Refining Infant Word Recognition Analysis:
Optimising Gaze Feature Selection based on Random Forest

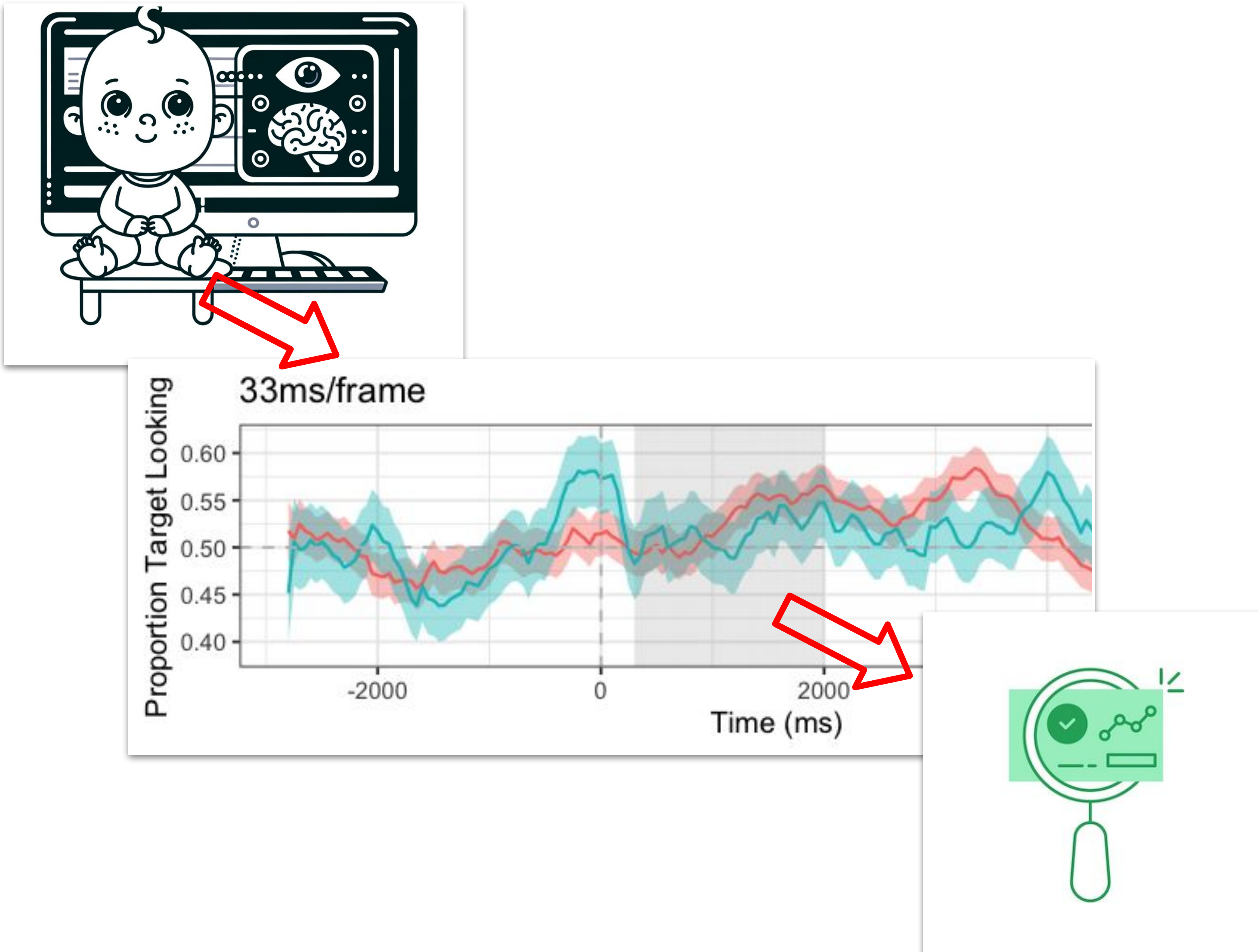
Jun Ho Chai[1], Minji Kim[2], Minkyu Shim[2], Youngki Lee[2], Eon-Suk Ko[1]

[1] Chosun University, [2] Seoul National University

CDS Center for Data Science in the Humanities Chosun University

조선대학교 인문데이터과학연구소

VERITAS LUX MEA


33ms/frame
Proportion Target Looking
0.60
0.55
0.50
0.45
0.40
-2000
0
2000
Time (ms)

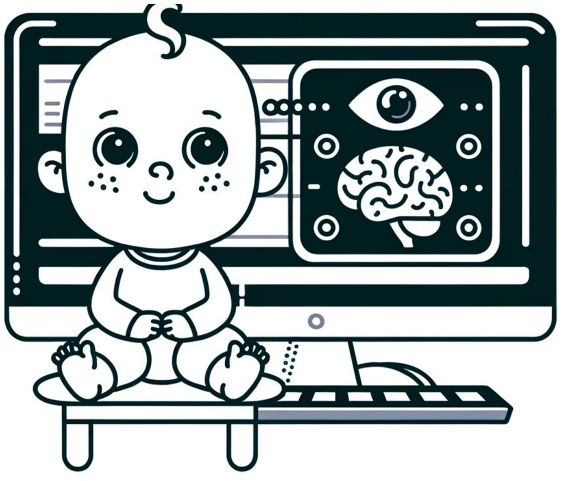
Challenges in Infant Eye-tracking Studies

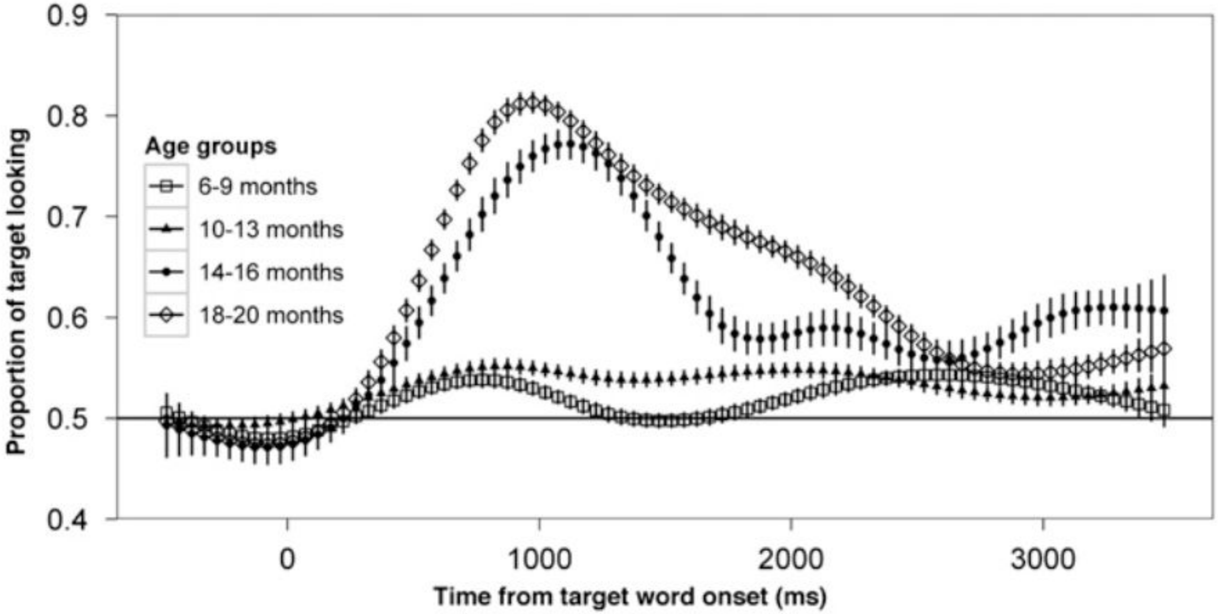
Time Window Reliability

- Debate over the reliability of different time windows

Age-Specific Analysis

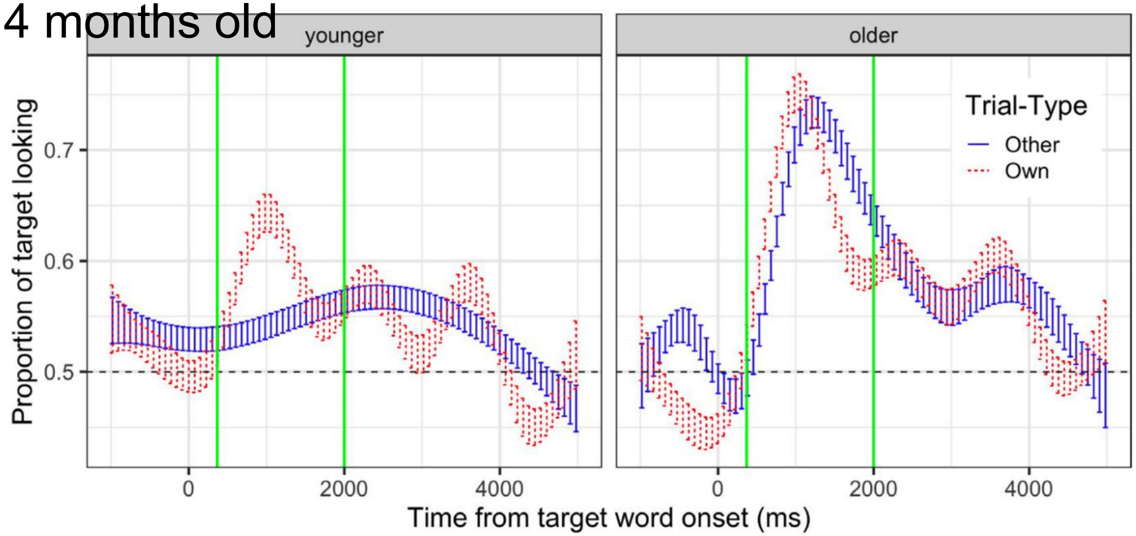
- Challenging to cater to a broad age range (10-36 months)
- Potential ceiling effects in older children
- Discrepancies in how younger and older kids respond to visual stimuli



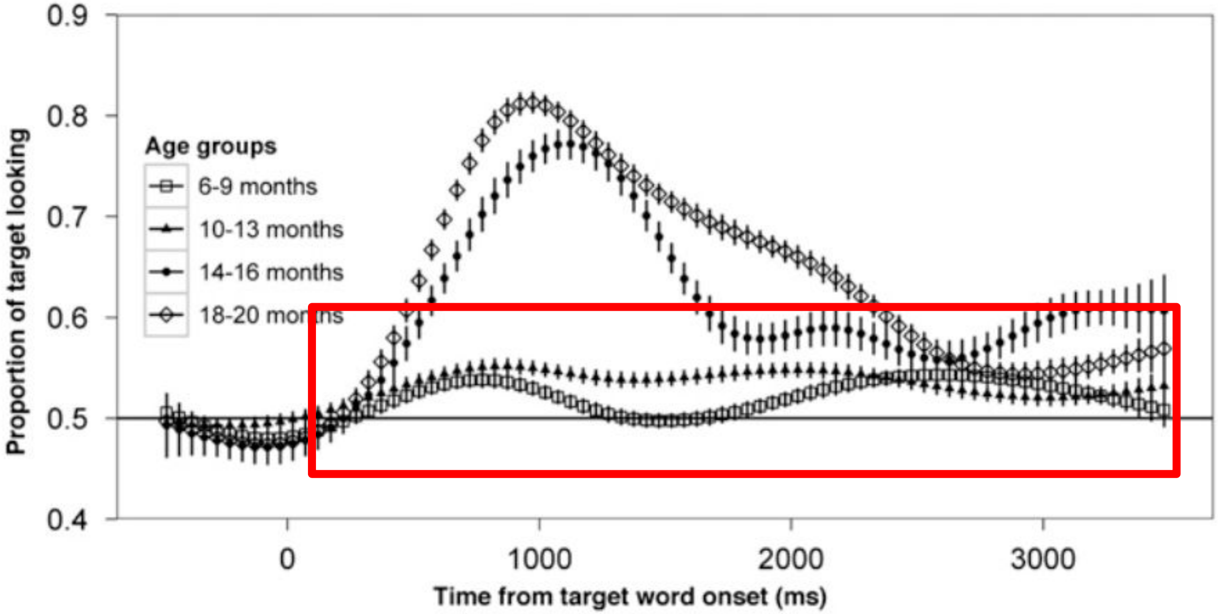


Bergelson & Swingley (2012)

~14 months old

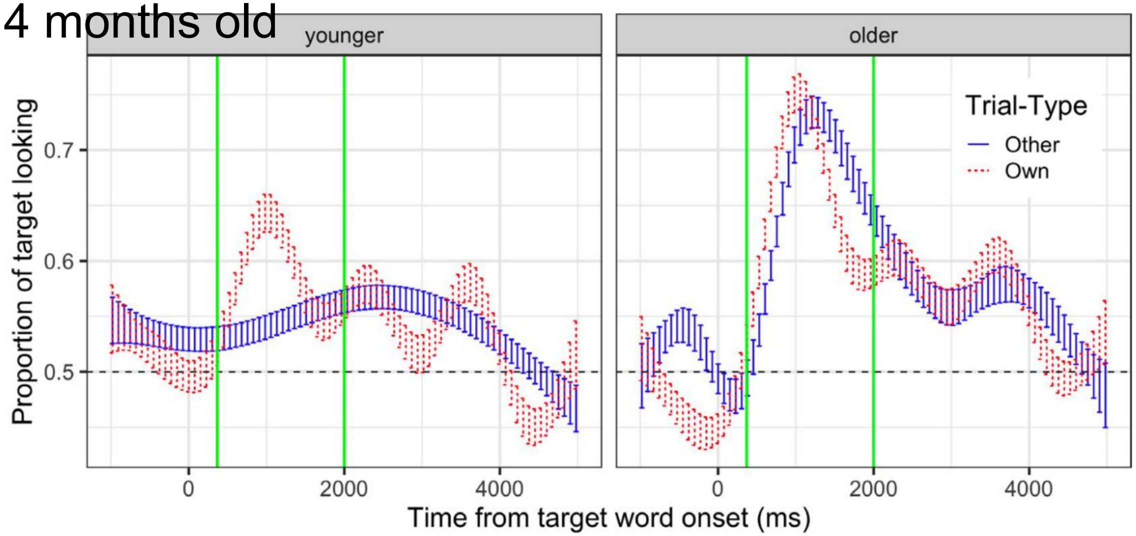


Garrison et. al. (2020)

