Statistical learning and language multimodality

Arianna Colombani

University of Potsdam, Macquarie University

November 14, 2024

First Language Acquisition – IECL Master

University of Potsdam

Introduction

- PhD candidate in **Linguistics and Cognitive Sciences** (supervised by Prof. Tuomainen and Prof. Boll-Avetisyan)
- I investigate **gesture learning** from infancy to adulthood using **brain, eye-tracking** and **behavioural** measures
- **Psycholinguist** and **Italian Sign Language** interpreter, and educator for the deaf

More info:
https://clmrnn.github.io

Topics

Part 1: Statistical learning

- Introduction to SL
- Cross-situational learning paradigm

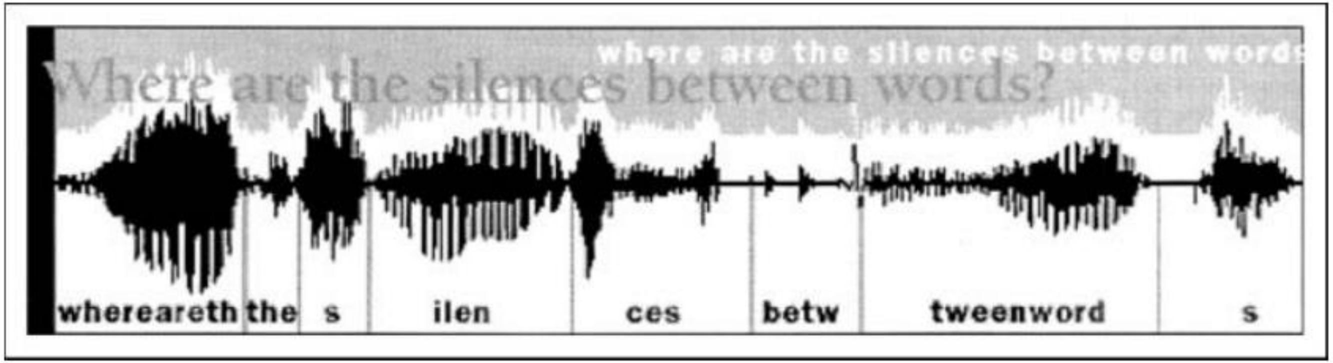
Part 2: Language multimodality

- Introduction to sign-like gestures
- Results of «Cross-situational learning of sign-like gestures» experiment

Statistical learning

Statistical learning

How do language learners recognize words in a stream of fluent speech? **(segmentation)**



(Aslin et al., 1998; Saffran et al., 1996; Saffran, 2003).

Statistical learning

Statistical learning (SL) → ability to detect the **statistical regularities** in the environment that is perceived by our senses, **without intention** or conscious awareness.

SL can be applied as a mechanism to dected word boundaries (and many other language features)

e.g. soundstream: *prettybaby* → *pret-ty-ba-by* → *pretty* is a word, *tyba* is not

pretty ✓
tyba ✗

the probability that *ty* will follow after *pre* is as high as 80%
the probability of *ba* given *ty* is less than 1%, **indicating a word boundary**

Seminal work by **Aslin, Saffran, and Newport (1996)**
infants can segment artificial continuous speech into its constituent syllabic sequences based on the transitional probabilities between the syllables

(Aslin et al., 1998; Saffran et al., 1996).

Saffran, Aslin, and Newport (1996)

Participants: 24 infants (8-month old)

Stimuli: **4 pseudowords x 2 minutes** of continuous speech stream: no pauses, stress differences, or any other acoustic or prosodic cues to word boundaries.

Transitional probability (TP): the likelihood that one syllable follows another in a sequence.

bidakupadotigolabubidaku → *bi-da-ku-pa-do-ti-go-la-bu-ti-bu-do…*

(repeated in random order)

$$Y|X = \frac{\text{frequency of } XY}{\text{frequency of } X}$$

$$P(da|bi) = \frac{\text{of bi followed by da}}{\text{of bi}} = 1.$$

$$P(ku|da) = \frac{\text{of da followed by ku}}{\text{of da}} = 1.$$

$$P(pa|ku) = \frac{\text{of ku followed by pa}}{\text{of ku}} = \frac{3}{10} = 0.3$$

Within-word TP = 1 (100%) **Between-word TP = 0.3 (33%)**

Saffran, Aslin, and Newport (1996)

bidakupadotigolabubidaku → *bi-da-ku-pa-do-ti-go-la-bu-ti-bu-do…* → *kupago* ✗ *padoti* ✓

(repeated in random order)

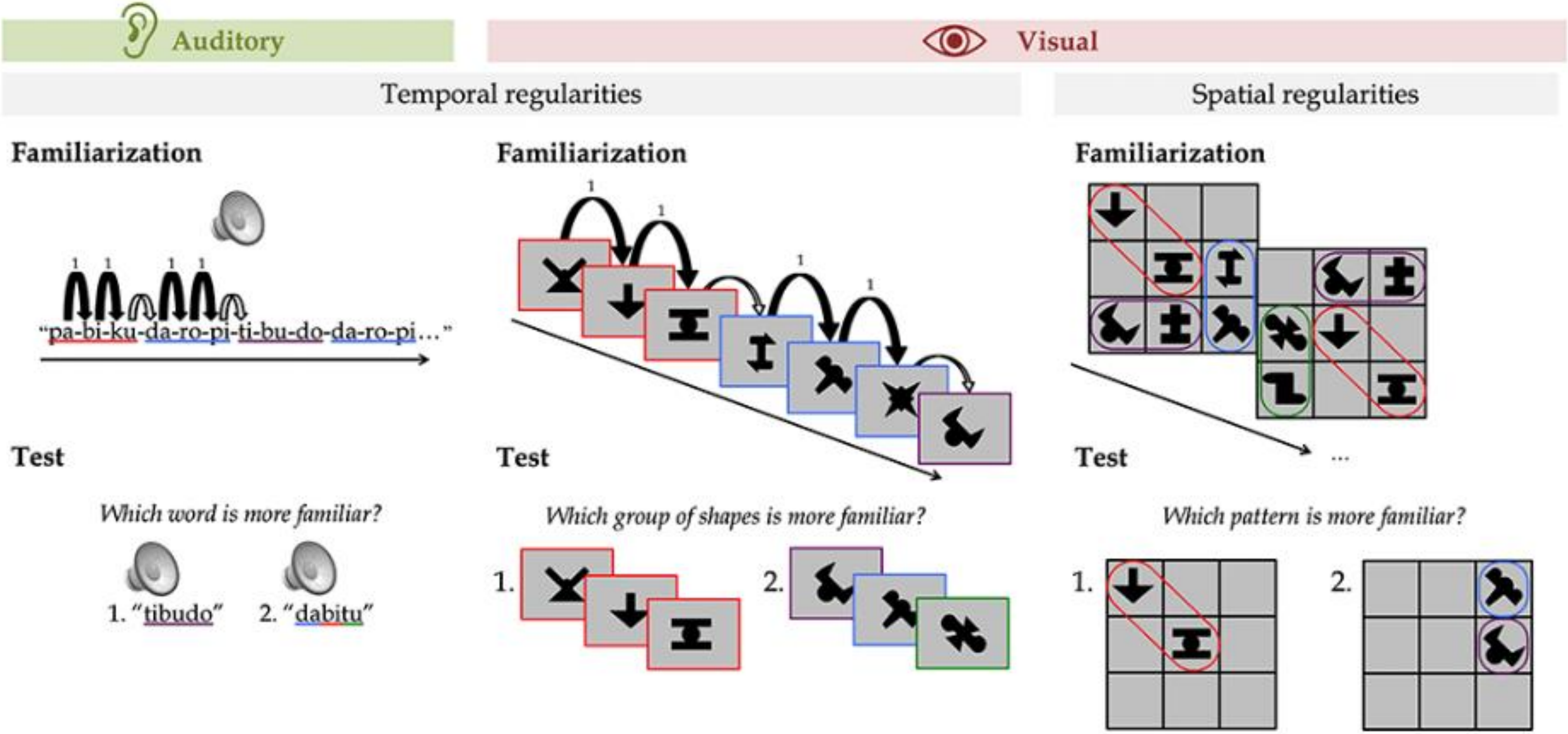
Measure: Head turn preference (novelty prefence: novel item = longer listening times)

Results: 8-month-old infants can **segment artificial continuous speech** into its constituent syllabic sequences based on the **transitional probabilities** between the syllables.

Infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also the acquisition of other aspects of language (phonolgy, morphology, syntax, word learning)

(Saffran 2003).

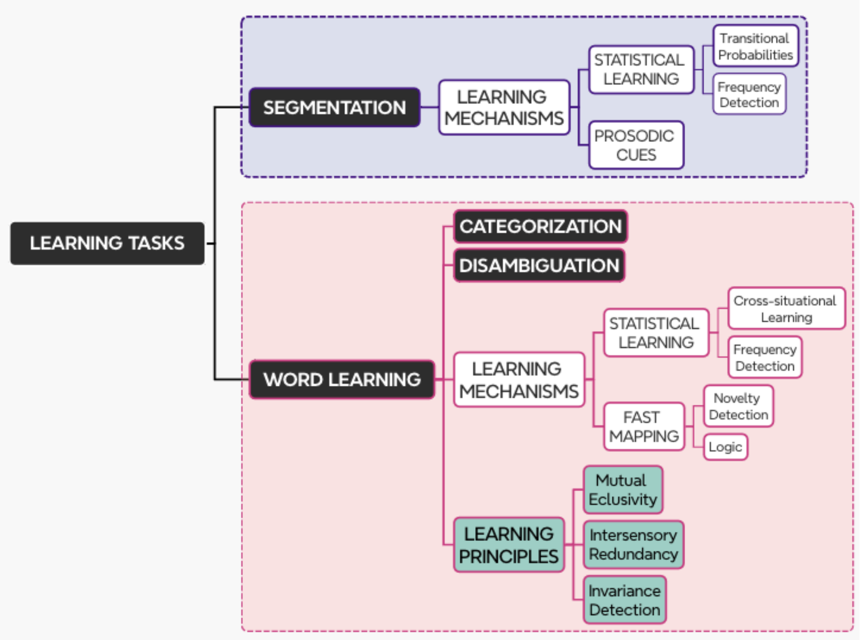
Statistical learning



(Review: Brooks & Kempe, 2014).

