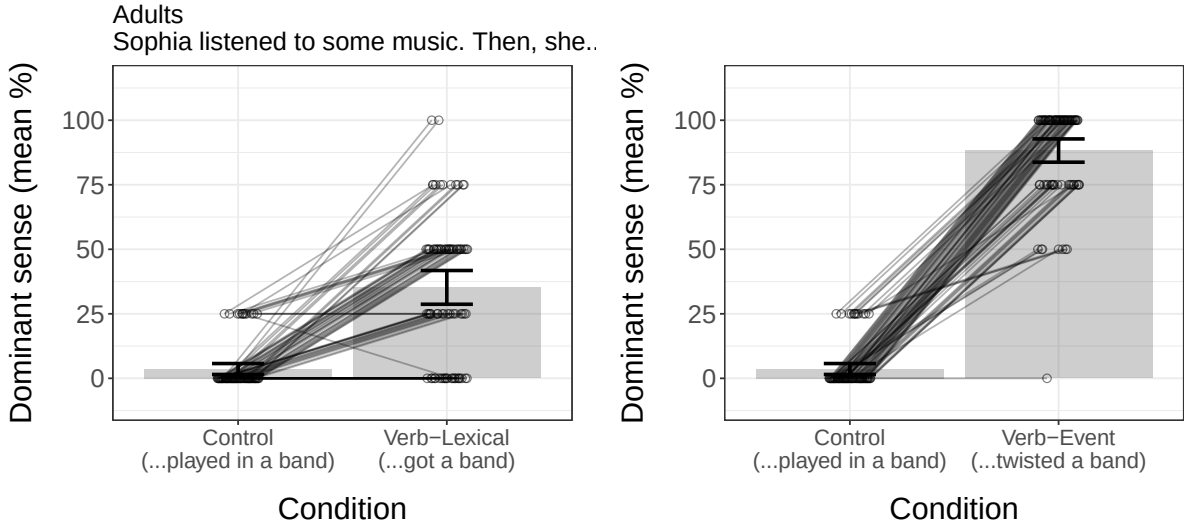
Data collection and power estimation

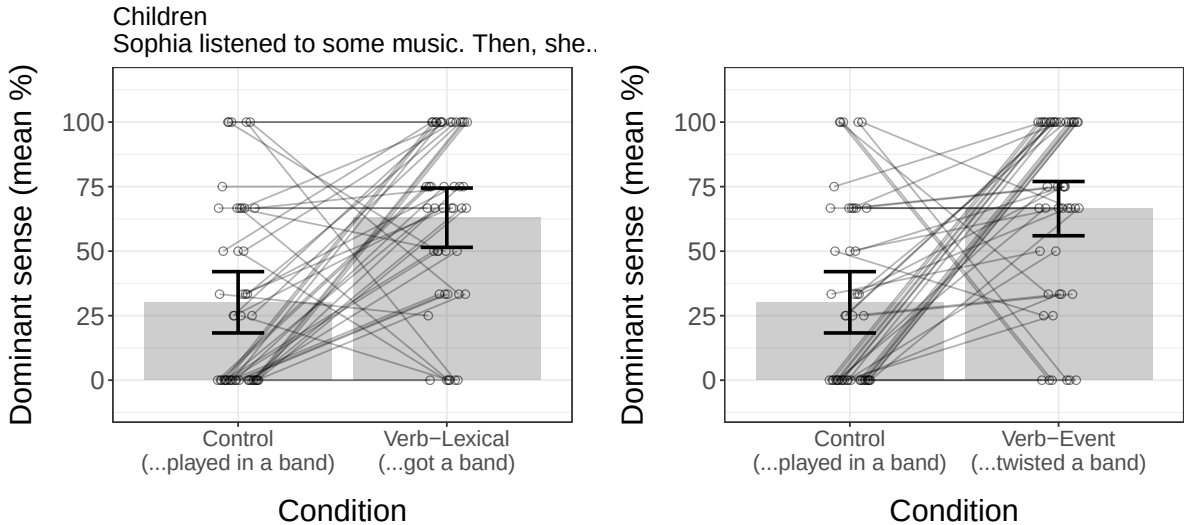


- Final child sample of 45 children aged between 48 to 59 months (*Mean Age* = 52 months, *SD* = 3 months).