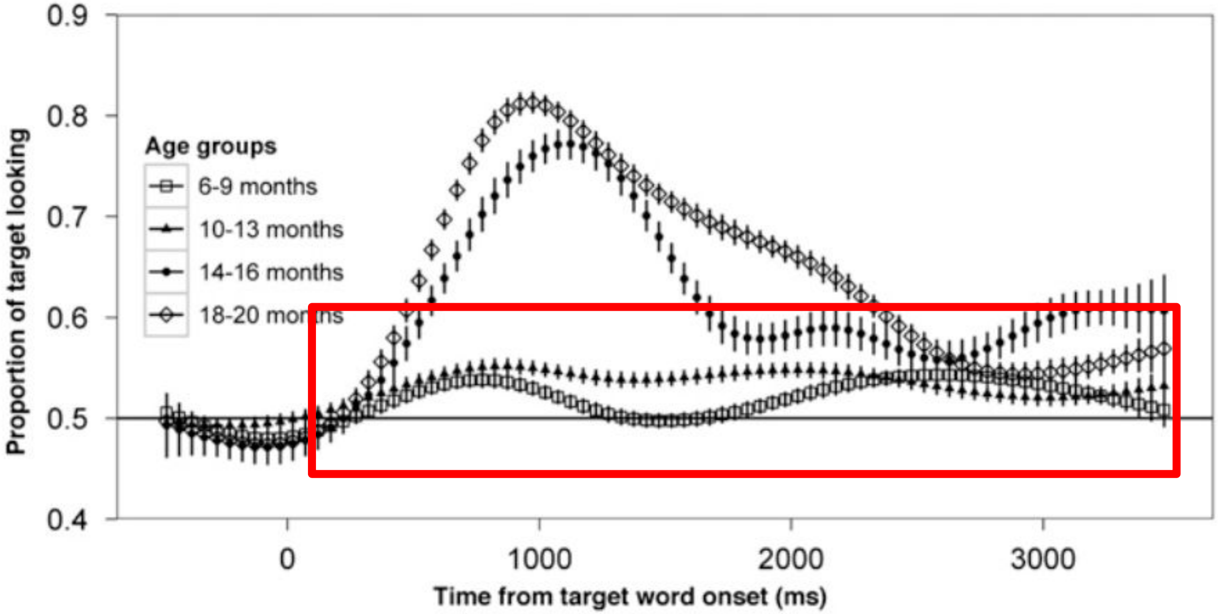


Bergelson & Swingley (2012)

~14 months old

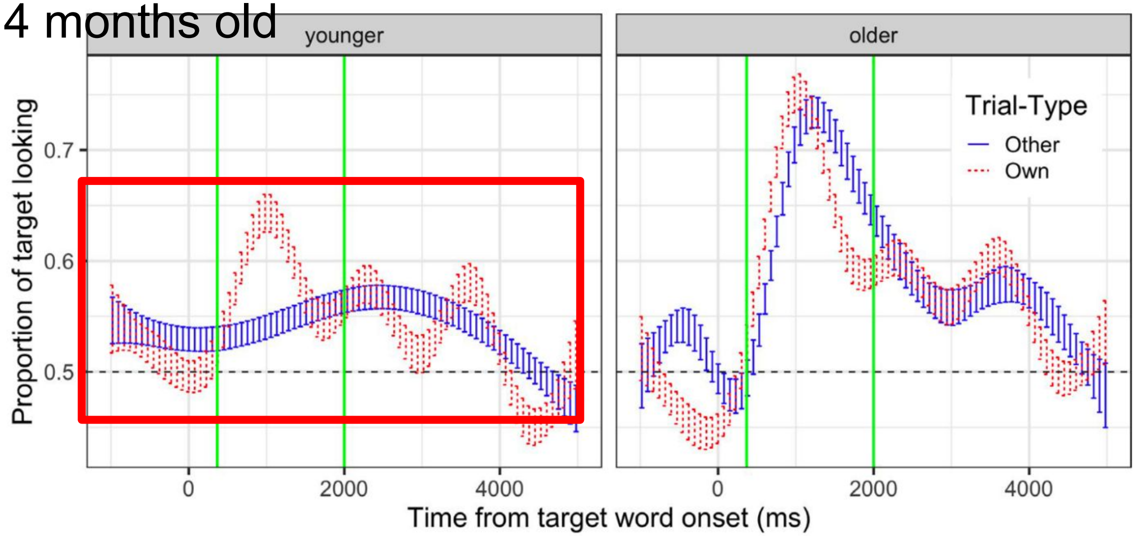


Garrison et. al. (2020)



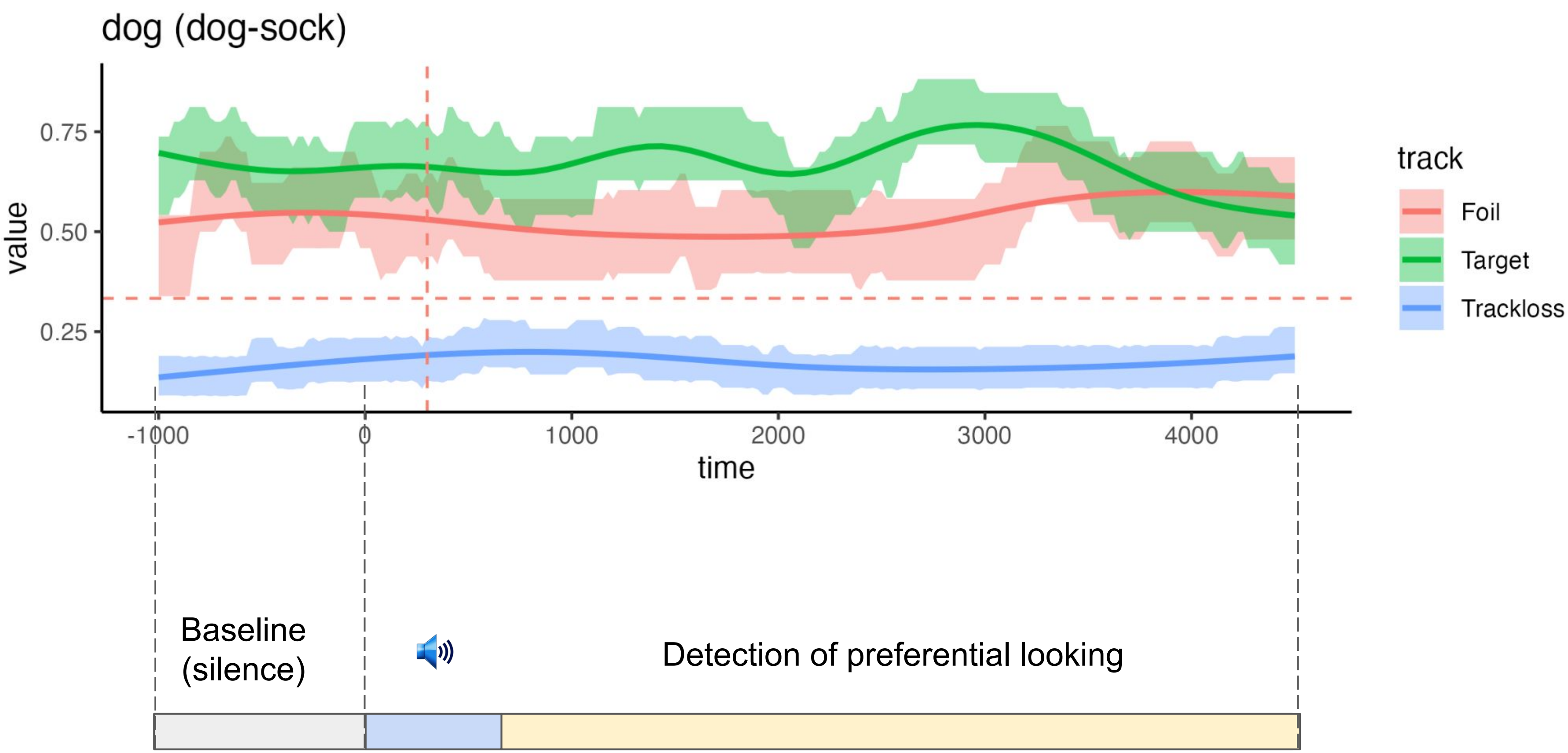
Bergelson & Swingley (2012)

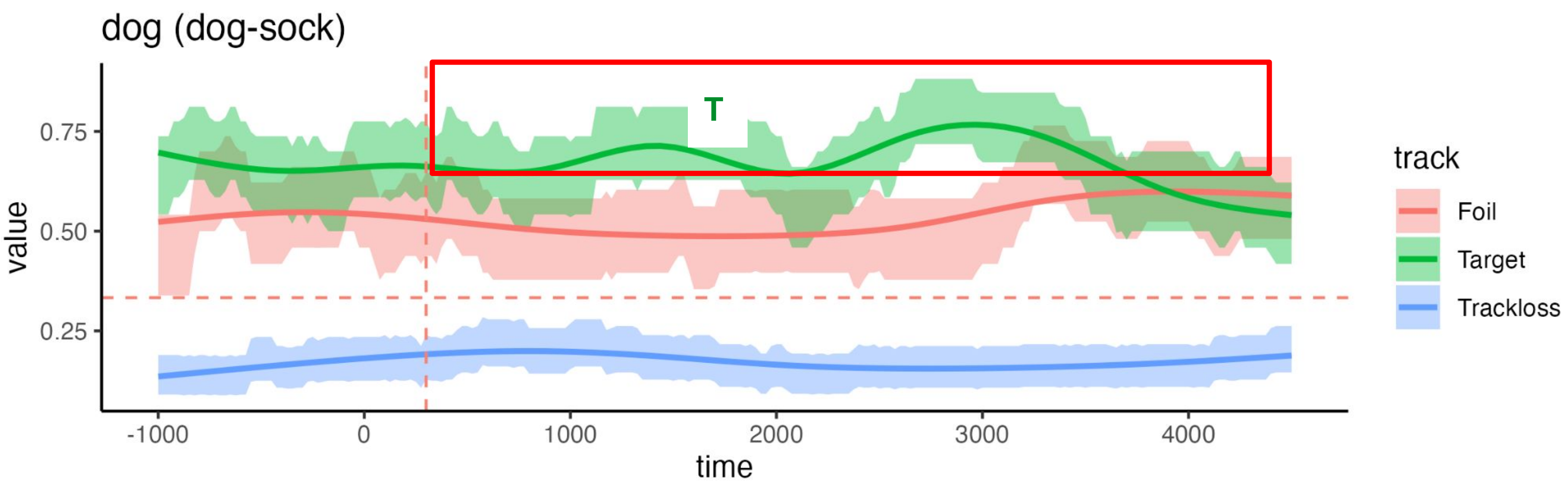
~14 months old



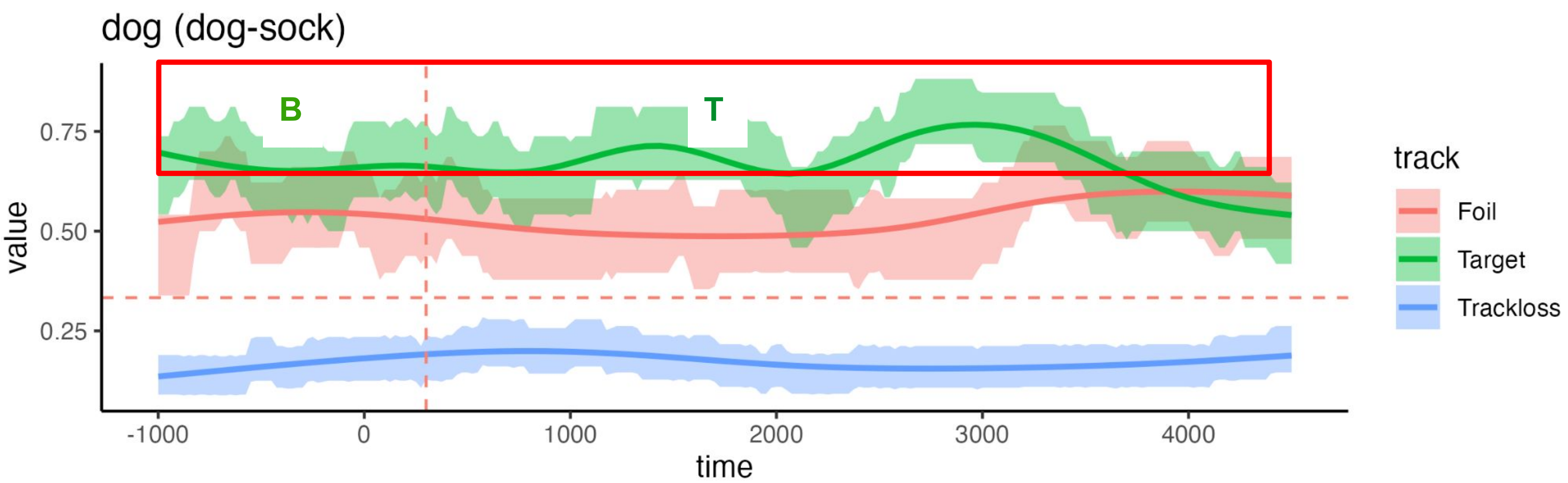
Garrison et. al. (2020)

Common gaze adjustment method

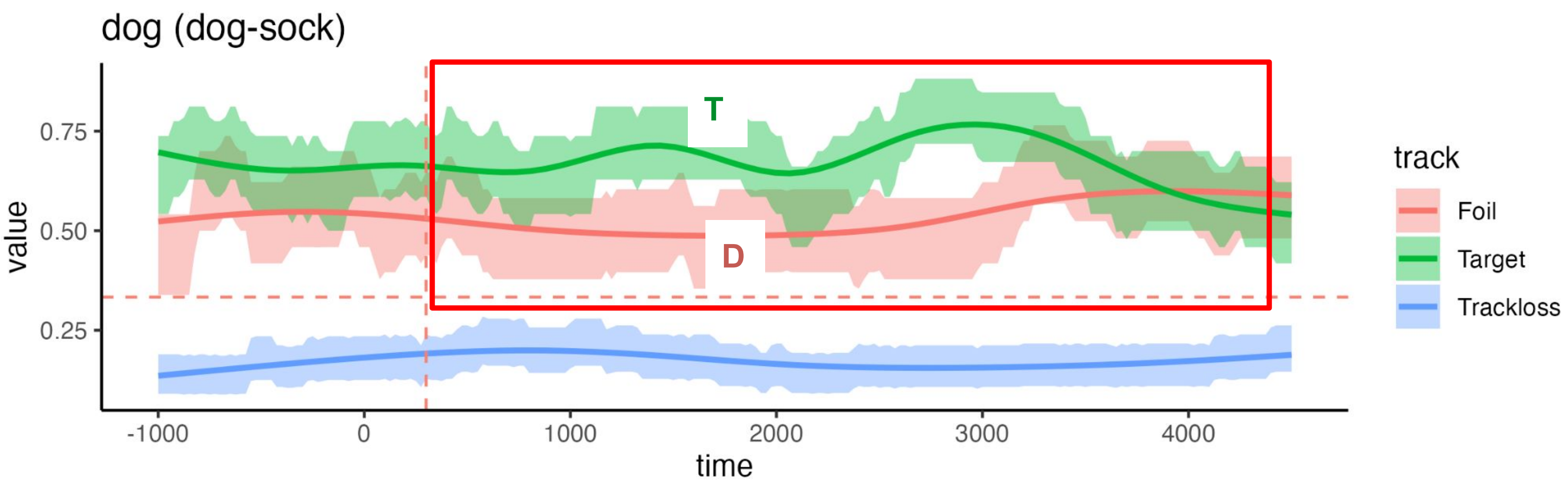




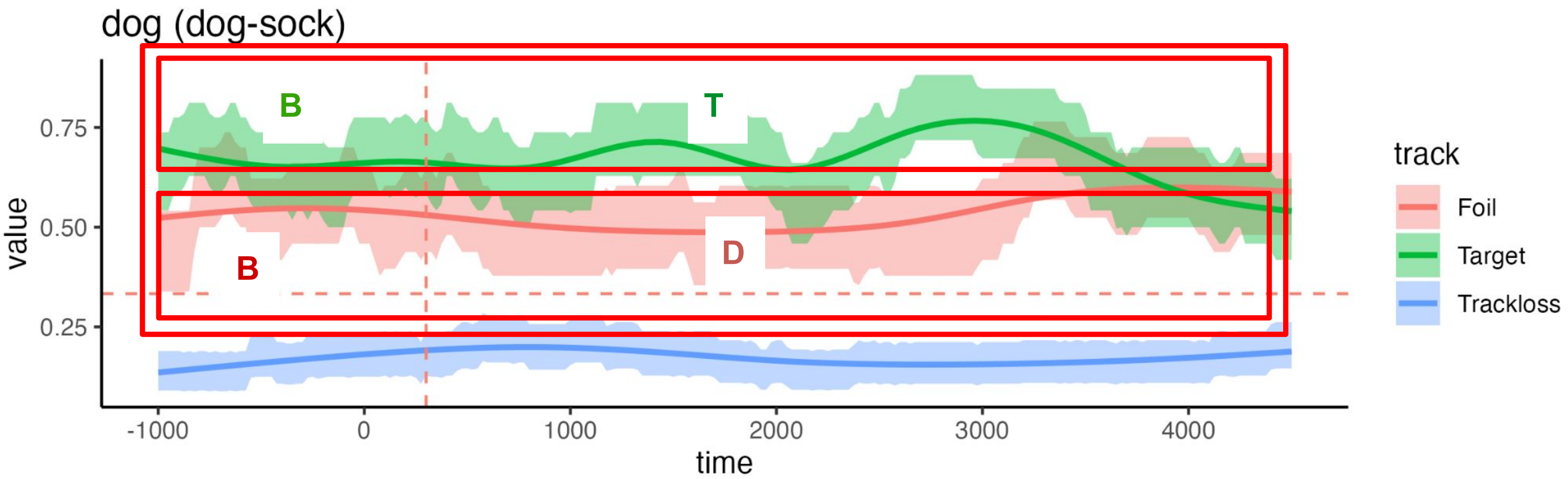
- Proportion of look to target (**raw**): **T**