Language learning challenges

Segmentation is far from being the only challenge of language learners.



Examples of learning mechanisms and principles that we can test as researchers if we want to investigate language learning.

My focus:

- **Multimodal integration** (auditory and visual modality)
- Word-referent mapping = **word learning**

Cross-situational statistical learning

Word-referent mapping

How could a learner who understand how to **map** (associate) a names onto its correct **referent in the environment**? (*word-to-world mapping* problem)

- Social cues (e.g. gaze)
- Attention direction (e.g. pointing)
- **Statistical learning** → statistics are calculated across different learning instances to determine the most likely word-referent mappings across multiple experiences

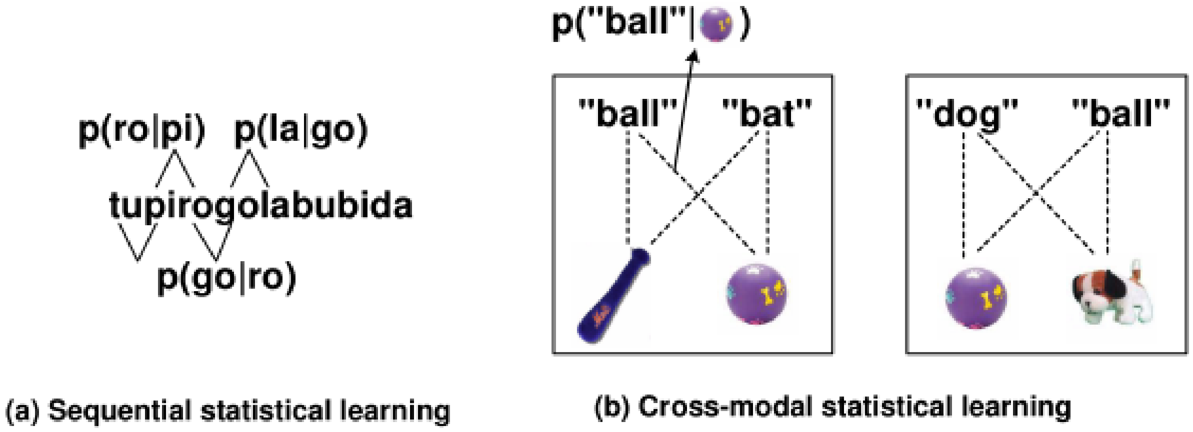
↓

In the lab: Cross-situational learning paradigm

(Gleitman, 1990)

Word-referent mapping

Sequential statistics	Cross-modal statistics
Conditional probabilities of **adjacent elements** from the same stream of repeating elements.	Conditional probabilities of **co-occurring elements** in a trial from two streams of data.



(Smith & Yu, 2008b)

Cross-situational learning paradigm

«Cross-situational learning paradigms simulate ***language learning in the real world*** *by presenting participants with* ***words or phrases*** *that can potentially map onto multiple* ***competing objects or scenes*** *in the environment (e.g., an infant hears the word "ball," which can map onto numerous toys in their field of vision).*

Over time, learners can capitalize on these ***co-occurrences*** *to acquire words, phrases, and their meanings , just as they do in natural settings (e.g., the infant consistently hears the word "ball" in relation to small round objects and eventually surmises that such objects are the referents for this word)»*

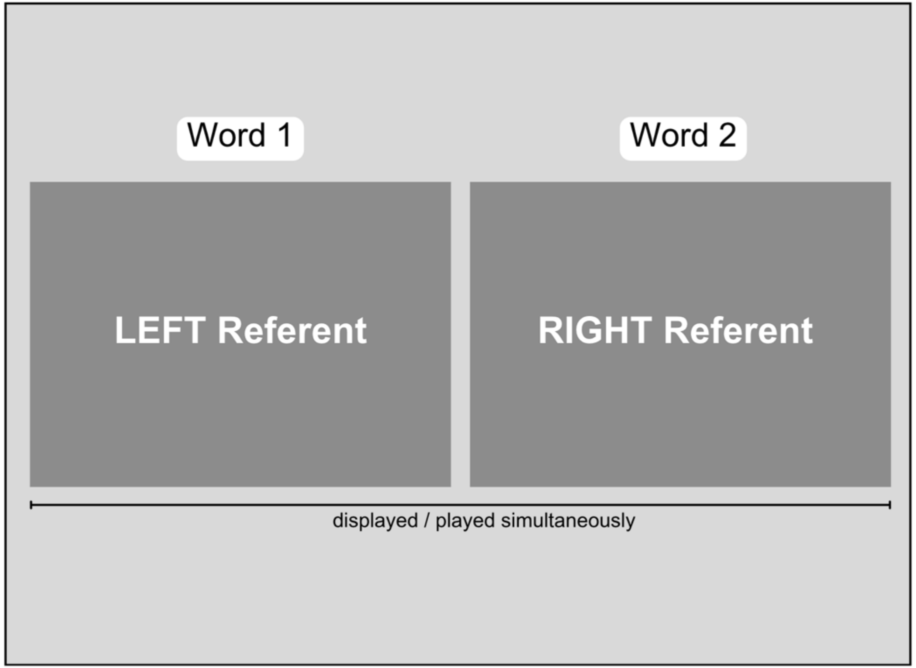
(Isbilen & Christiansen, 2022:4)

(Isbilen & Christiansen, 2022)

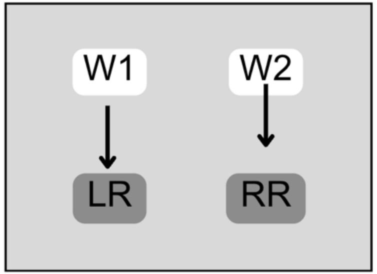
Cross-situational learning paradigm

- ambiguous learning trials
- multiple words and referents
- no explicit indication of word-referent correspondences

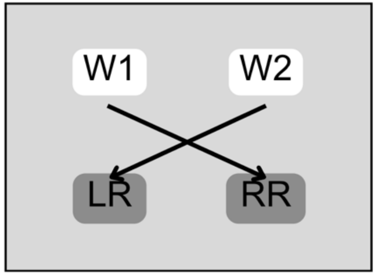
ability to identify the correct association by **implicitly detecting co-occurrences** across the trials



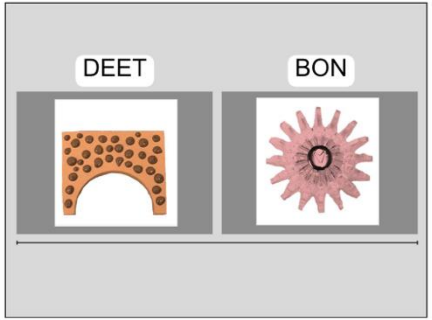
Aligned trial



Crossed trial



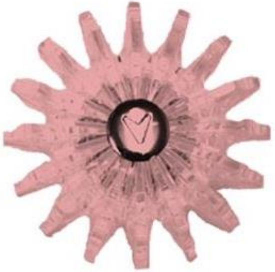
Cross-situational learning paradigm



?

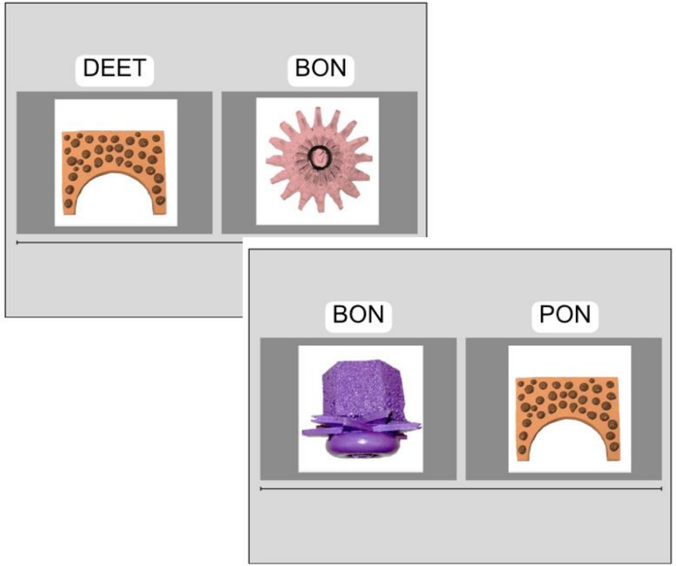
?

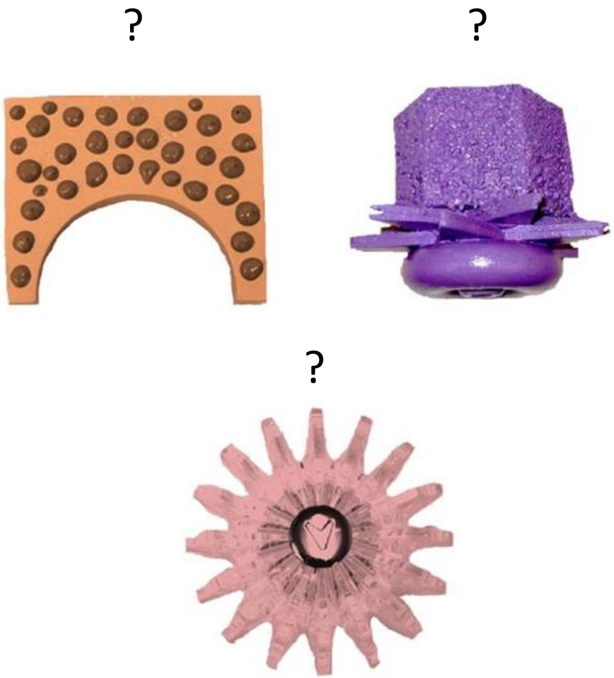
?