Cabiddu et al., 2021

Results





Cabiddu et al., 2022; 2024

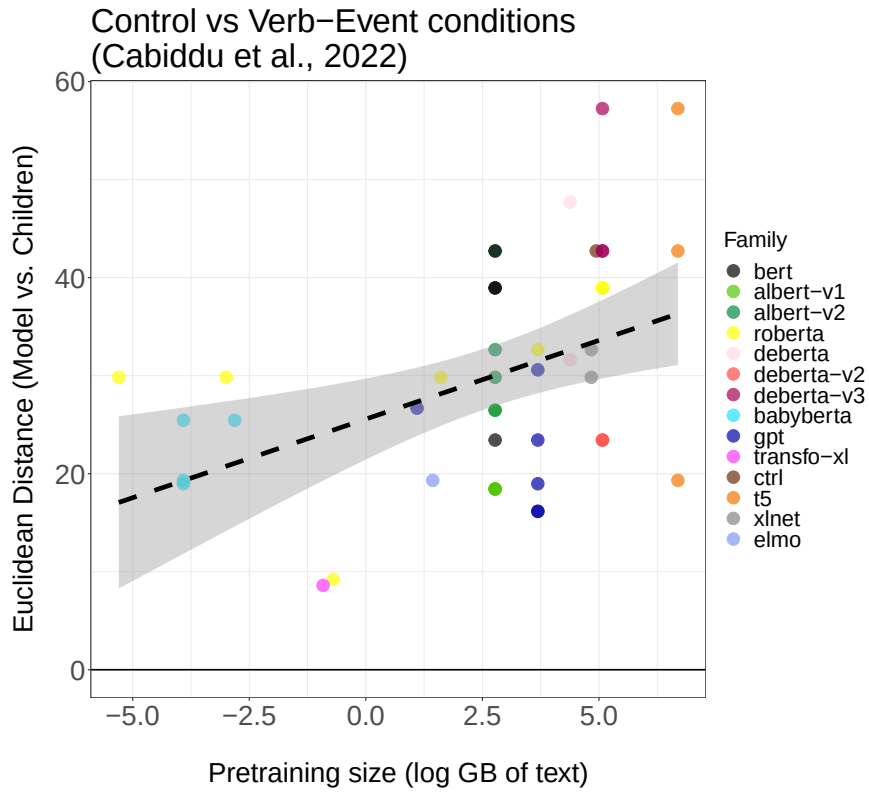
Discussion

- First evidence that 4-year-old children primary rely on top-down cues (verb-event structures) in a lexical disambiguation task, supporting an informativity account to sentence parsing (Trueswell & Gleitman, 2007)
- Child word representations are contextualised (Srinivasan & Rabagliati, 2021)
- Advantages (and disadvantages) of using naturalistic language corpora to disentangle the effects of bottom-up associations and top-down structural cues

Next step Cabiddu et al., 2023: Simulating learning mechanisms and the role of language input

- What learning mechanisms might capture performance in different tasks and age groups?
  - Usage-based learning via domain-general associative and analogy mechanisms (e.g., Alishahi & Stevenson, 2007; Bybee, 2010)

Word prediction via self-attention (e.g., Vaswani et al., 2017; Devlin et al., 2019; Radford et al., 2019):
- Contexualised word representations based on child or adult-directed speech
- Sensitive to word associations and sentence structure
- Large pool of models that can be studied as participants



Thank you!

Francesco Cabiddu

francesco.cabiddu@ucl.ac.uk

Presentation OSF