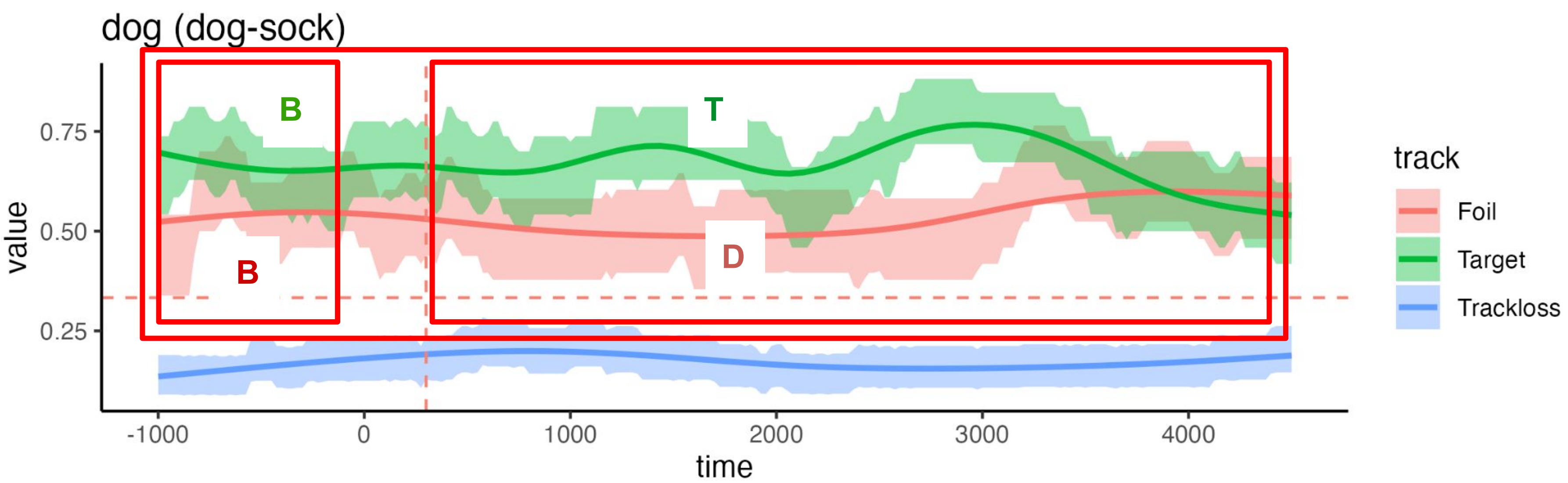
- Proportion of look to target (**raw**): **T**
- Proportion adjusted by **baseline**: **T - B**



- Proportion of look to target (**raw**): **T**
- Proportion adjusted by **baseline**: **T - B**
- Proportion adjusted by **image: T - D** (requires at least a yoked pair)

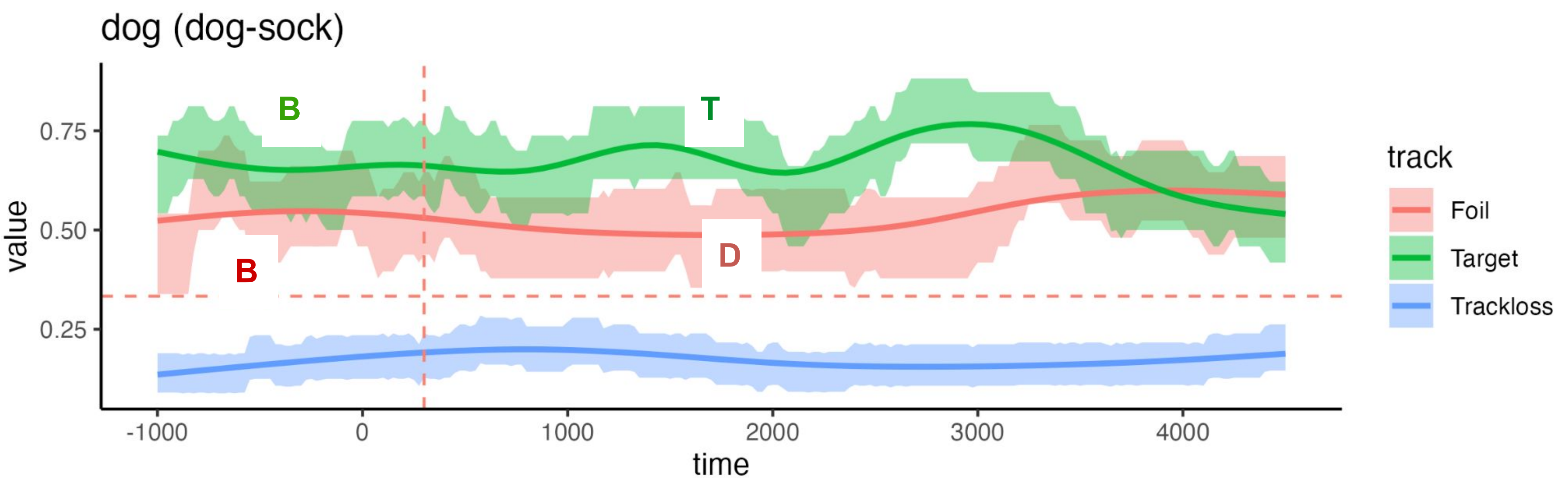


- Proportion of look to target (**raw**): **T**
- Proportion adjusted by **baseline**: **T - B**
- Proportion adjusted by **image: T - D** (requires at least a yoked pair)
- Proportion adjusted by **image and baseline: (T - B) - (D - B)**



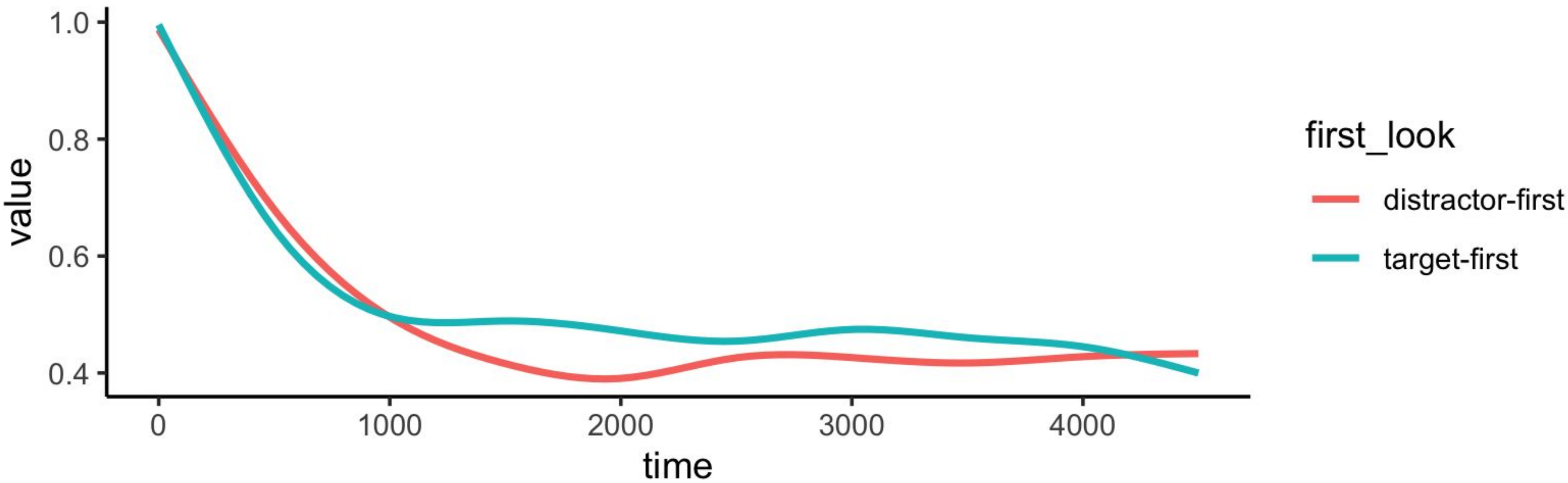
- Proportion of look to target (**raw**): **T**
- Proportion adjusted by **baseline**: **T - B**
- Proportion adjusted by **image: T - D** (requires at least a yoked pair)
- Proportion adjusted by **image and baseline: (T - B) - (D - B) or**

(T - D) - (B - B)



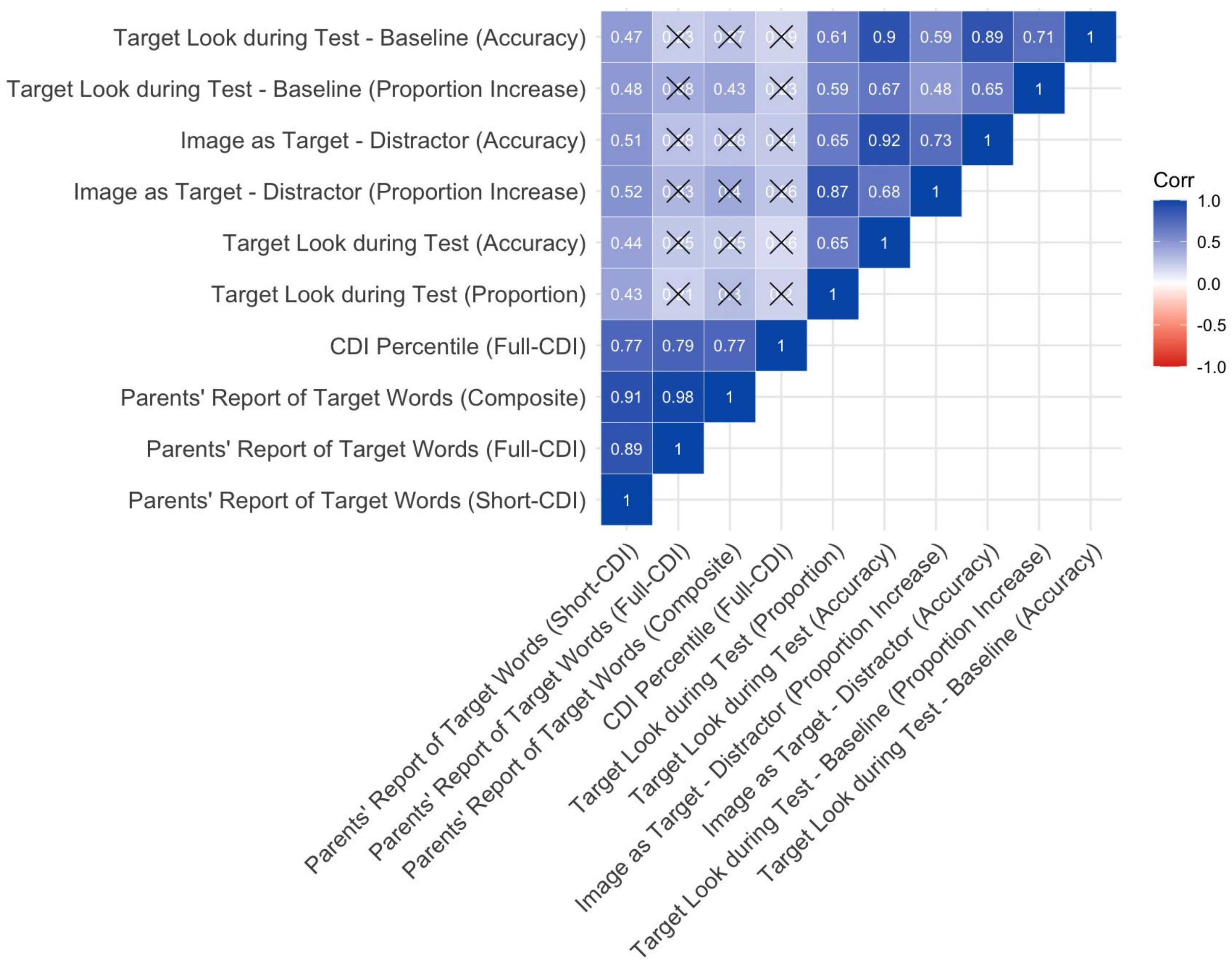
- Proportion of look to target (**raw**): **T**
- Proportion adjusted by **baseline**: **T - B**
- Proportion adjusted by **image: T - D** (requires at least a yoked pair)
- Proportion adjusted by **image and baseline: (T - B) - (D - B) or**

  (T - D) - (B - B)
- Target Time Window: 0 - 1000ms:1000 - 2000ms:2000 - 3000ms…etc