(Stimuli from Escudero, Mulach, Vlach, 2016)

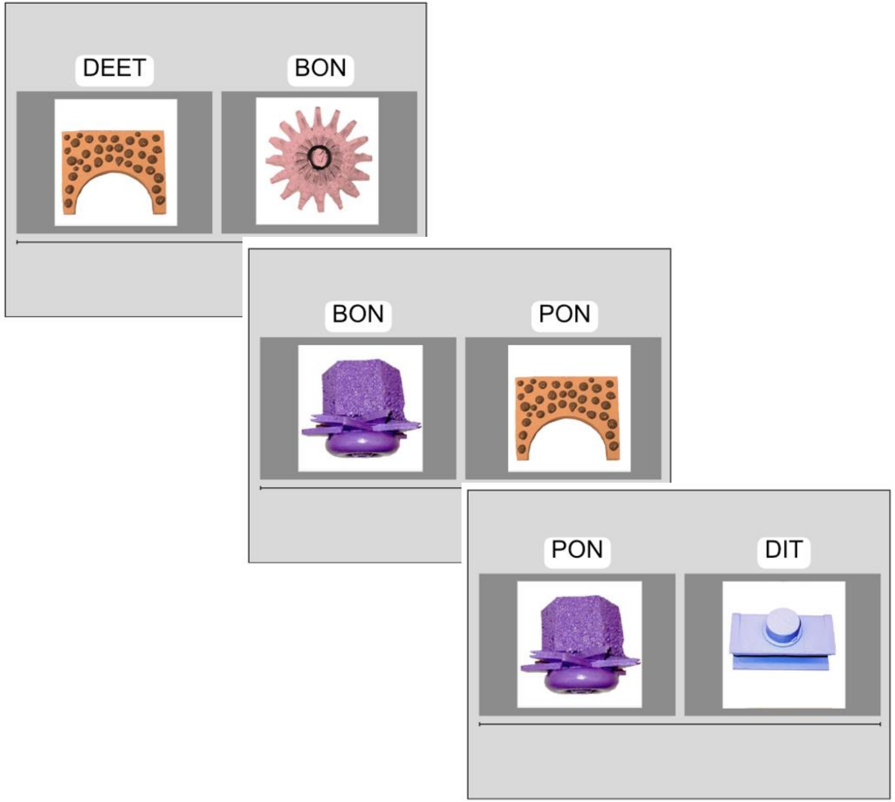
Cross-situational learning paradigm

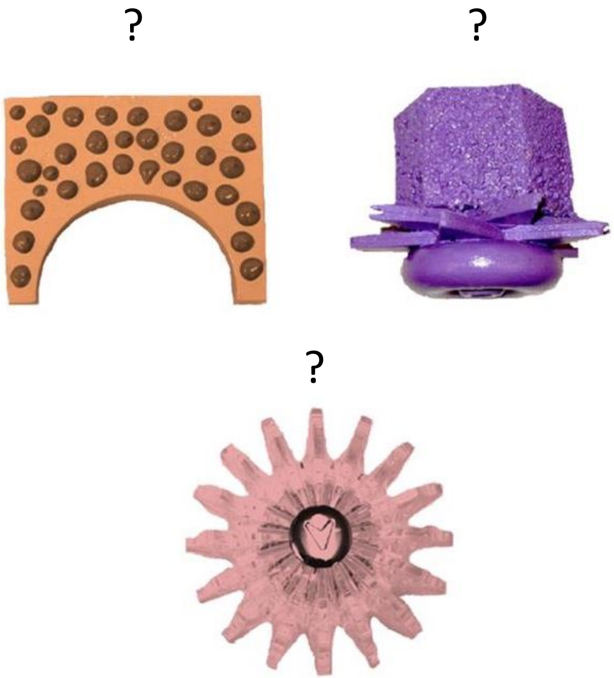




(Stimuli from Escudero, Mulach, Vlach, 2016)

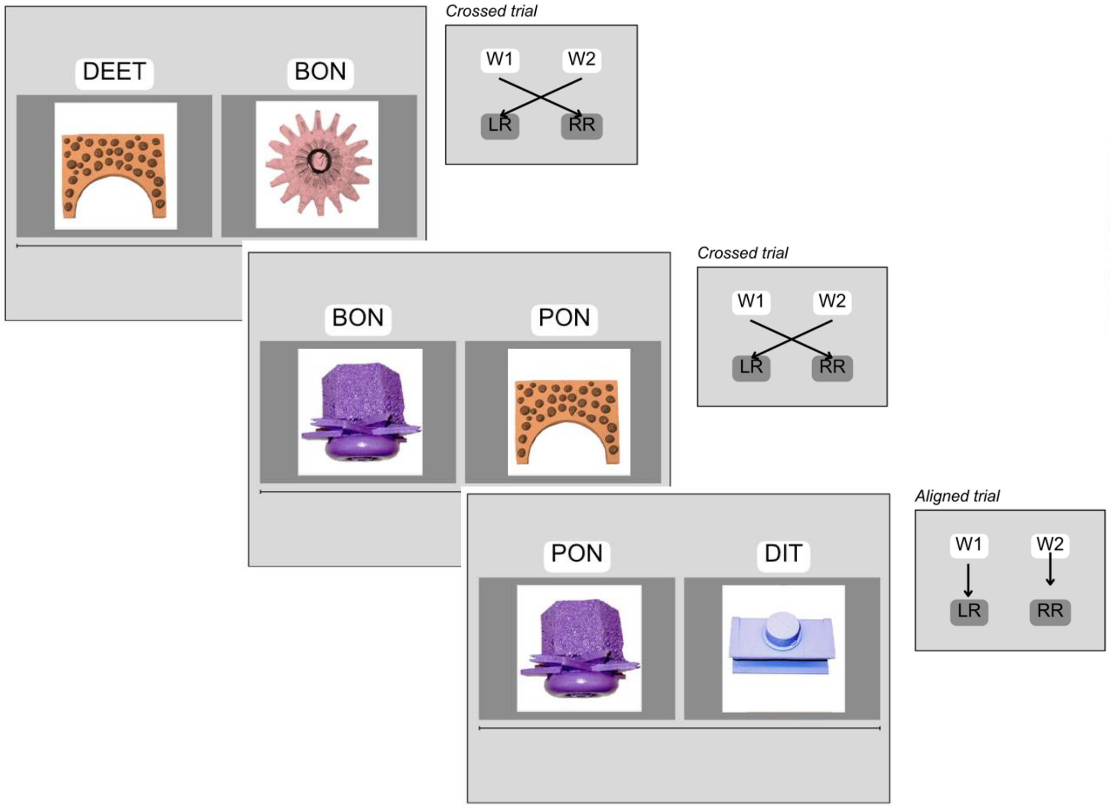
Cross-situational learning paradigm

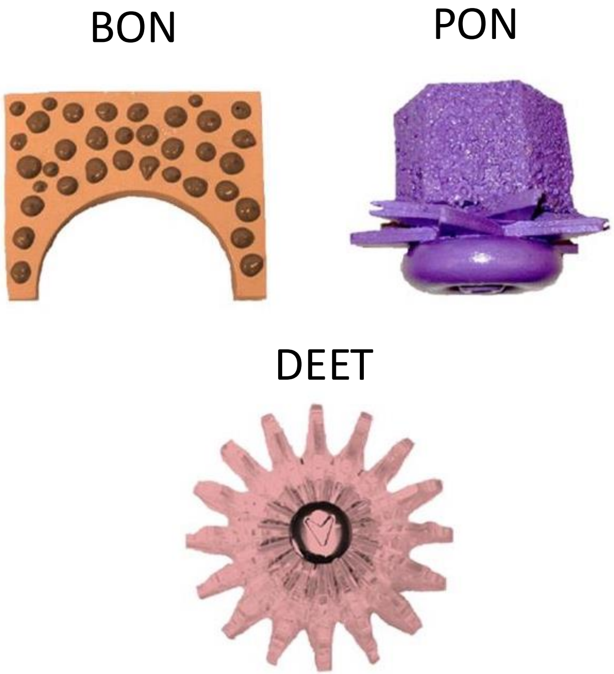




(Stimuli from Escudero, Mulach, Vlach, 2016)

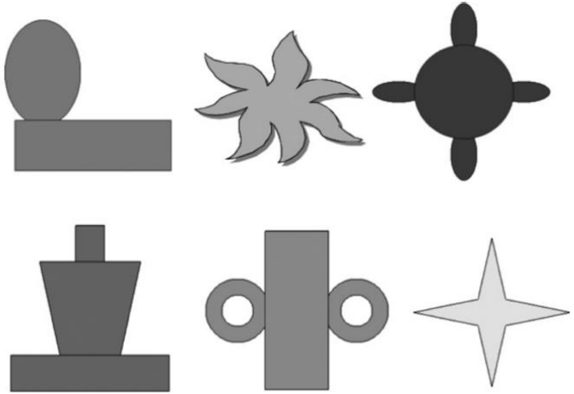
Cross-situational learning paradigm





(Stimuli from Escudero, Mulach, Vlach, 2016)

Smith & Yu (2008)



Participants: 28 infants (12 m.o) + 27 infants (14 m.o)

Stimuli: 6 pseudowords and 6 novel shapes
bosa, gasser, manu, colat, kaki and *regli*

Procedure: Familarization phase (6 x 10 times)
Preferencial-looking (correct referent = longer looking times)

Results: 12–14-month-old infants can infants rapidly learn multiple word-referent pairs across multiple and ambiguous scenes via statistical learning.

These findings suggest that learning the statistical regularities of the environment may be a critical part of the cognitive system by which children make sense of the world.

Language multimodality

Multimodality

Multimodality of language

Gesture-speech continuum

– “developmental”
– “linguistic”

- Gestures and words share a common neural system
- They develop on a gesture-speech continuum

(Xu et al., 2009; Fabbri-Destro et al., 2015; Bates et al., 1979; Capirci & Volterra, 2008;)

Gesture-speech continua

CO-EVOLVE

Before spoken language

DEICTIC GESTURES
- pointing
  - imperative
  - declarative

SYMBOLIC GESTURES
- symbolic play *(pretend to use the phone)*
- social routine *(BYE-BYE,...)*

After spoken language

GESTURE + WORD COMBINATIONS
- complementary *(pointing at doll + “doll”)*
- supplementary *(point at doll + “sleep”)*

SPOKEN WORDS

degree of gesture-word ratio

CO-EXIST

Context-dependent

CO-SPEECH GESTURES
- gesticulation

DEICTIC GESTURES
- pointing

Context-indipendent

SYMBOLIC GESTURES
- cultural based *(THUMBS UP)*
- iconically motivated *(CAR)*

Natural language

SIGN LANGUAGES

degree of ‘linguisticity’

(Adapted from Kendon, 1988; Bates et al. 1979)

Signs vs. gestures

VEASYT Tour Guida accessibile
YouTube link

Link to
Video 2

Link to
Video 3

Signs vs. gestures

SIGN LANGUAGE SIGNS

- language system
- related to one another to convey **complex meaning**
- mostly produced **without speech**

CONTEXT-INDIPENDENT GESTURES

- rely on vocal language
- produced in isolation to convey a **specific meaning**
- usually associated **with** speech

Signs vs. gestures

SIGN LANGUAGE SIGNS

- language system
- related to one another to convey **complex meaning**
- mostly produced **without speech**

SIGN-LIKE GESTURES

- arbitrary
- standardized

- non-iconic
- non-cultural based

- in isolation
- with speech

CONTEXT-INDIPENDENT GESTURES

- rely on vocal language
- produced in isolation to convey a **specific meaning**
- usually associated **with** speech

Sign languages vs. exact signing

American Sign Language

e.g., My house look-like what?

Signing Exact English

e.g., What does my house look-like?

- signs + speech
- English syntax
- **each word → one sign**

VISUAL SUPPORT

(ASL vs. PSE vs. SEE - House Description – YouTube link)

Sign languages vs. baby signing

[Link to video](#)

e.g. You ate the YOUGURT!

[Link to video](#)

e.g., MILK – SLEEP

- signs + speech
- **Target word = one sign**

VISUAL SALIENCE

(Videos courtesy of Baby Signs Italia©)

Literature gap

- Scarce psycholinguistic evidence
- Sign language literature cannot fully explain sign-like gestures, due to key differences

⇩

Preliminary investigation of the mechanisms underlying gesture + speech communication.

Cross-situational learning of word-pseudosign pairs in children and adults: a behavioral and event-related potential study

Research aim

Using…

- Statistical learning (cross-situational learning)
- Sign-like gestures

We aim to understand…

- the feasibility of learning **sign-like gestures as lexical labels for referents** in the context of bimodal communication (speech + gestures)

To ultimately…

- support the validity of bimodal communication (speech + gestures)

Research questions

- Is it possible to associate familiar spoken words with **novel sign-like gestures**?
- Is this possible to do this association rapidly through **statistical learning**

- Is it possible to **build semantic categories** of novel sign-like gestures?
- In case of category violation, do sign-like gestures **elicit (electro)physiological responses similar** to spoken words/sign language signs (i.e **language input)**?

Methods

Participants

- Children (8–11 y.o) N = 24
- Adults (18–35 y.o)   N = 19

Stimuli: 8 word-gesture pairs

- 8 words (8 semantic categories) matched with 8 novel symbolic gestures

Measures:

- Behavioral tasks (yes/no task)
- Electrophysiological responses (ERP - N400)

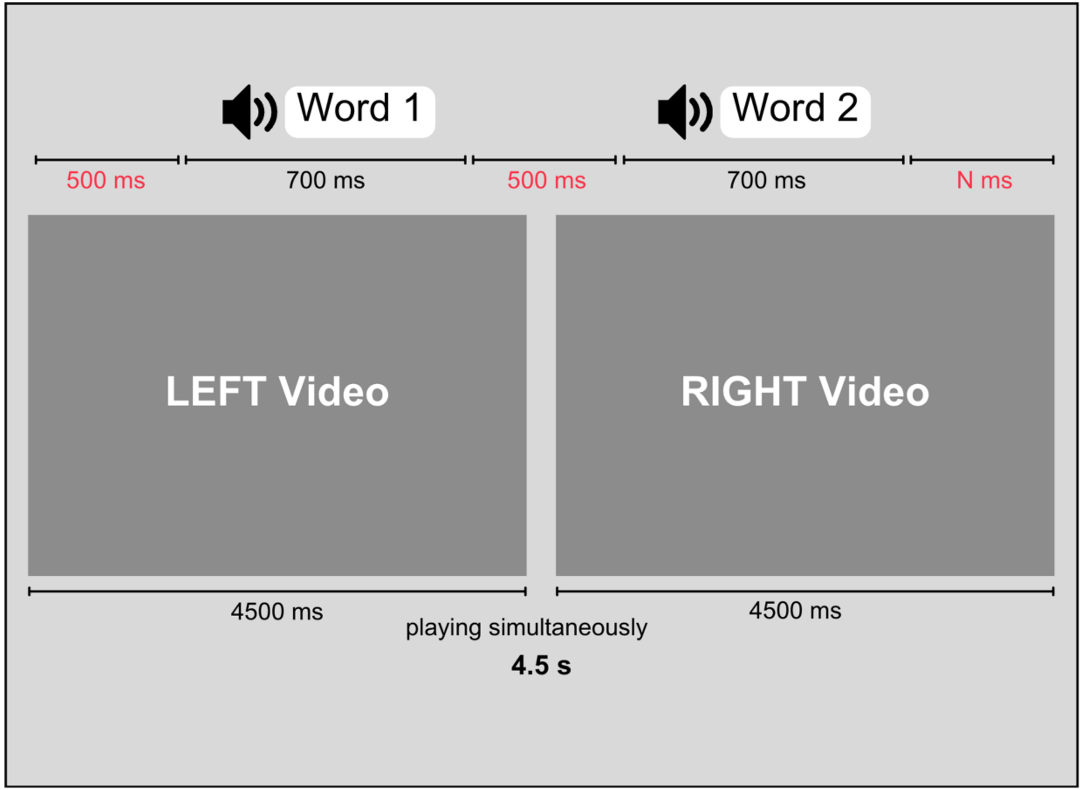
Paradigm:

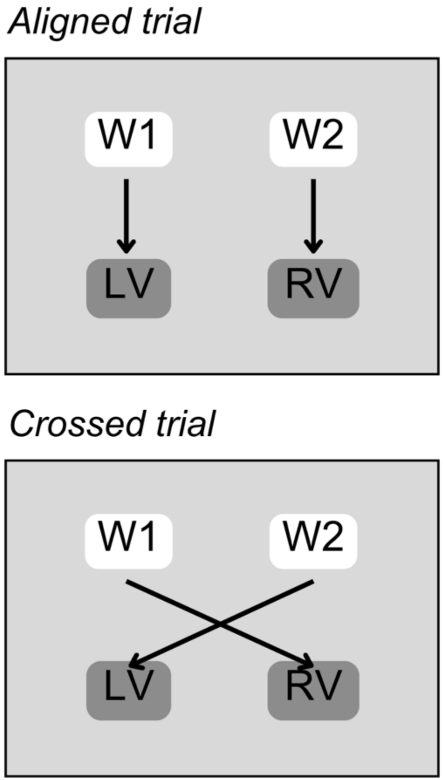
- Cross-situational statistical learning

Static depiction of the 8 symbolic gestures and matched target words

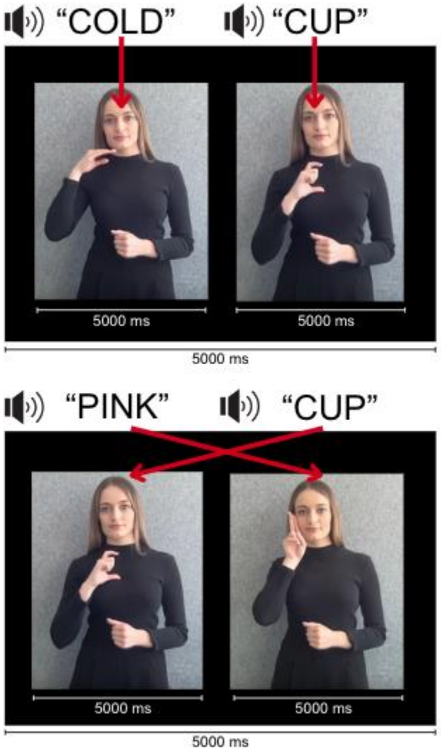
(Stimuli from Colombani et al., under review)