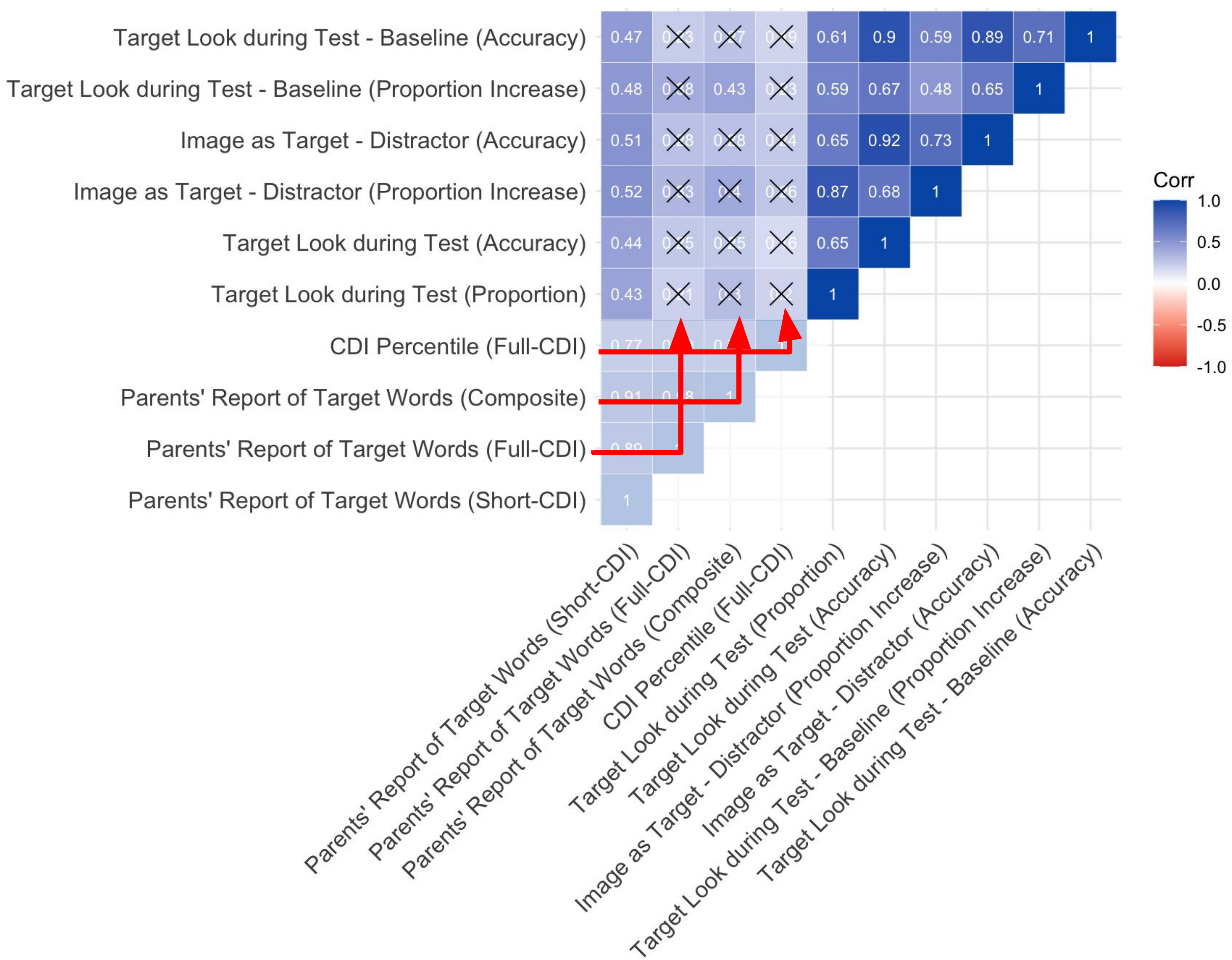


- **Switch** (latency:switch proportion)

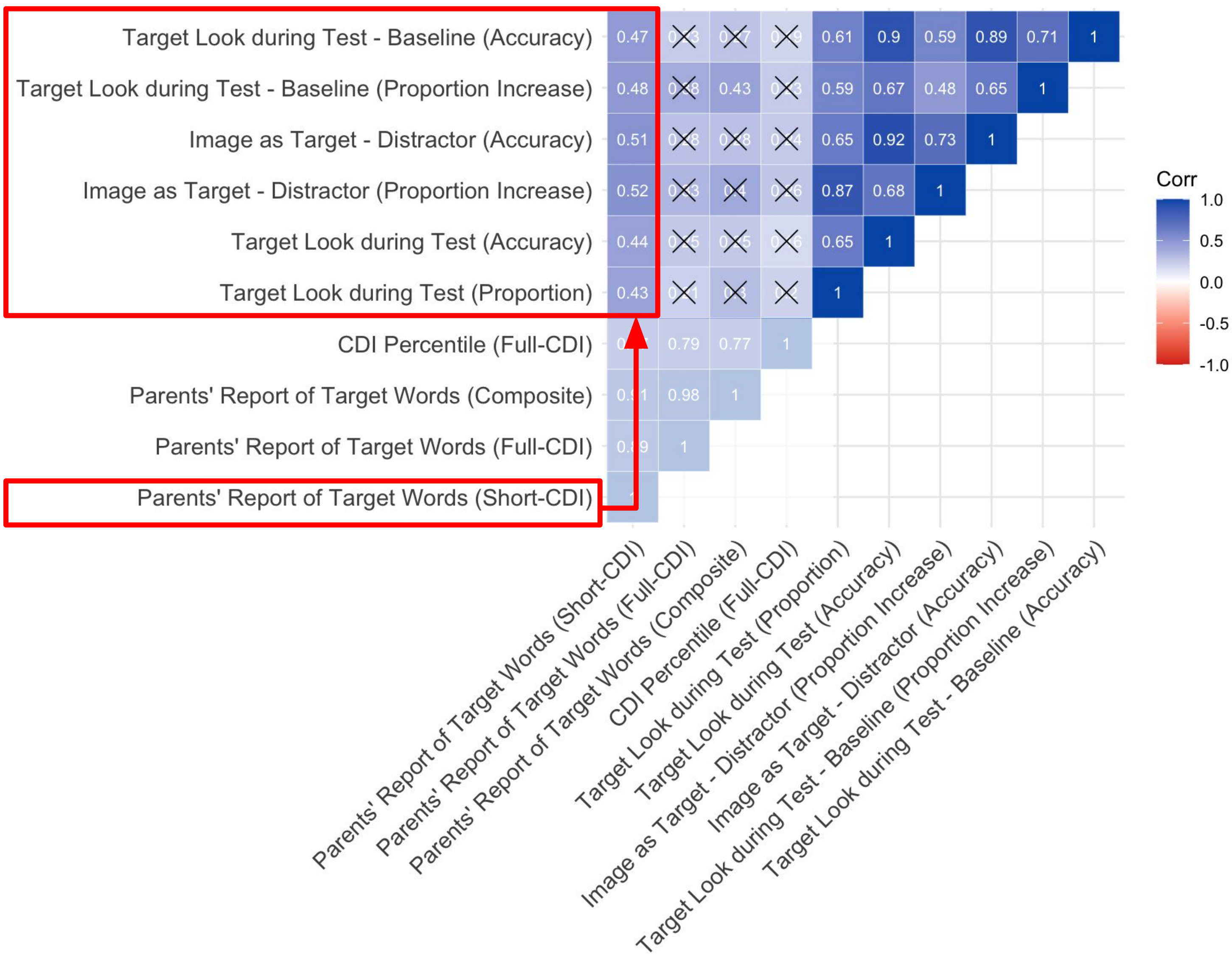
Infant eye-tracking studies with 14mo Korean infants



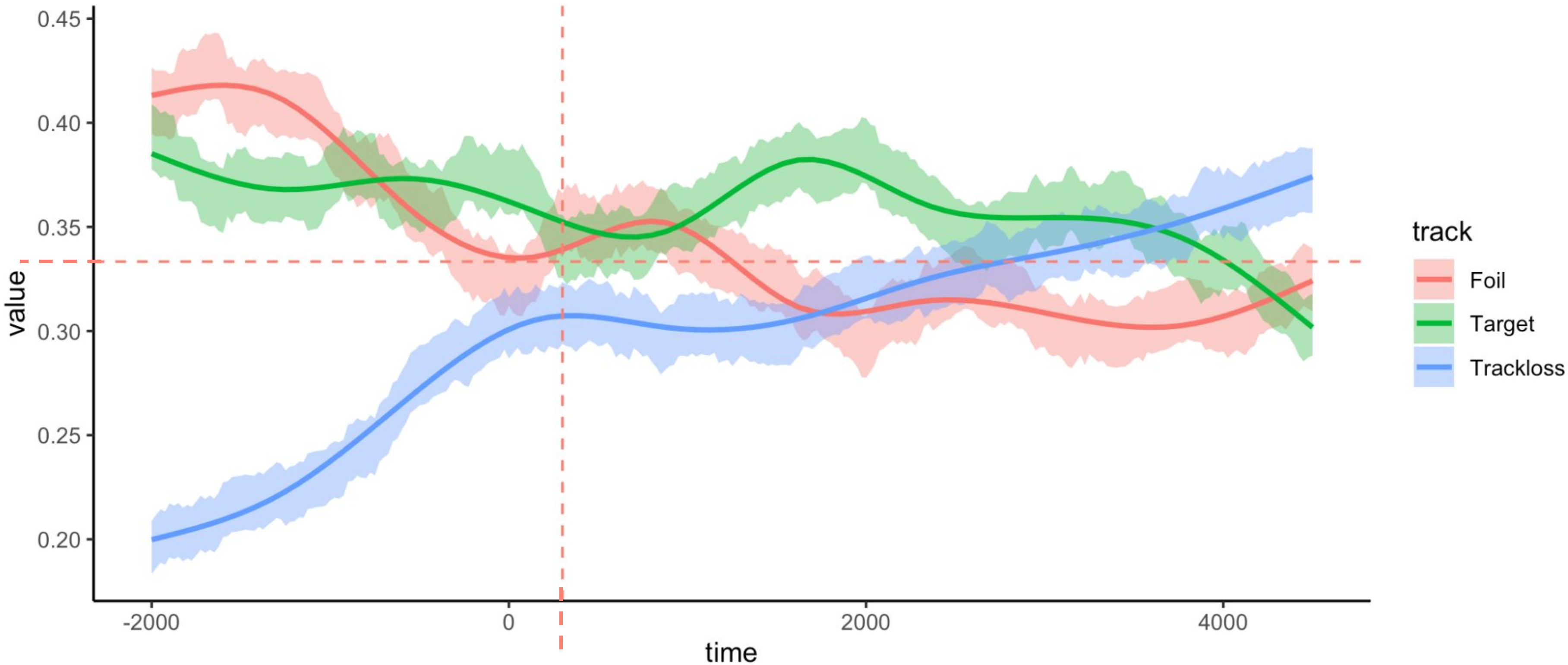
Infant eye-tracking studies with 14mo Korean infants



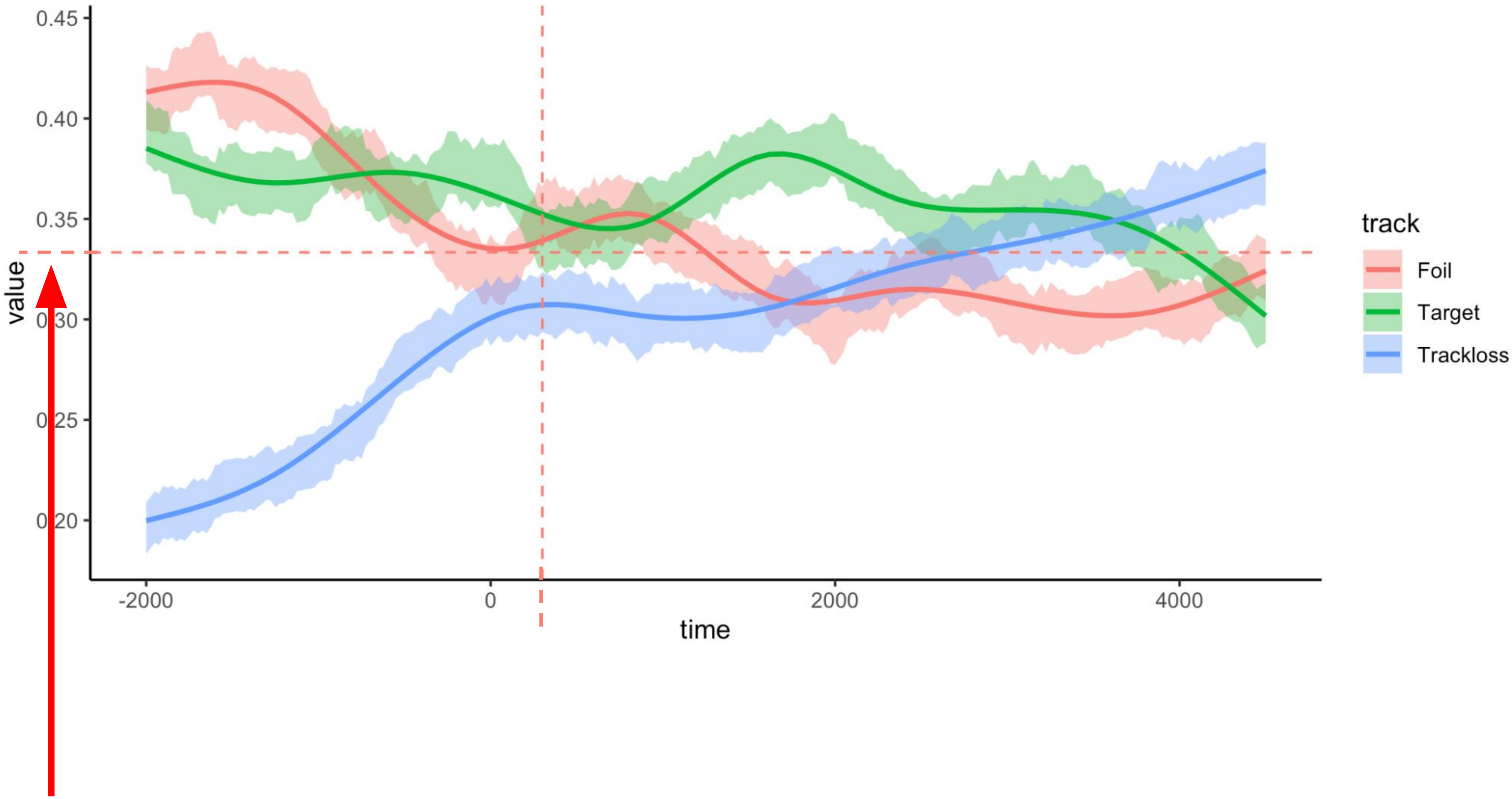
Infant eye-tracking studies with 14mo Korean infants



Infants' gaze patterns

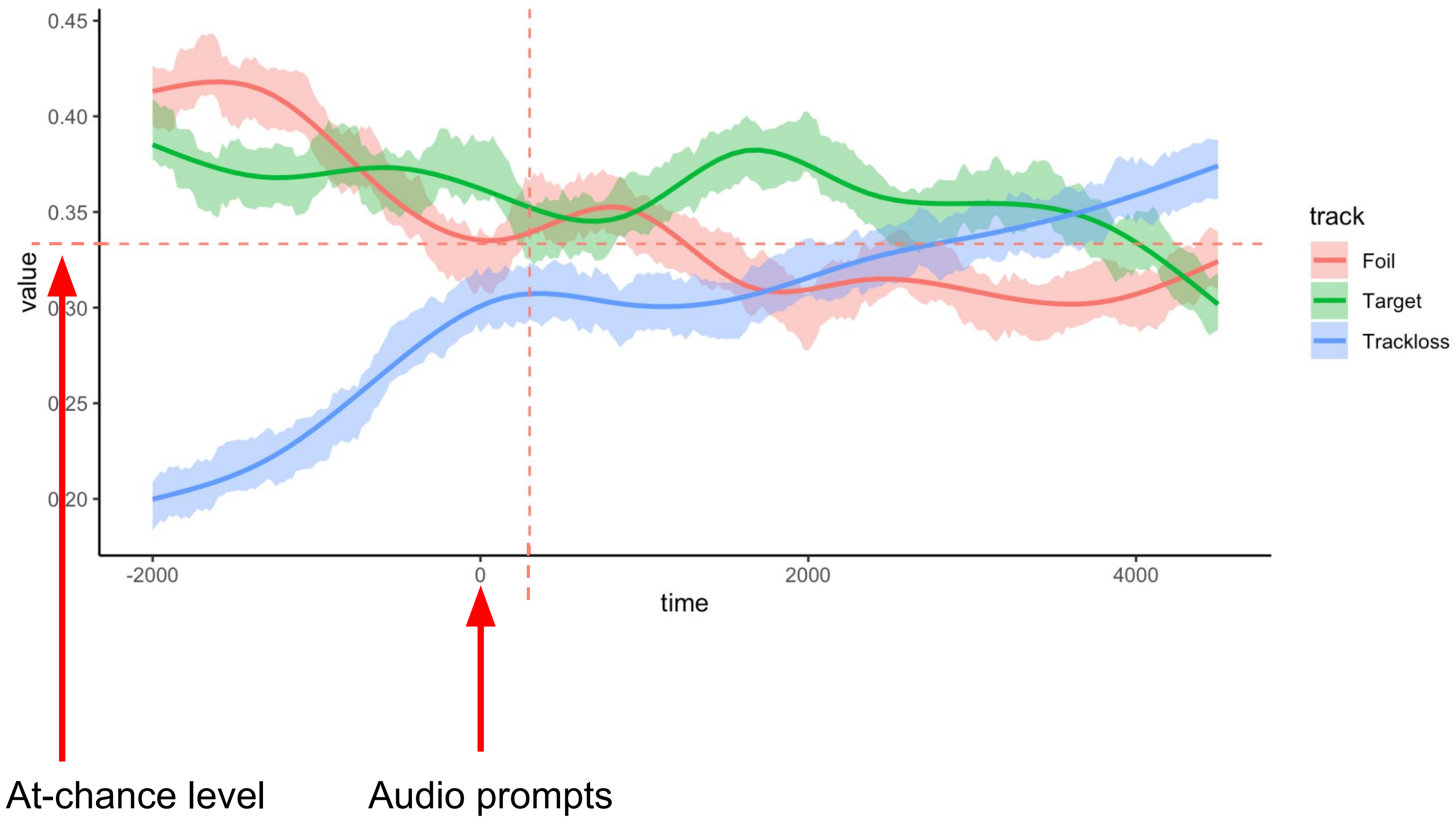


Infants' gaze patterns

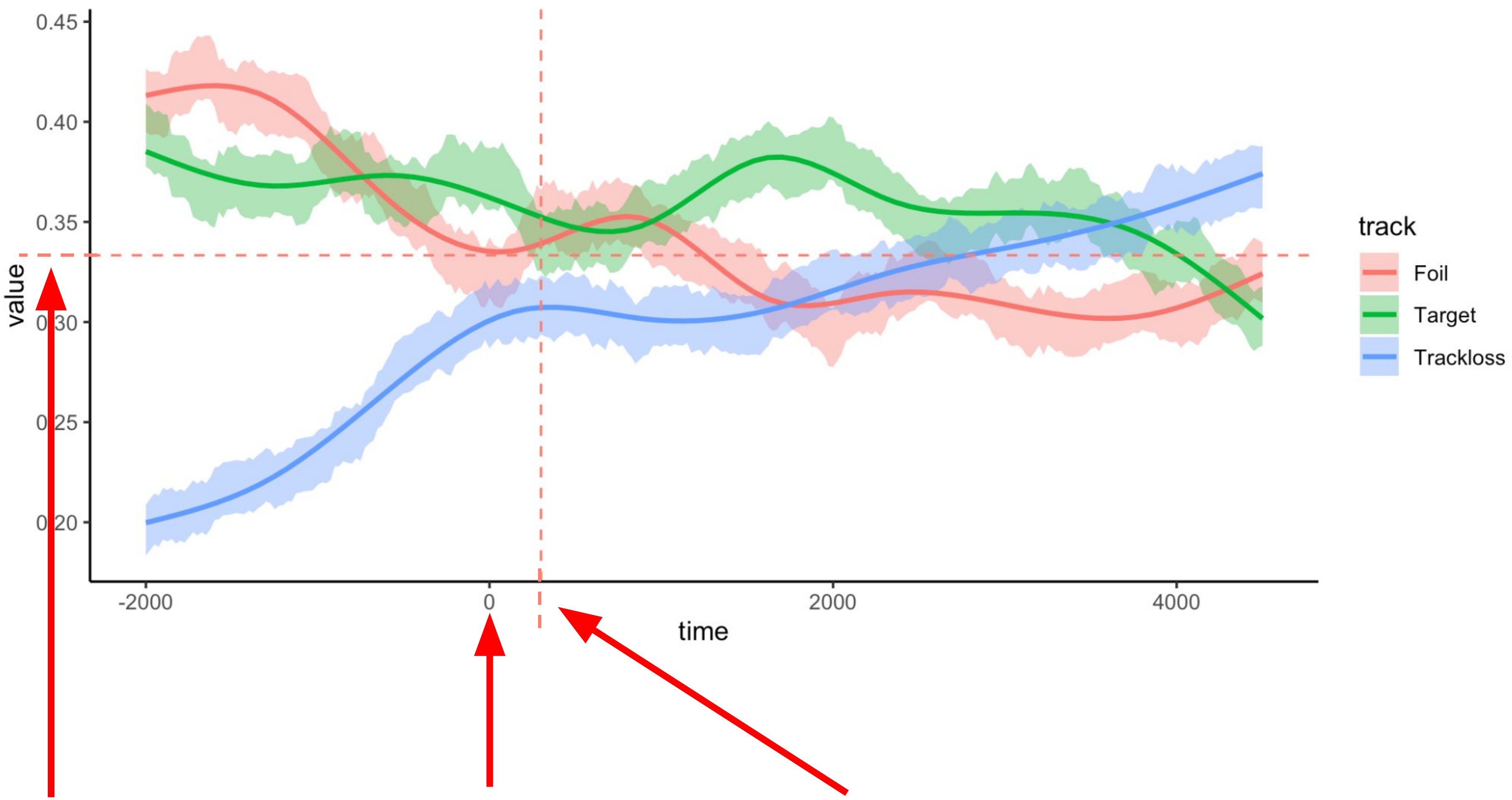


At-chance level

Infants' gaze patterns

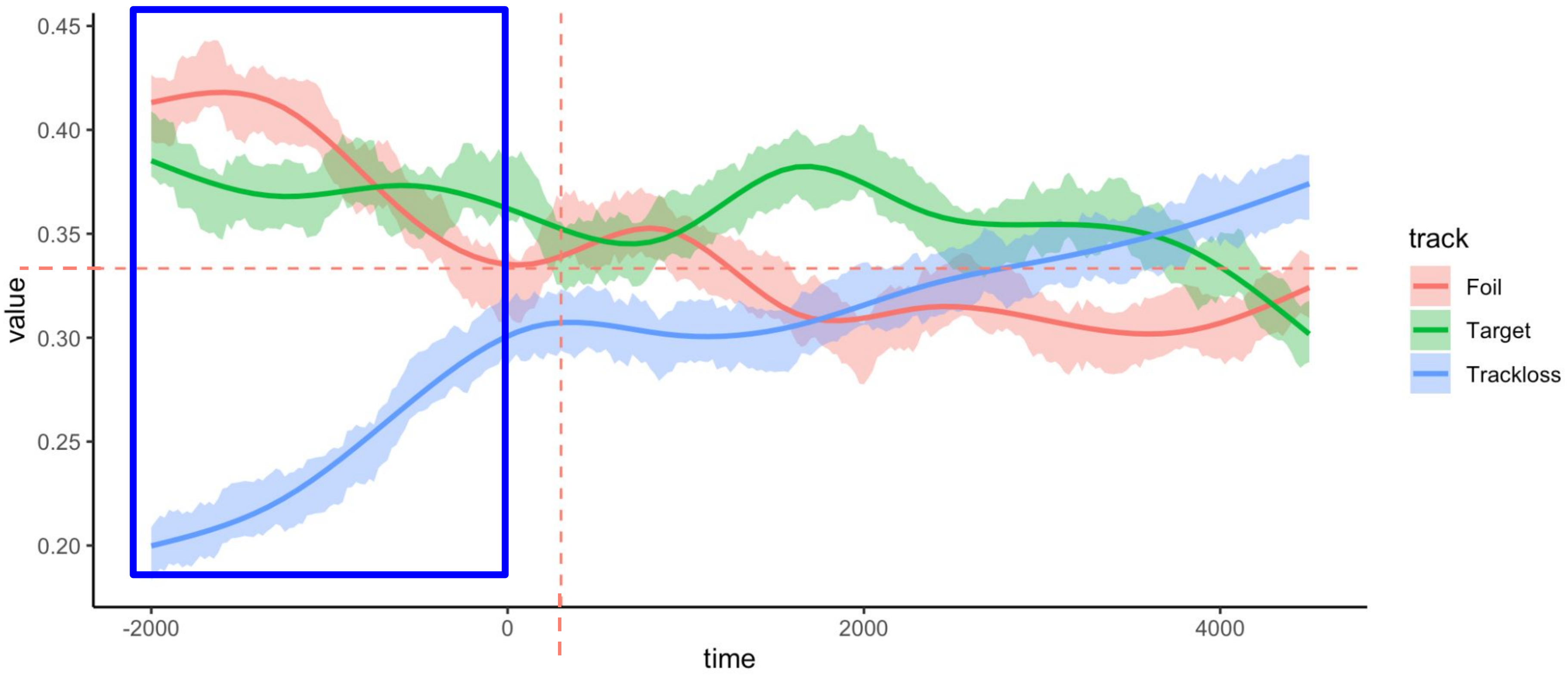


Infants' gaze patterns



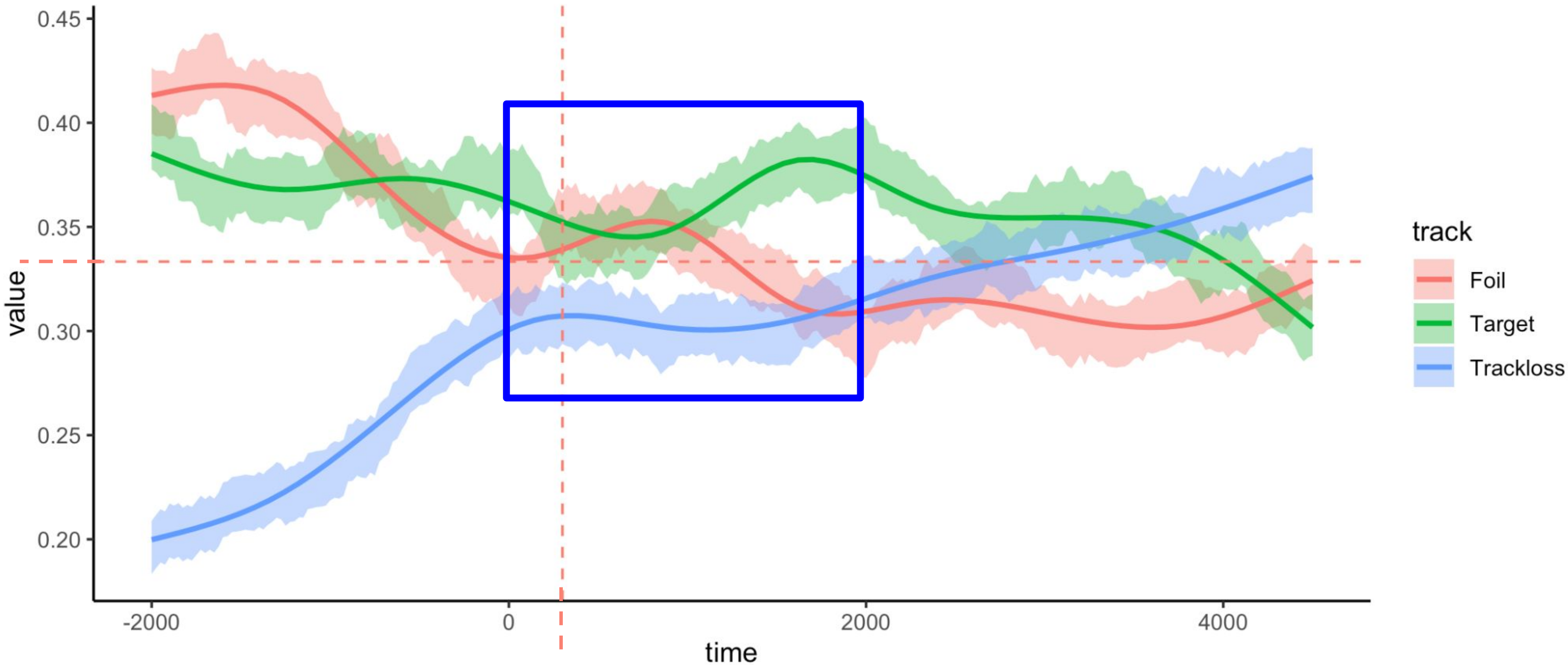
At-chance level   Audio prompts   ~300-400ms