Cross-situational statistical learning





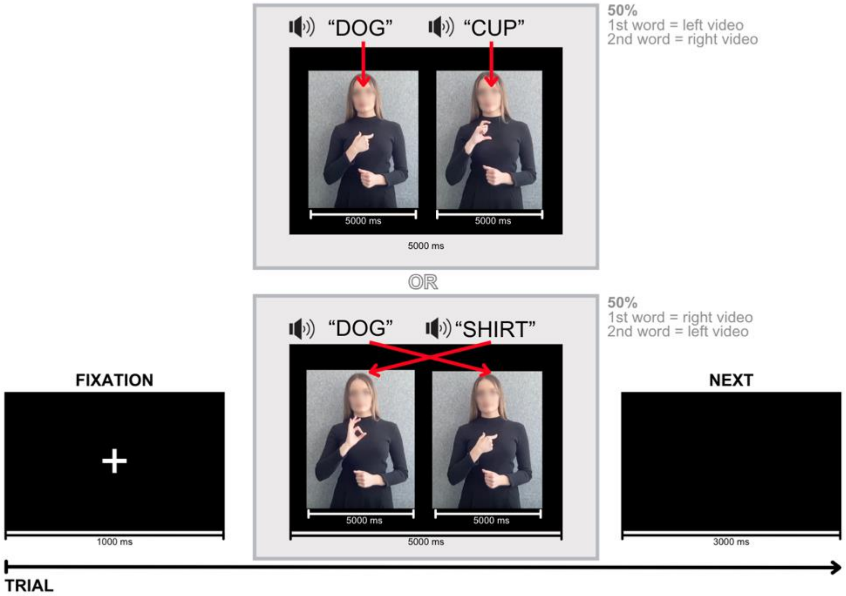
Cross-situational statistical learning



(Stimuli from Colombani et al., under review)

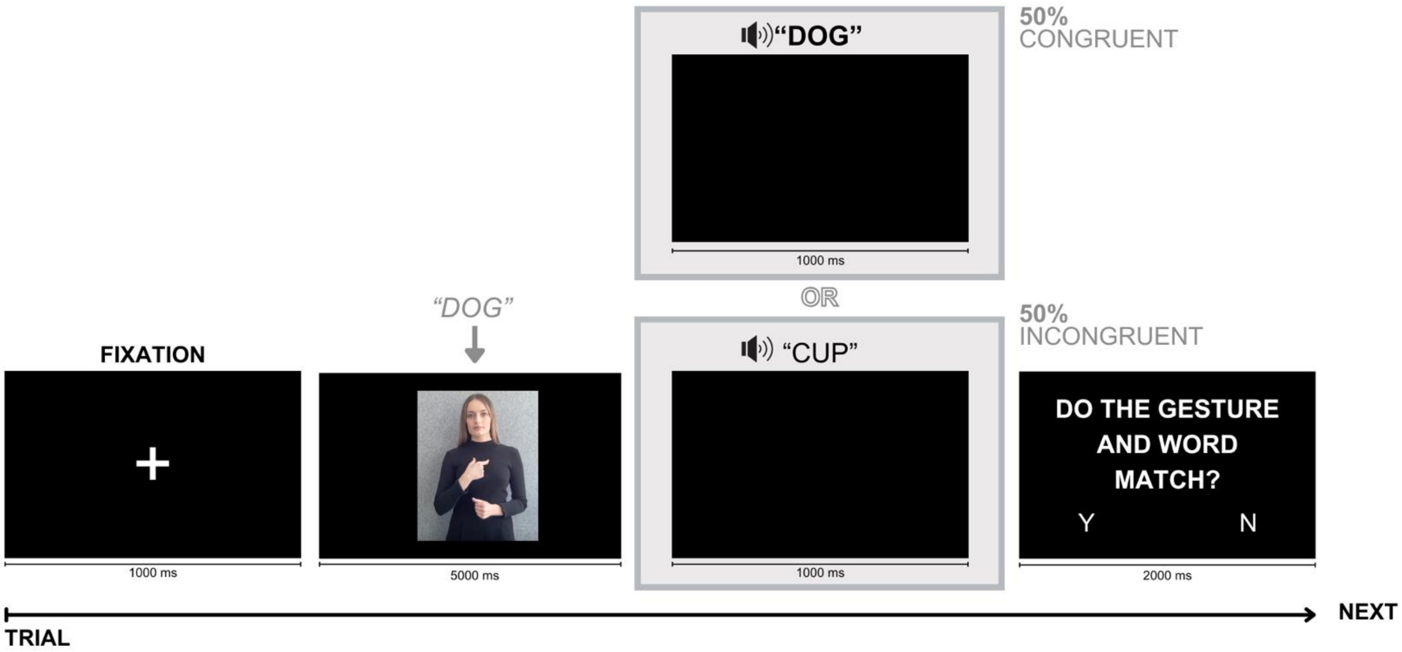
Procedure

1. Familiarization	2. Recognition task	3. Categorization task
8 target items X 12 repetitions Trial N = 48	check learning of gesture forms (yes/no task + EEG and N400) Trial N = 96	check semantic learning (yes/no task + EEG and N400) Trial N = 96



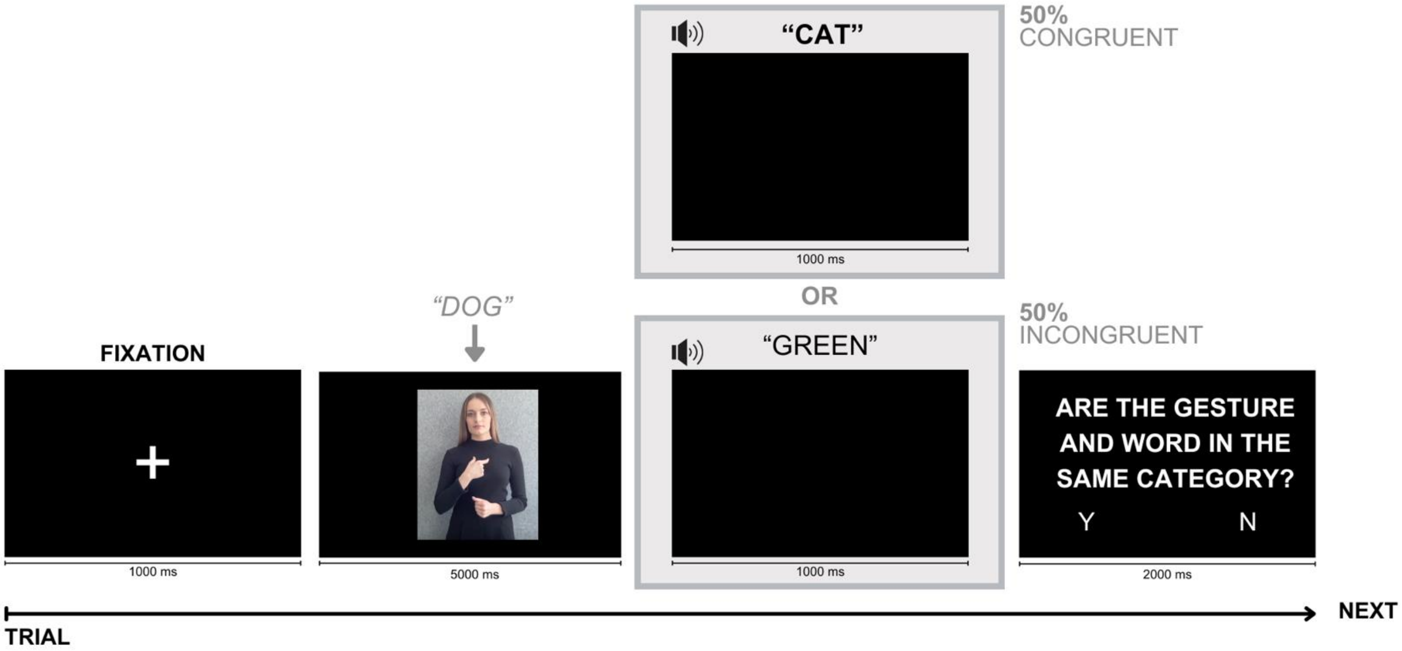
Procedure

1. Familiarization	2. Recognition task	3. Categorization task
8 target items X 12 repetitions Trial N = 48	check learning of gesture forms (yes/no task + EEG and N400) Trial N = 96	check semantic learning (yes/no task + EEG and N400) Trial N = 96



Procedure

1. Familiarization	2. Recognition task	3. Categorization task
8 target items X 12 repetitions Trial N = 48	check learning of gesture forms (yes/no task + EEG and N400) Trial N = 96	check semantic learning (yes/no task + EEG and N400) Trial N = 96



Analysis

Behavioral data:

- Accuracy = percentage of correct answers on total number of trials
- D-prime = measure of sensitivity that takes into account participants' response strategy

ERP data:

N400

- associated to semantic access/retrieval of the meaning of a word form
- interpreted as a mark of semantic processing (Kutas & Federmeier, 2011)

D-prime scores

D-prime (d') = quantifies the ability to distinguish between signal and noise

	Response: Yes	Response: No
Signal	***HIT***	***MISS***
No Signal	***FALSE ALARM***	***CORRECT REJECTION***

$$d' = Z(\text{Hit Rate}) - Z(\text{False Alarm Rate})$$

D-prime offers a continuous scale where higher values indicate better sensitivity:

d'= 0 : No sensitivity (cannot distinguish signal from noise).

d' > 0 : Increasing sensitivity as the value rises.

d' < 0 : Worse than chance (e.g., misunderstanding the task).

Accuracy is confounded by response bias. It only gives a percentage, which doesn't reveal the underlying detection dynamics.