Infants' gaze patterns



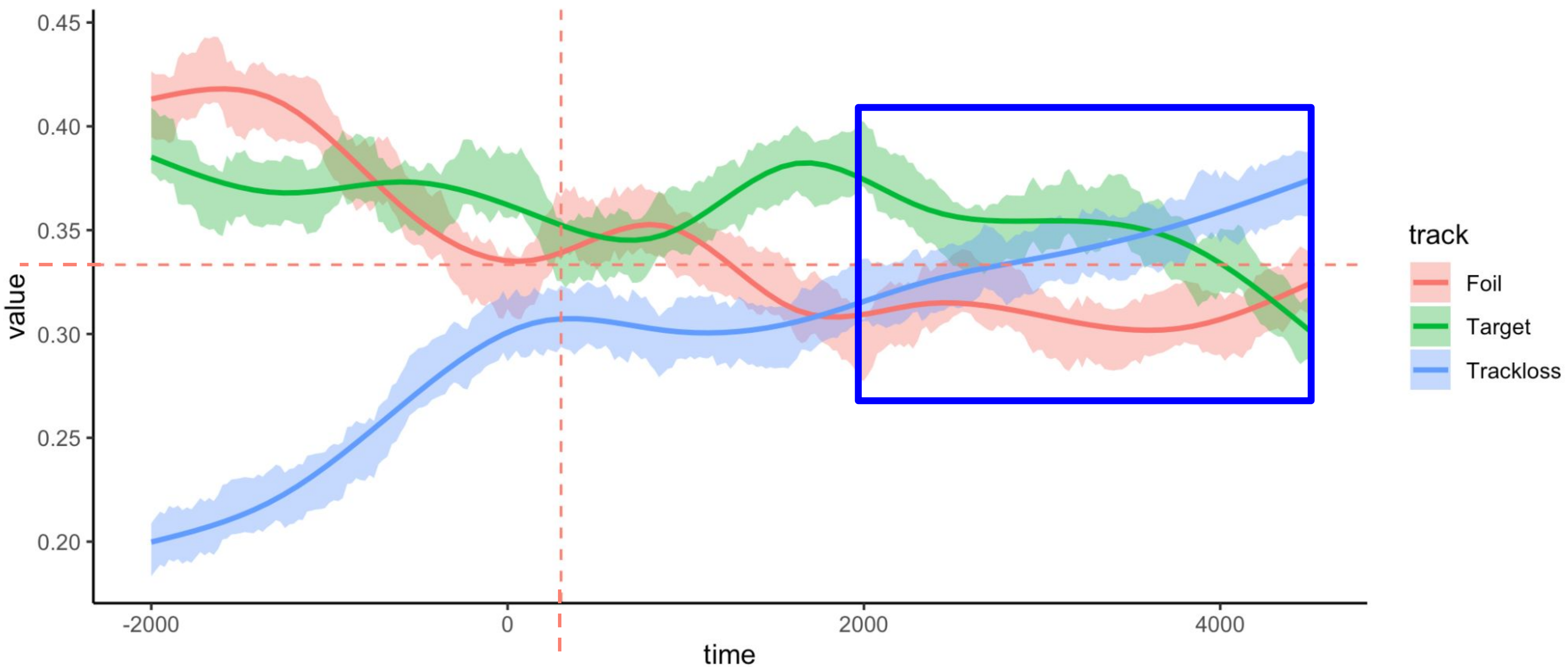
- Trackloss are below chance before audio played
- Attentive to the images:but increasingly inattentive

Infants' gaze patterns



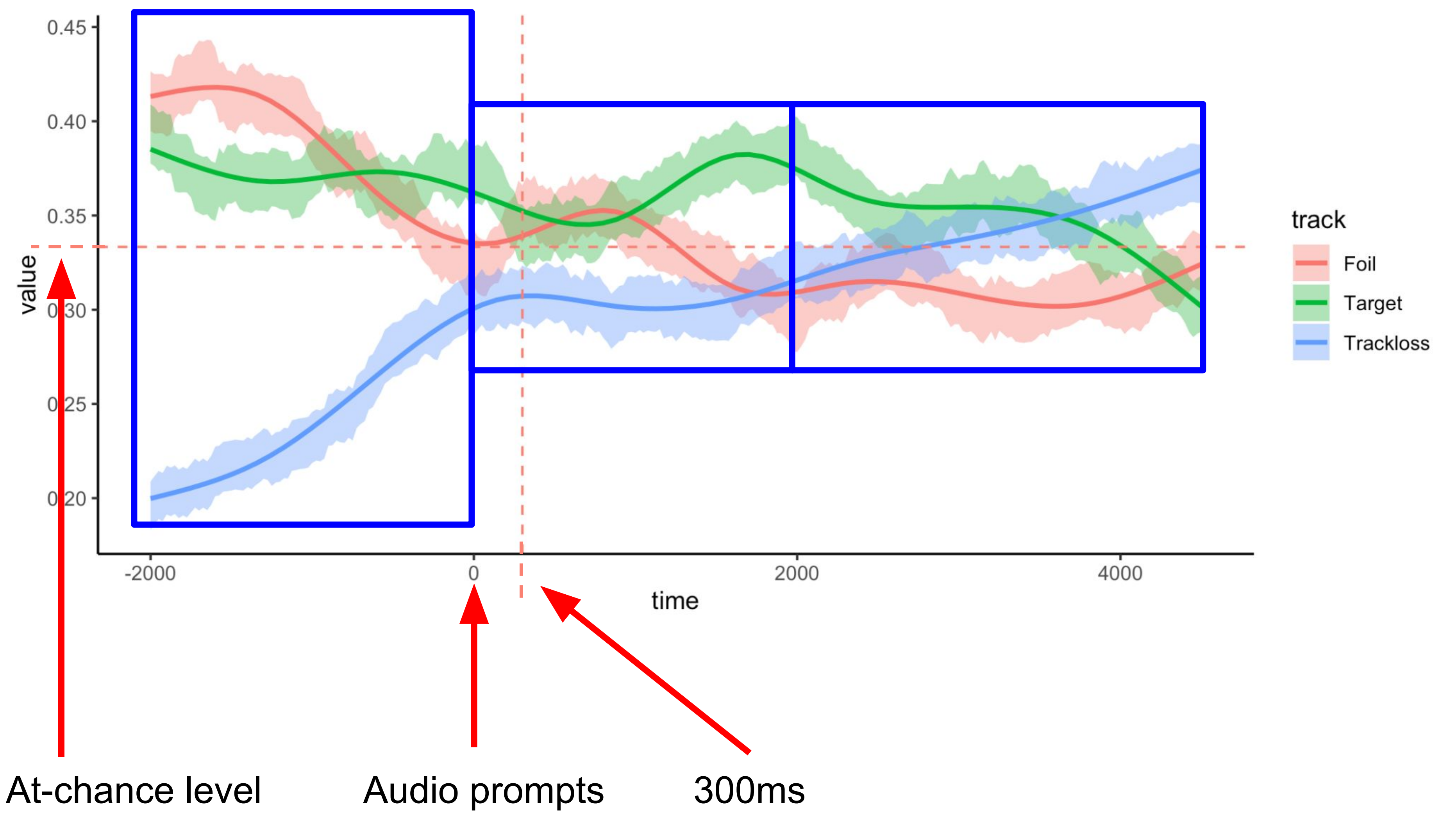
- Trackloss stabilises around 0 - 2000ms:
- Look to target increases

Infants' gaze patterns

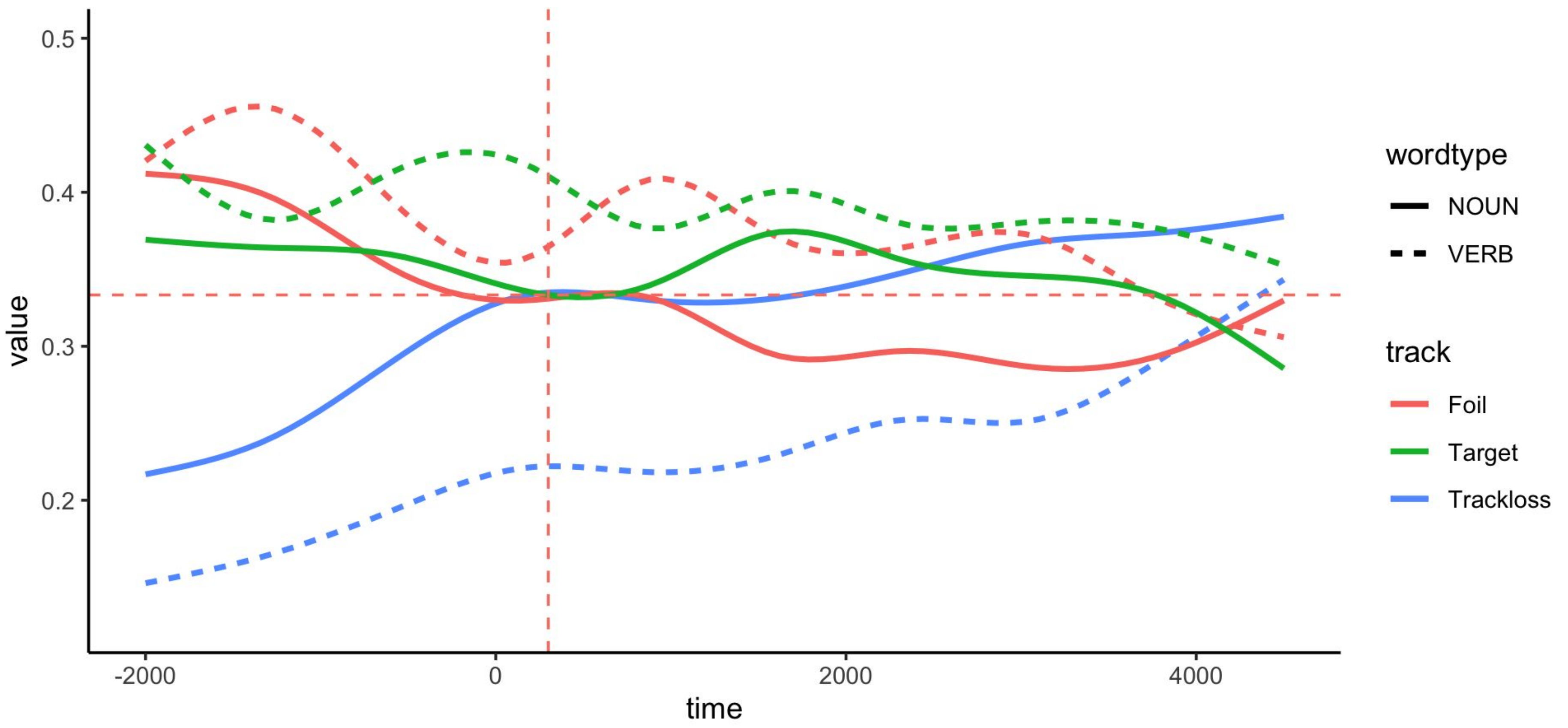


- Increasingly inattentive around 2000ms
- Trackloss generally above chance around 3000-4000ms

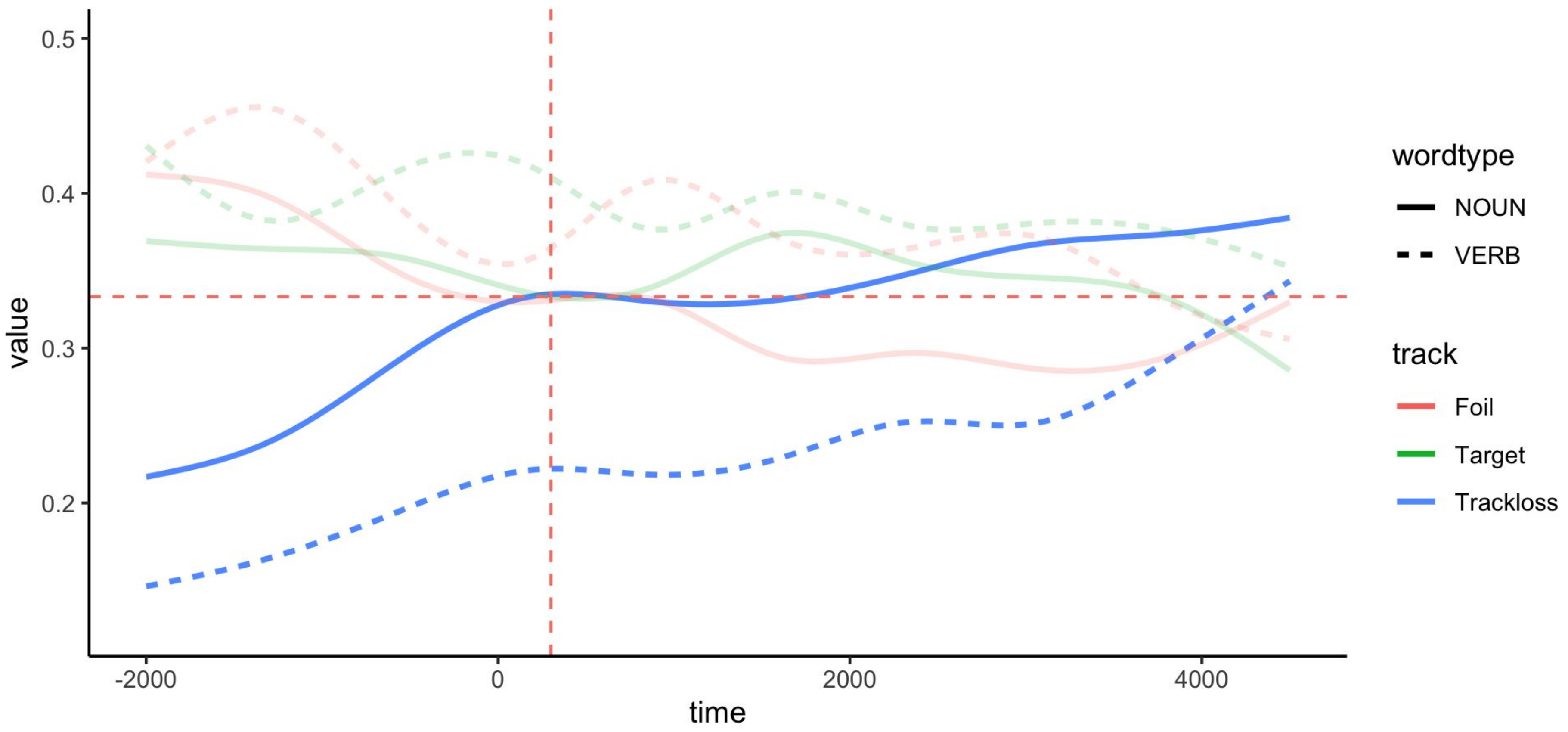
Infants' gaze patterns: overall



Infants' gaze patterns: word type

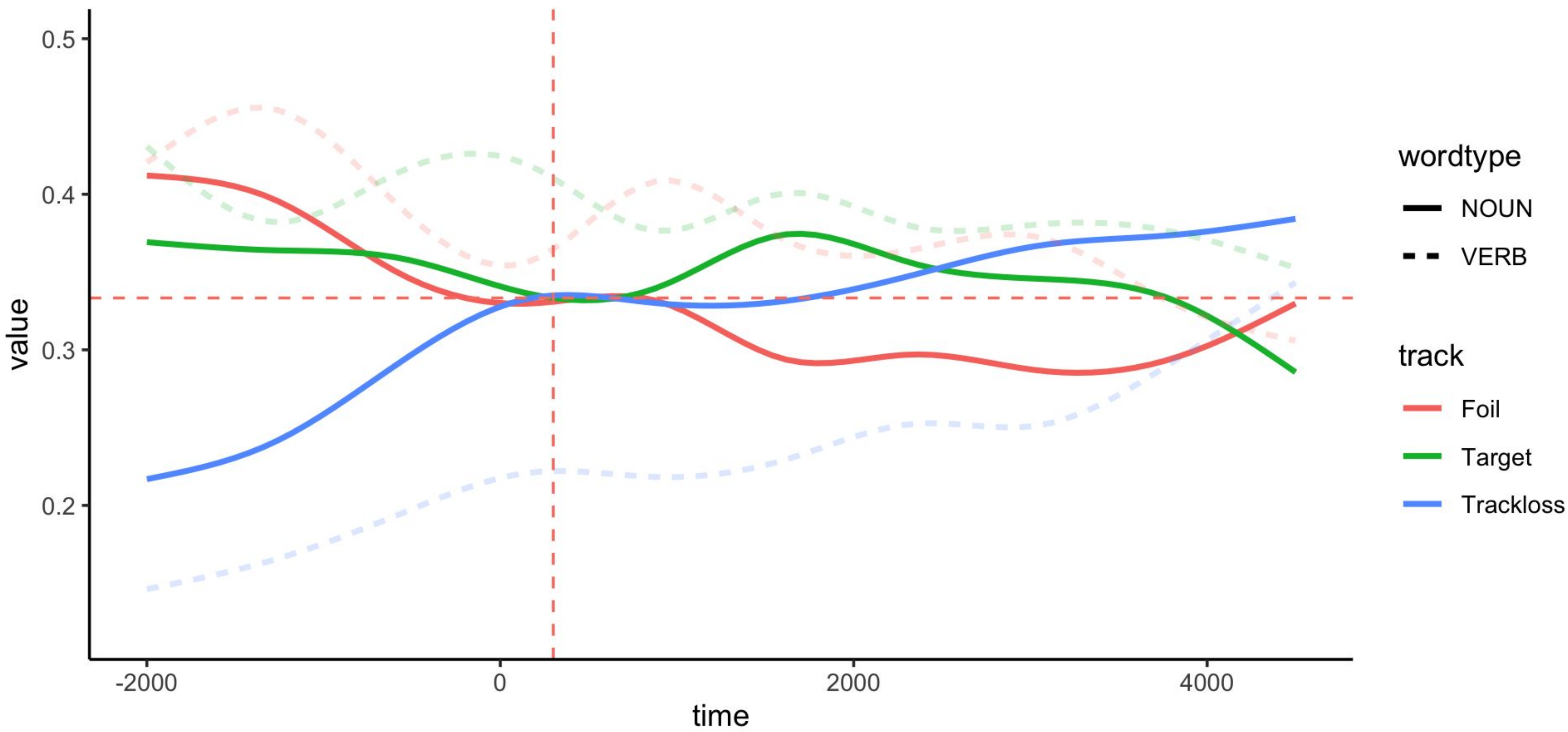


Infants' gaze patterns



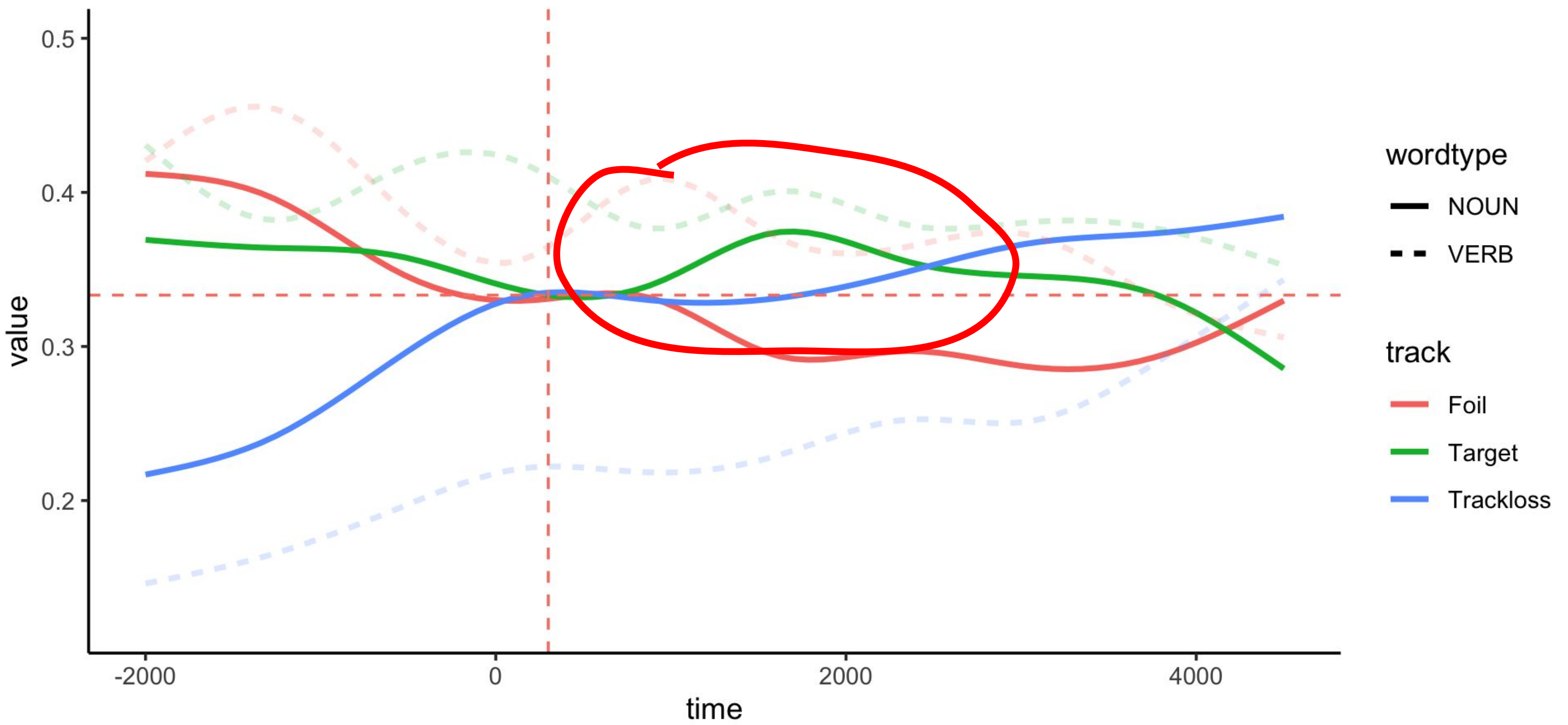
- More attentive to verb (lower trackloss) than noun
- Could be artifact due to surprisal or novelty (5 out of 20 word pairs are verb pairs)

Nouns



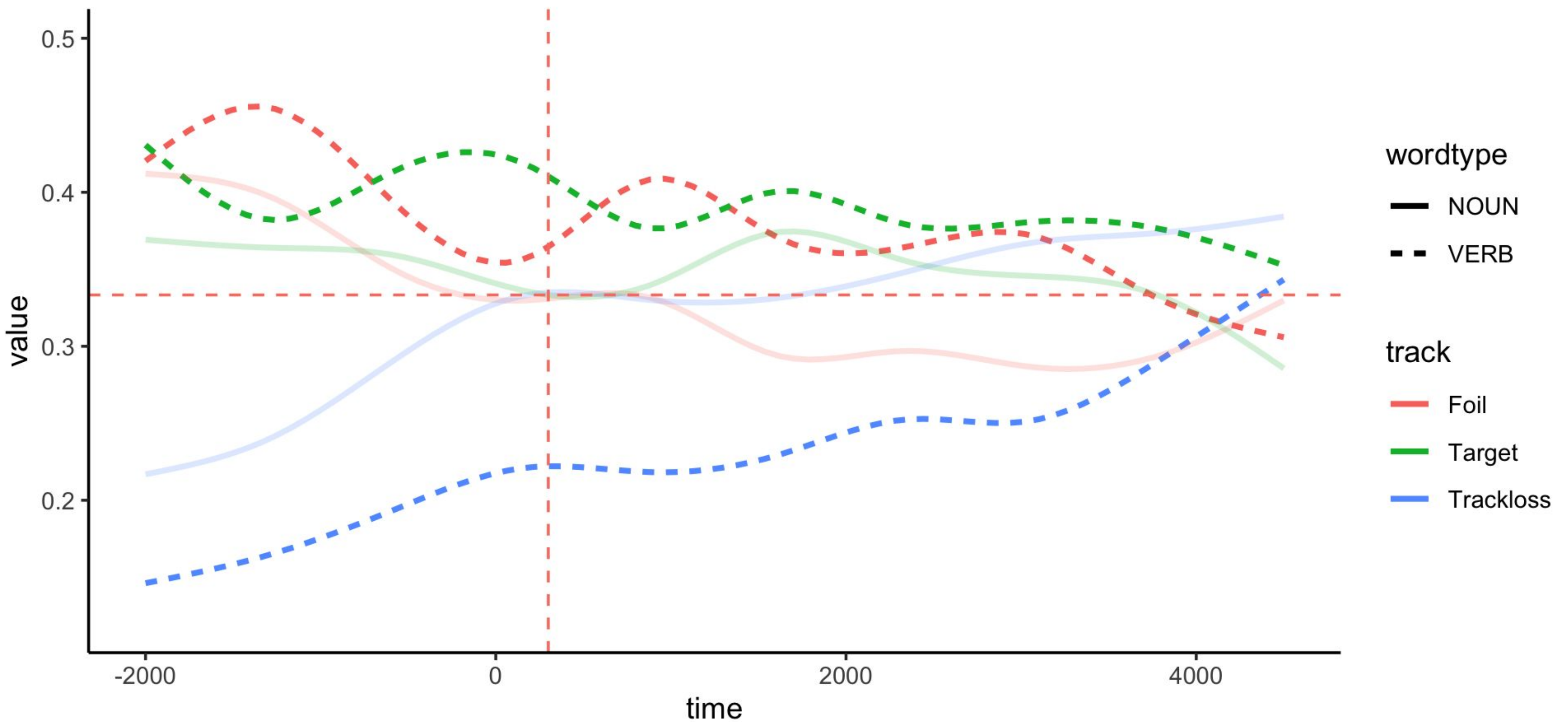
- Target generally above chance

Nouns



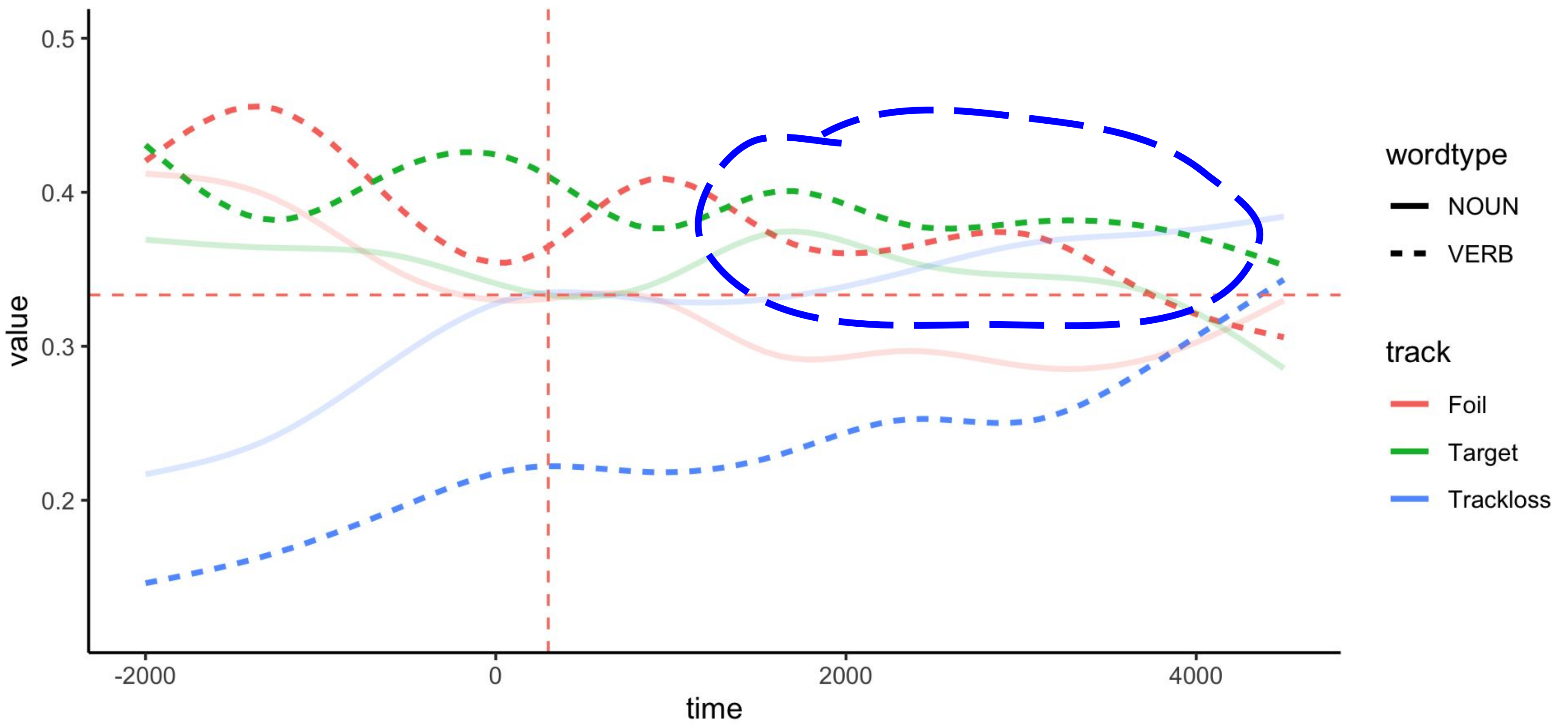
- Target generally above chance

Verbs



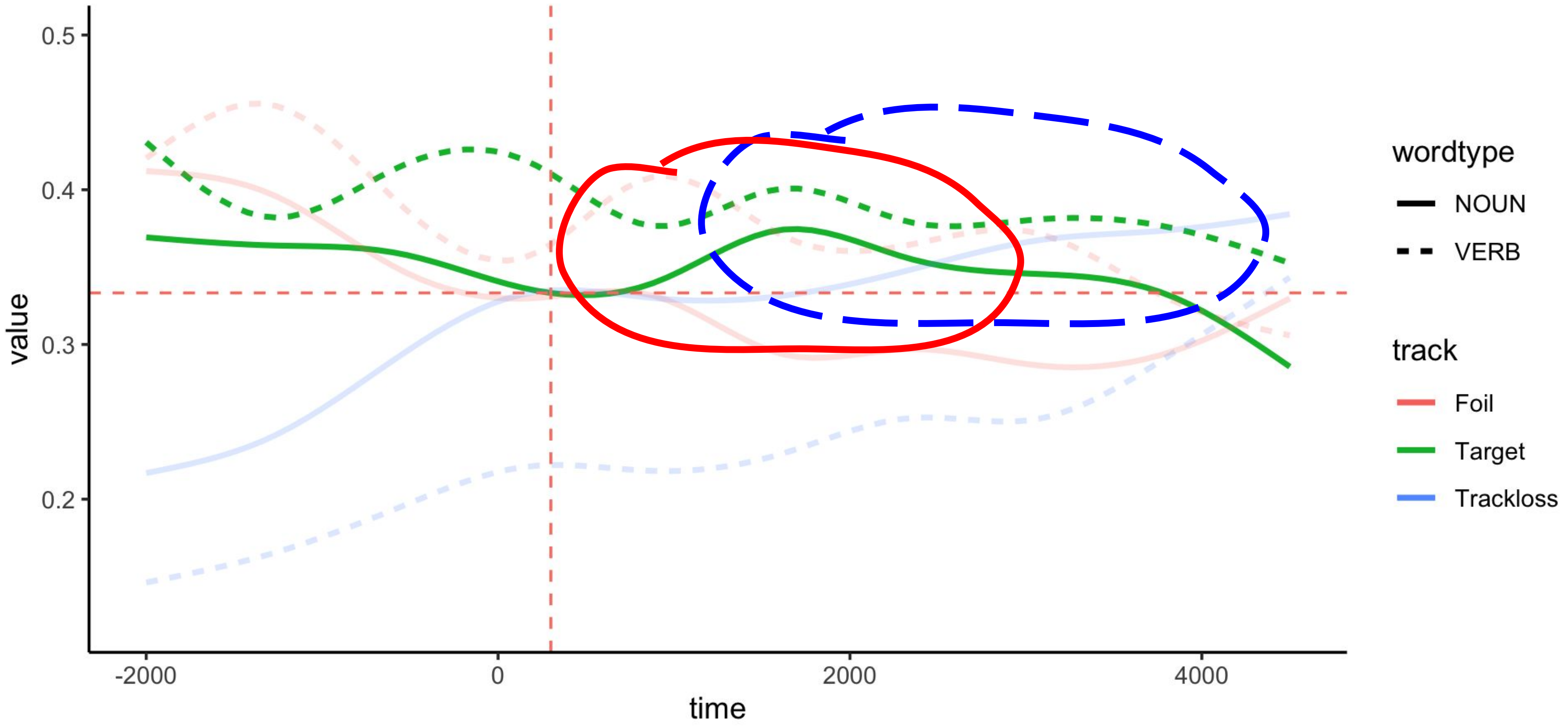
- Target generally above chance but more difficult to tell