Results

Behavioural results

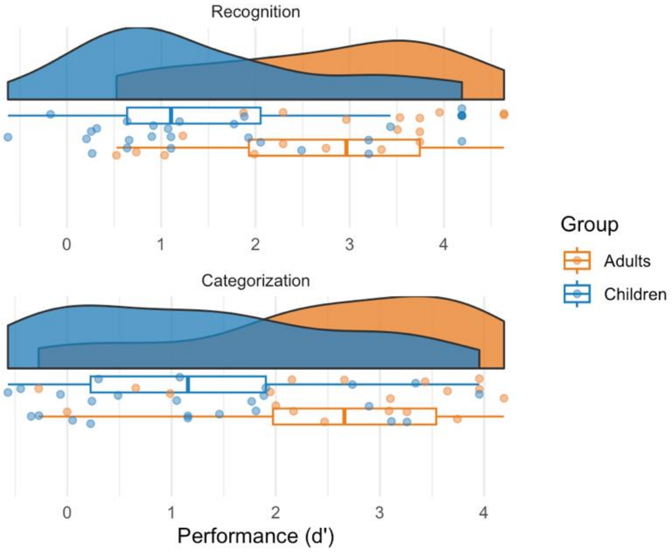
		Accuracy (%)			D-prime			Correlation	Bias
Group	Task	M	SD	WST	M	SD	WST	r	M
Adults (N=19)	Rec	87.4%	13.5	V = 190, p < .001*	2.76	1.27	V = 190, p < .001*	.93*	.880
	Cat	83.7%	16.6	V = 170, p < .001*	2.48	1.34	V = 170, p < .001*		
Children (N=24)	Rec	72.0%	17.5	V = 314.5, p < .001*	1.47	1.33	V = 318, p < .001*	.89*	.688
	Cat	69.0%	18.2	V = 295, p < .001*	1.29	1.34	V = 293, p < .001*		

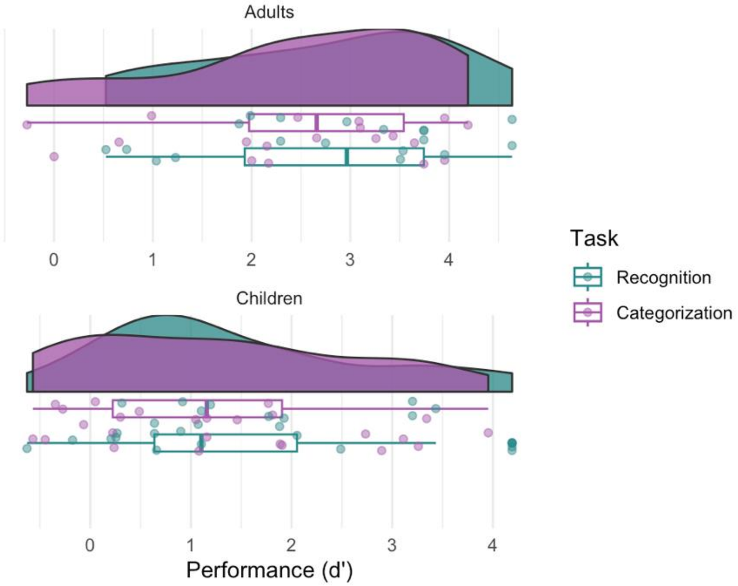
*Behavioral performance of the two groups in recognition and categorization tasks. M: mean, SD: standard deviation. r: correlation coefficient. Wilcoxon signed-rank test was performed to test significance against chance. *p < .001*

1. In both groups:
   - accuracy and d-prime scores above chance
   - scores in the two tasks were strongly and positively correlated
2. Between groups:
   - no difference in the response strategy

Statistical analysis

d-prime ~ (Task * Group)





- Significant effect of group

(adults performed better than children)

- No effect of task

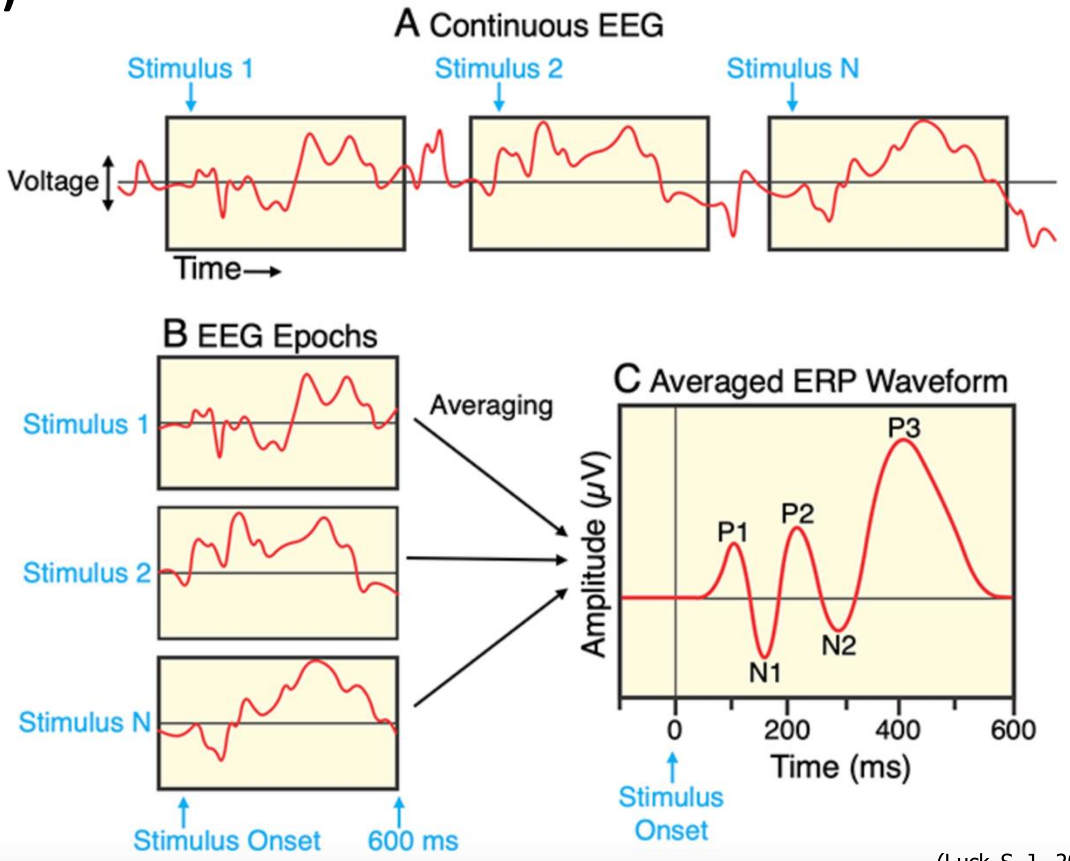
(equally learnt the gestures and their meaning)

Event-related potentials (ERPs)

Event-related potentials are very small voltages generated in the brain structures in response to specific events or stimuli.

To extract average waveforms from the EEG,it is necessary to sum and average the results of multiple trials.

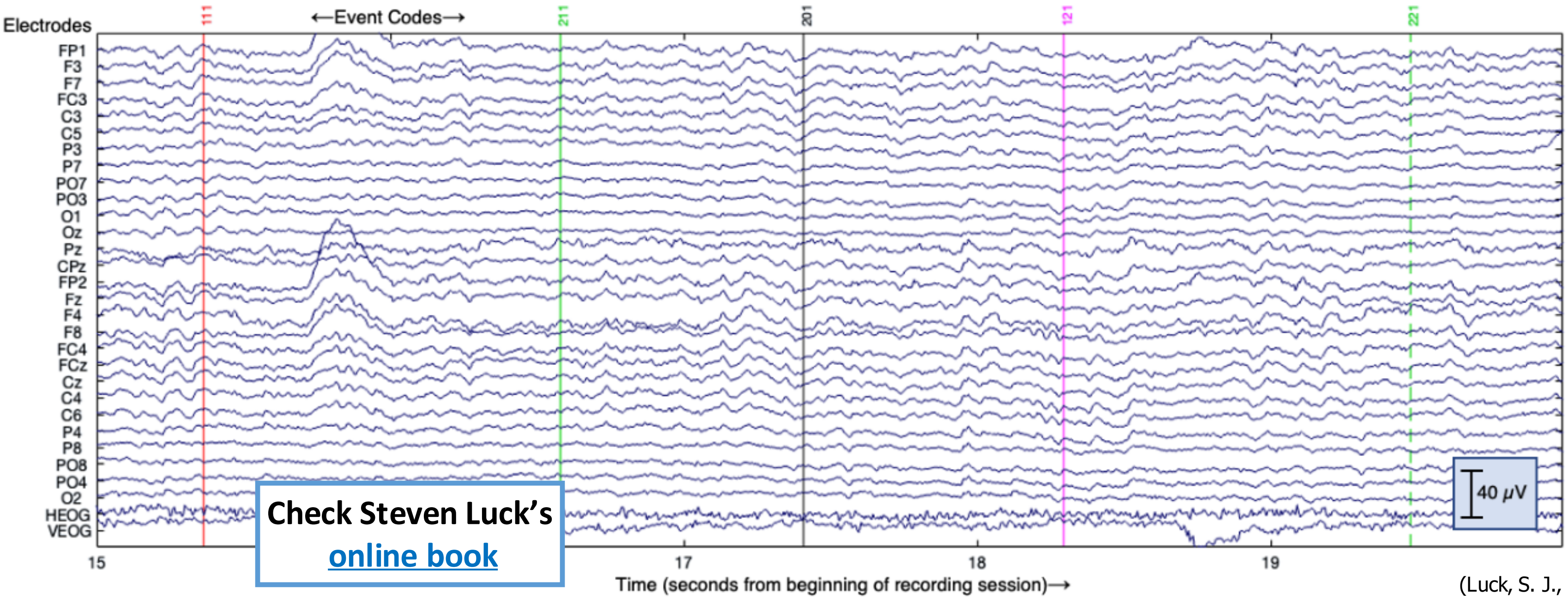
Check Steven Luck's [online book]



(Luck, S. J., 2022)

Event-related potentials (ERPs)

The raw, continuous EEG is a complex signal and includes all the activity generated by thousands of processes in the cortex. It is **impossible to extract brain activity** associated with a particular cognitive process by **viewing only the raw EEG**.



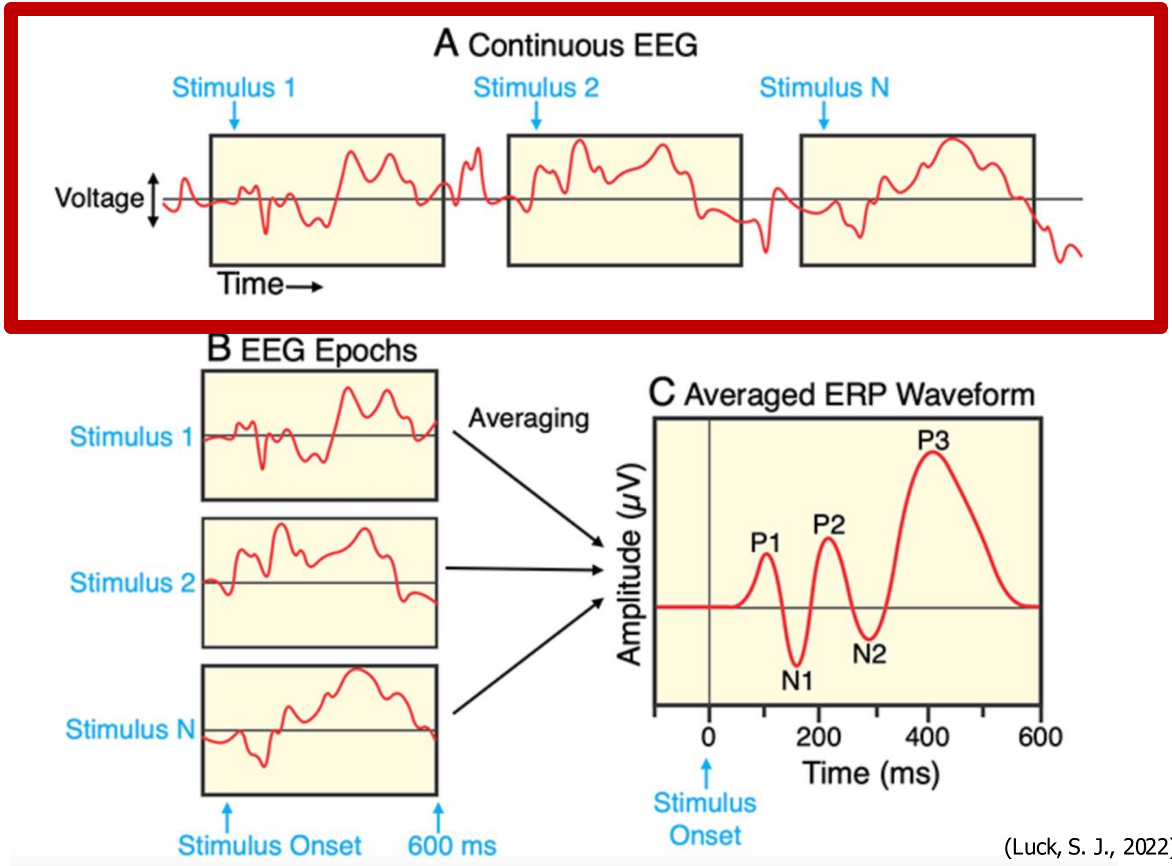
Check Steven Luck's [online book]

(Luck, S. J., 2022)

Event-related potentials (ERPs)

The first step is to divide the continuous EEG into specific time windows (called EPOCHS) surrounding the event presentation, i.e. the experimental stimulus.

In each of these epochs there will be event-related brain activity (ERP), plus all other ongoing activity in the brain, which we call NOISE.



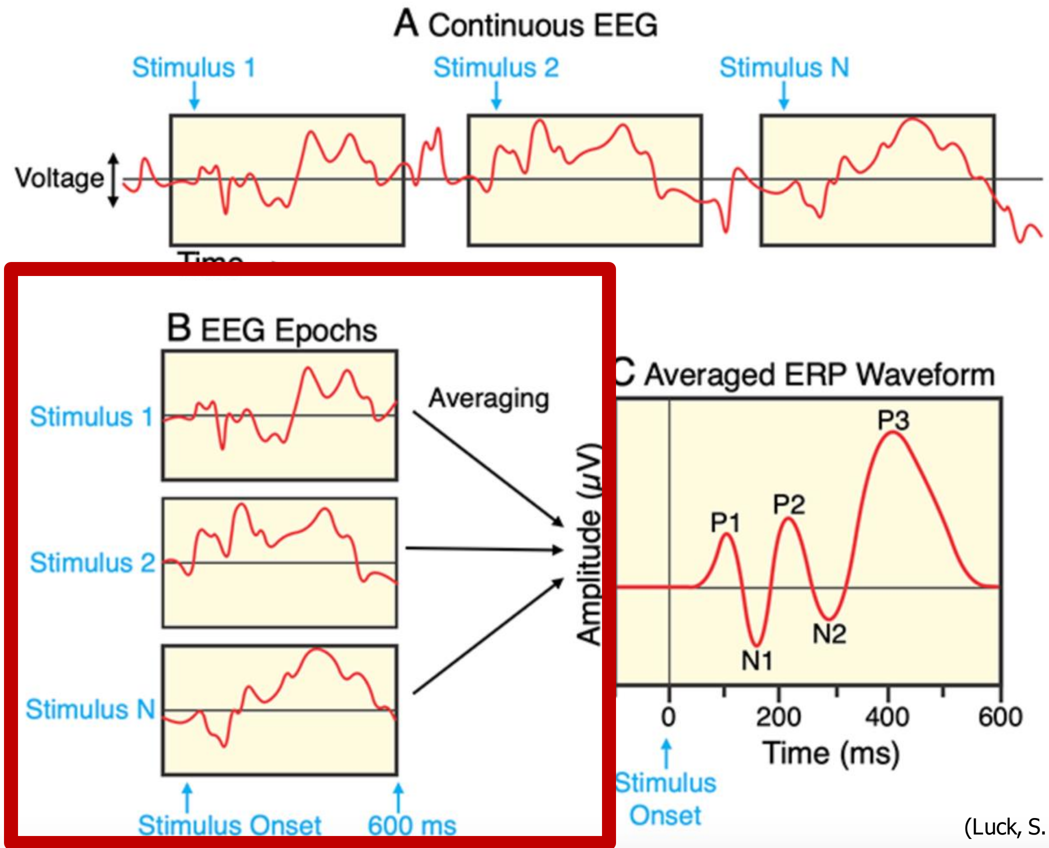
(Luck, S. J., 2022)

Event-related potentials (ERPs)

The second step is to align all epochs with respect to the start of the stimulus (0 ms) and average the voltage.

The signal connected to the start of the stimulus (i.e. the ERP), will be consistent between epochs.

The noise (i.e. the voltages connected to other brain activities), will be so different that they will be averaged between them.



(Luck, S. J., 2022)

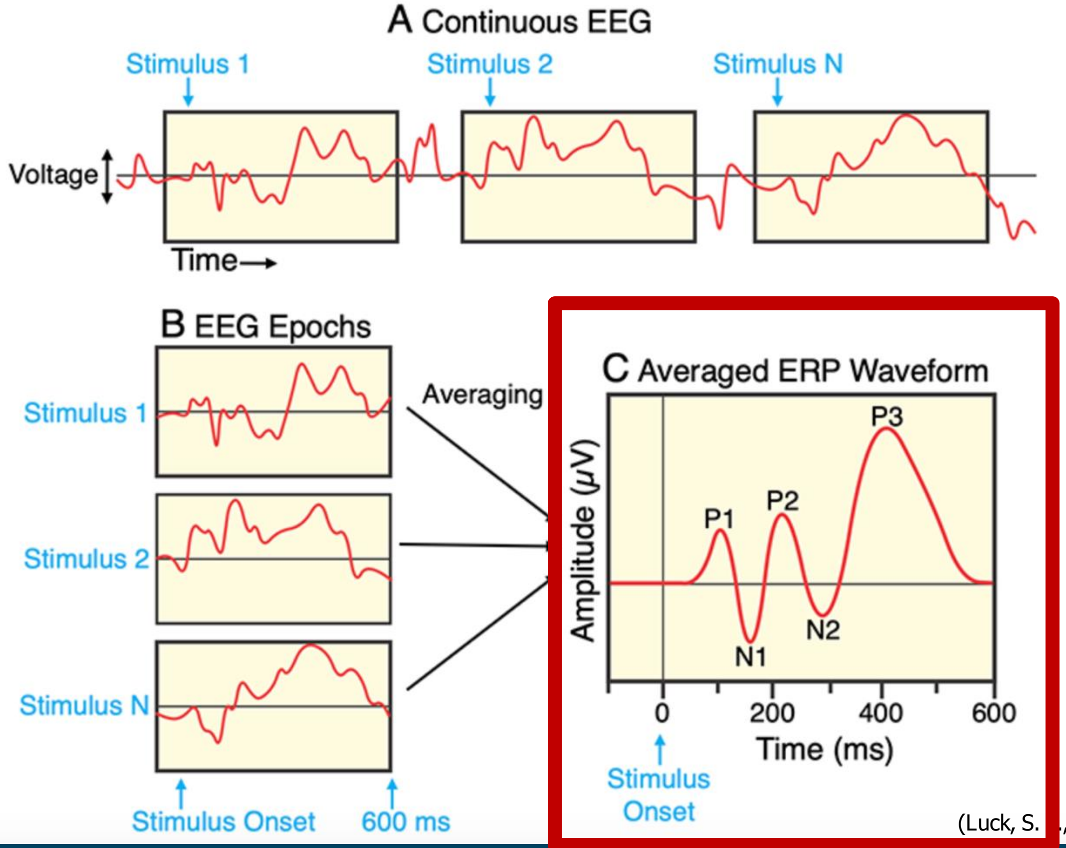
Event-related potentials (ERPs)

The ERP waveform consists of a set of positive and negative waves (peaks), related to a set of underlying components in the brain that reflect specific neurocognitive processes.

N400 = negative spike ~400 ms after stimulus onset.

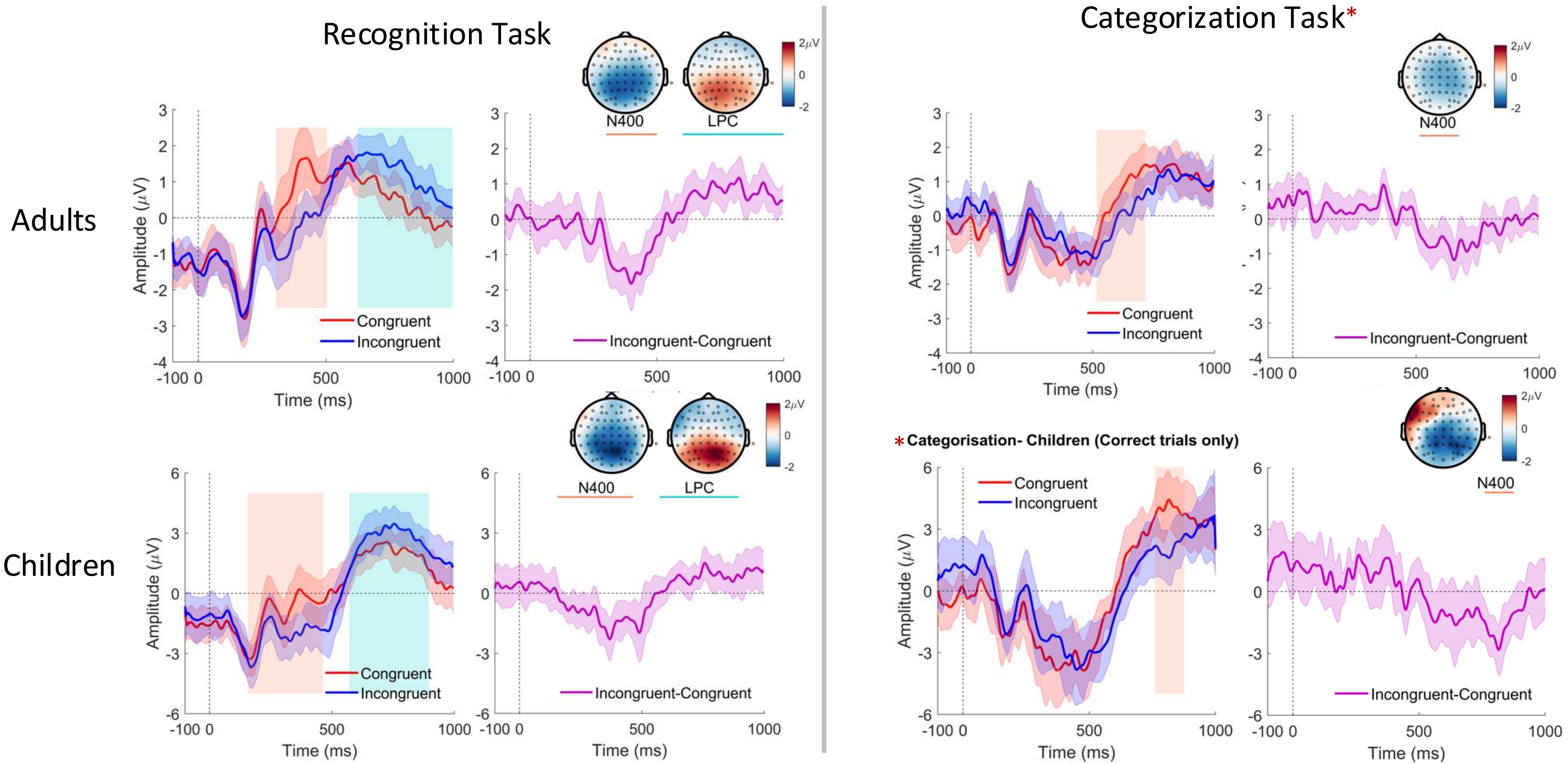
One of the most studied components. Discovered in the 1980s, it has been widely associated with semantic processing.

(Kutas and Hillyard, 1980; Kutas and Federmeier, 2011)



(Luck, S. ., 2022)

ERP results – Overview



Discussions

Behavioral results:

	RECOGNITION	CATEGORIZATON
Group	Adults better than children	
Task	No effect of task	

Sign-like gesture can be

- **learned through statistical learning** (recognition)
- **rapidly given a meaning** (categorization)
- advantage of adults (e.g. memory, attention)

ERP results:

	RECOGNITION	CATEGORIZATON
Adults	N400 - P600	N400
Children	N400 - P600	N400 (correctly identified trials)

Sign-like gesture

- elicit **brain responses** (N400) similar to spoken words and sign laguages **(i.e., similar to language input)**

Conclusions

- Despite the ambiguous learning context
- Naive to gestural communication languages (i.e., sign languages / baby signing,)
- No instruction on the task
- No associative cues

⇩

Sign-like gesture can:

- be **learned through statistical learning**
- be **rapidly given a semantic representation**
- elicit **brain responses (N400) similar to spoken words**

⇩

Sign-like gestures can be perceived as **linguistic, meaningful referents.**

⇩

Test this hypothesis on a younger age group (Exp.2)

To wrap up…

1. Statistical learning (SL) is a learning mechanism applied since infancy to acquire a varied set of language properties.

2. Cross-situational learning is a SL-based paradigm that recreates an ambiguos learning environment to test word-to-world mapping in the lab setting.

3. Sign-like gestures are similar to sign language vocabulary but used as context-indipendent gestures to refer to specific referents and meaning.

4. Sign-like gestures can be learned by adults and older children through statistical computations in a cross-situatonal learning setting, as gestural labels for referents.

Thank you!

arianna.colombani@uni-potsdam.de

http://clmrnn.github.io

Additional readings

Statistical learning reviews

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical Learning of Language: A Meta-Analysis Into 25 Years of Research. Cognitive Science, 46(9). https://doi.org/10.1111/cogs.13198

Roembke, T. C., Simonetti, M. E., Koch, I., & Philipp, A. M. (2023). What have we learned from 15 years of research on cross-situational word learning? A focused review. Frontiers in Psychology, 14(July), 1–9. https://doi.org/10.3389/fpsyg.2023.1175272

Baby signing

Vallotton, C. D. (2011). Babies open our minds to their minds: How "listening" to infant signs complements and extends our knowledge of infants and their development. Infant Mental Health Journal, 32(1), 115–133. https://doi.org/10.1002/imhj.20286

The must-read guides to ERPs!

Luck, S. J. (2022). Applied Event-Related Potential Data Analysis. LibreTexts. https://doi.org/10.18115/D5QG92

Luck, S.J. (2014). An Introduction to the Event-Related Potential Technique, 2nd Edition. MIT Press.

N400

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647. https://doi:10.1146/annurev.psych.093008.131123

Junge, C., Boumeester, M., Mills, D. L., Paul, M., & Cosper, S. H. (2021). Development of the N400 for Word Learning in the First 2 Years of Life: A Systematic Review. Frontiers in Psychology, 12(June). https://doi.org/10.3389/fpsyg.2021.689534

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. Psychological Science, 9, 321–324. https://doi:10.1111/1467-9280.00063

Brooks, P., & Kempe, V. (2014). Statistical Learning. Encyclopedia of Language Development, August, 1–21. https://doi.org/10.4135/9781483346441.n191

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). The Emergence of symbols: cognition and communication in Infancy. New York, NY: Academic Press.

Capirci, O., & Volterra, V. (2008). Gesture and speech. Gesture, 8(1), 22–44. https://doi.org/10.1075/gest.8.1.04cap

Colombani, A., Peter, V., Quian, Y. M., Saksida, A., Boll-Avetisyan, N., Tuomainen, O., & Sharma, M. (under review). Cross-situational learning of word-pseudosign pairs in children and adults: a behavioural and event-related potential study. Language, Cognition and Neuroscience.

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016). Cross-Situational Learning of Minimal Word Pairs. Cognitive Science, 40(2), 455–465. https://doi.org/10.1111/cogs.12243

Fabbri-Destro, M., Avanzini, P., De Stefani, E., Innocenti, A., Campi, C., & Gentilucci, M. (2015). Interaction Between Words and Symbolic Gestures as Revealed By N400. Brain Topography, 28(4), 591–605. https://doi.org/10.1007/s10548-014-0392-4

Gleitman, L. (1990). The structural sources of verb meanings. Language Acquisition, 1, 1–55. http://www.jstor.org/stable/20011341

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical Learning of Language: A Meta-Analysis Into 25 Years of Research. Cognitive Science, 46(9). https://doi.org/10.1111/cogs.13198

Kendon, Adam. 1988. How gestures can become like words. In Fernando Poyatos (ed.), Crosscultural Perspectives in Nonverbal Communication, 131–141. Toron- to: C. J. Hogrefe Publishers.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205. Cognitive Science, 46(9). 10.1126/science.7350657

References

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647. https://doi:10.1146/annurev.psych.093008.131123

Luck, S. J. (2022). Applied Event-Related Potential Data Analysis. LibreTexts. https://doi.org/10.18115/D5QG92

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. Current Directions in Psychological Science, 12(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science (New York, N.Y.), 274(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. Cognition, 106(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Yu, C., & Smith, L. (2008b). Statistical cross-situational learning inadults and infants. In C. Press (Ed.), Proceedings of the 32th Annual Boston UniversityConference on Language Development (pp. 562–573).

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics: Research article. Psychological Science, 18(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. Proceedings of the National Academy of Sciences of the United States of America, 106(49), 20664–20669. https://doi.org/10.1073/pnas.0909197106

Additional slides

Response bias

RespBias ~ GroupF

The main effect of **Group is statistically not significant** and small

↓

both groups tended to be conservative