Verbs

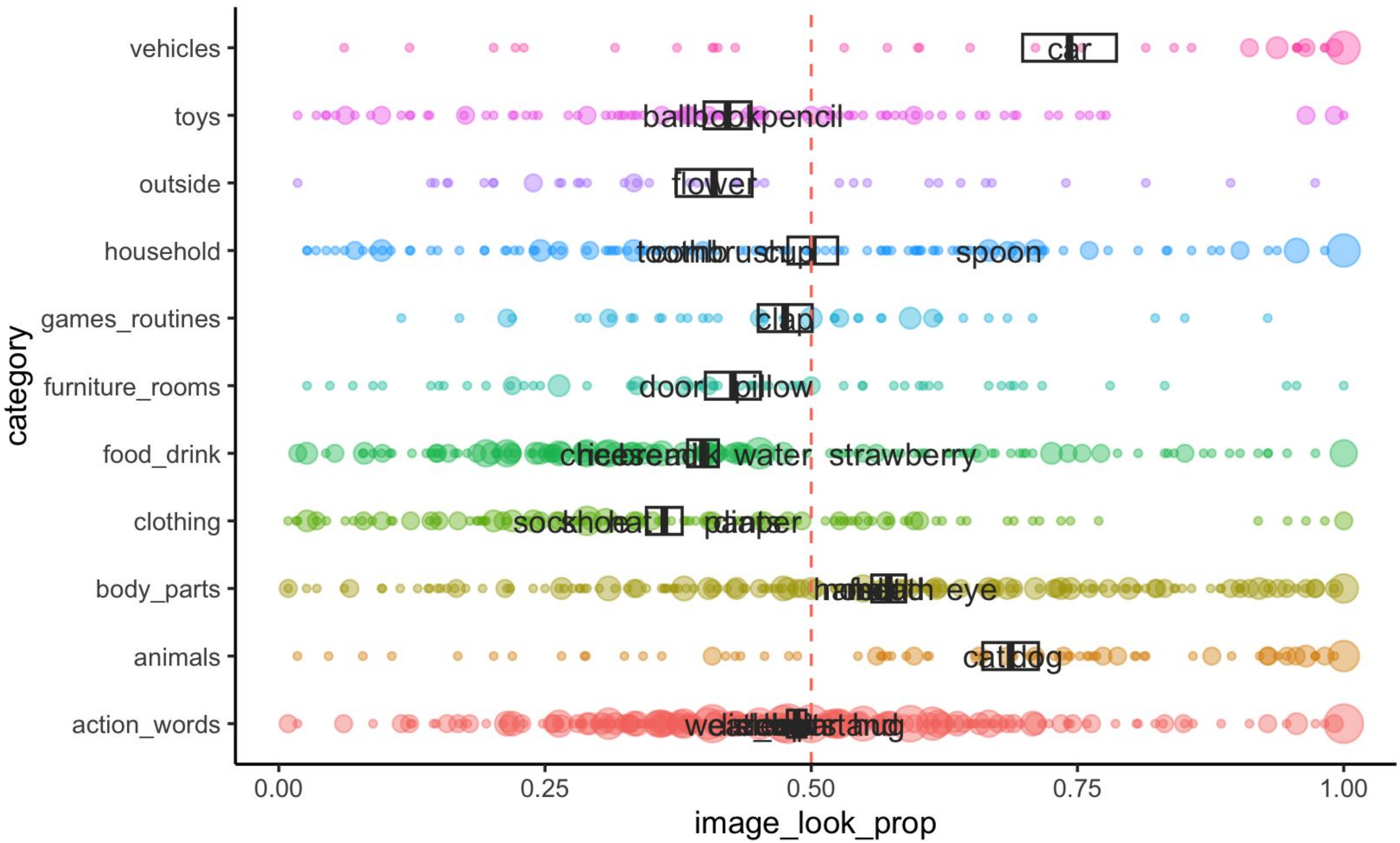


- Target generally above chance but more difficult to tell

Infants' gaze patterns



Perceptual biases



before audio prompt

Findings patterns in noisy data

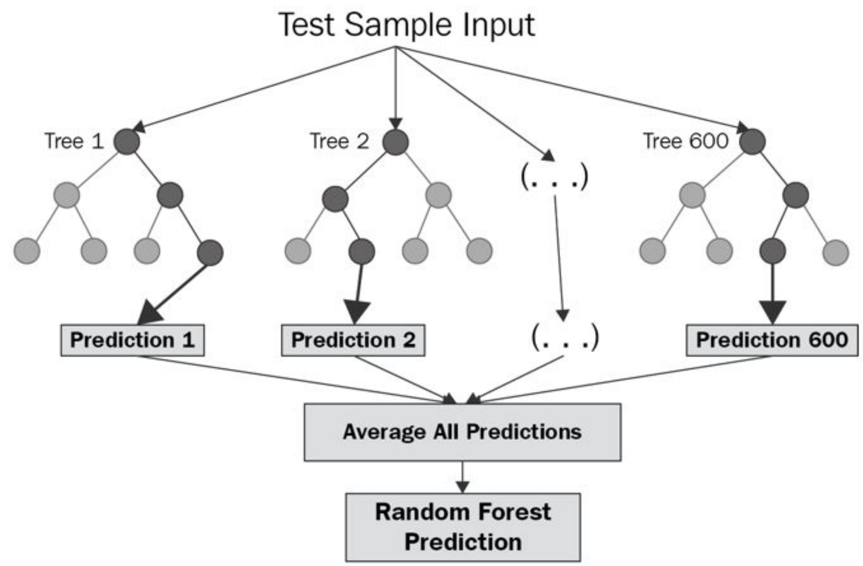
- Extract as much data as possible and quantifying gaze features
- Make gaze metric adjustments
- Rely on ground truth to infer relationship between gaze patterns and word recognition

Current Study

- Focusing on both raw and adjusted proportions of gaze directed toward the target during specific test windows.
- Utilising the Random Forest (RF)[1] method, we train the model to classify gaze patterns based on parental's report as the ground truth
- Trained using a dataset derived from eye-tracking tests on 25 Korean infants, approximately 14 months old

[1] Breiman, L. (2001)

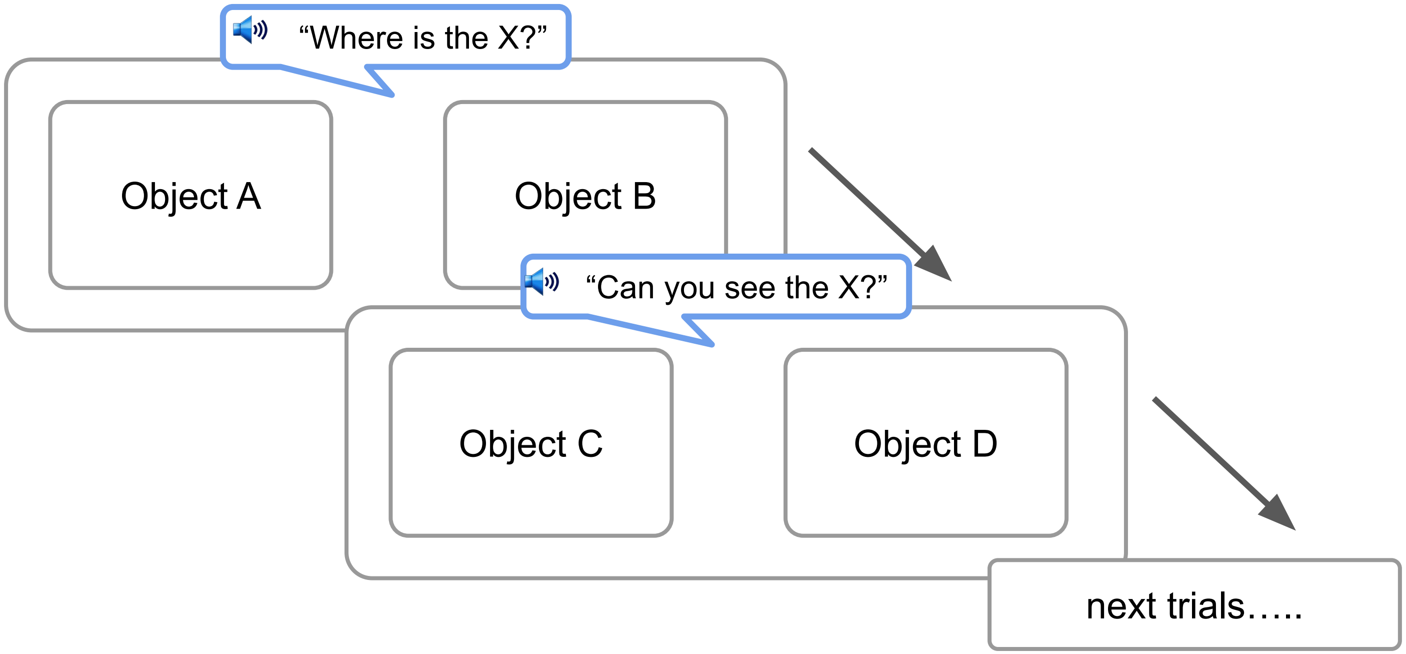
Random Forest Algorithm

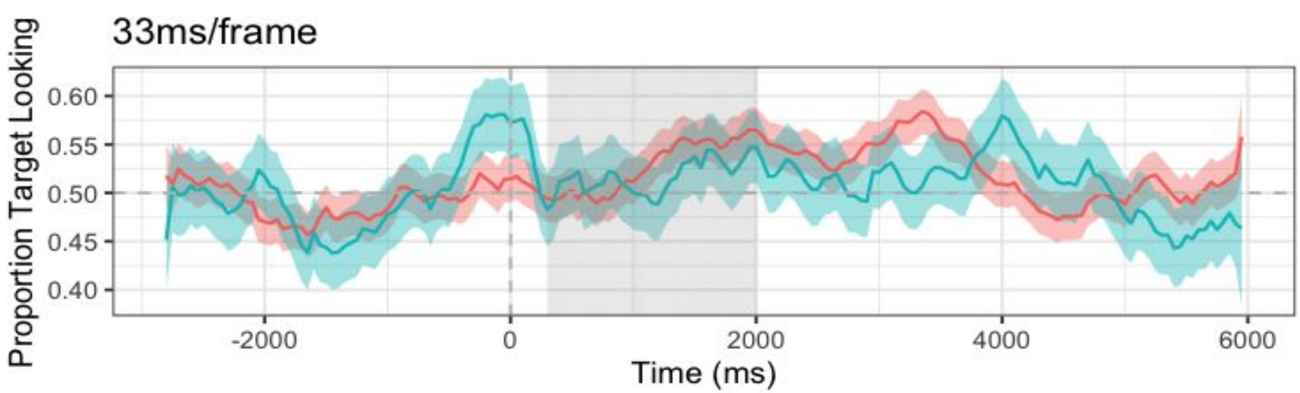


source.

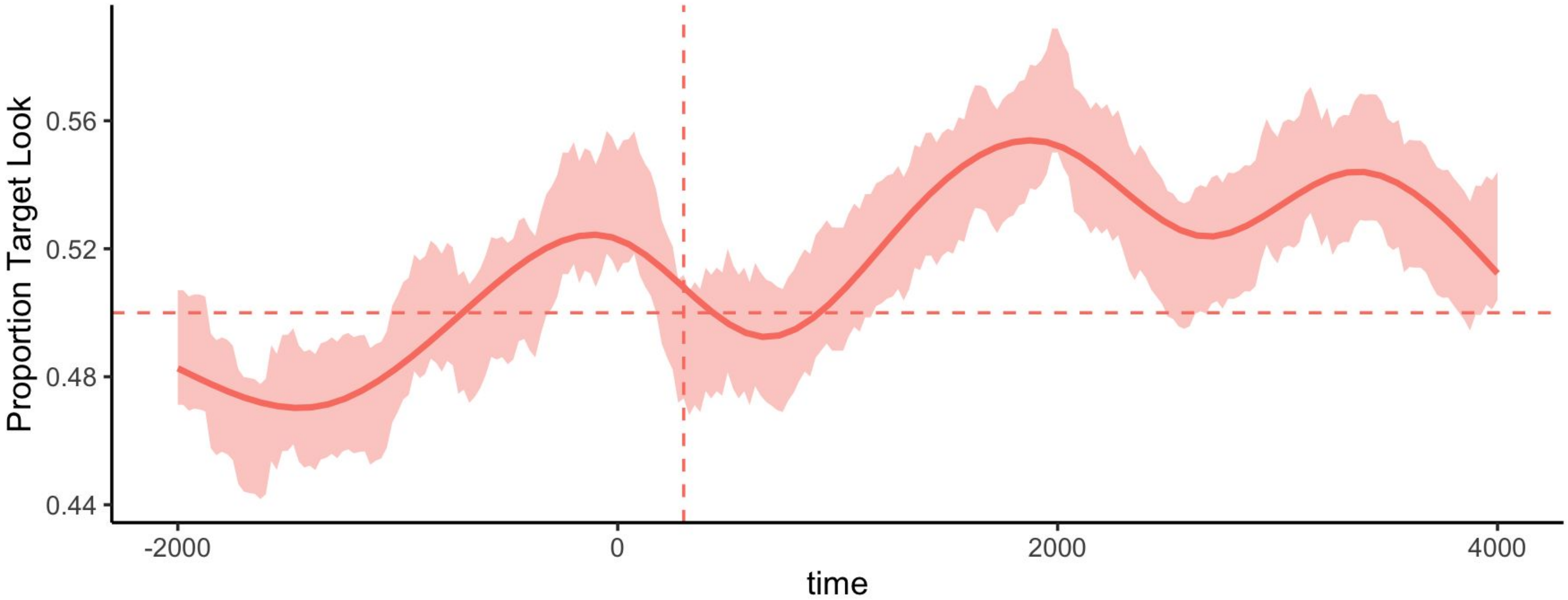
- An **Ensemble** Machine Learning Method
- Builds multiple **trees**, each using different **feature subsets**.
- Handles high-dimensional, complex data.
- We incorporate a wide range of gaze metrics for a more refined analysis of target recognition.

Training data from word recognition task

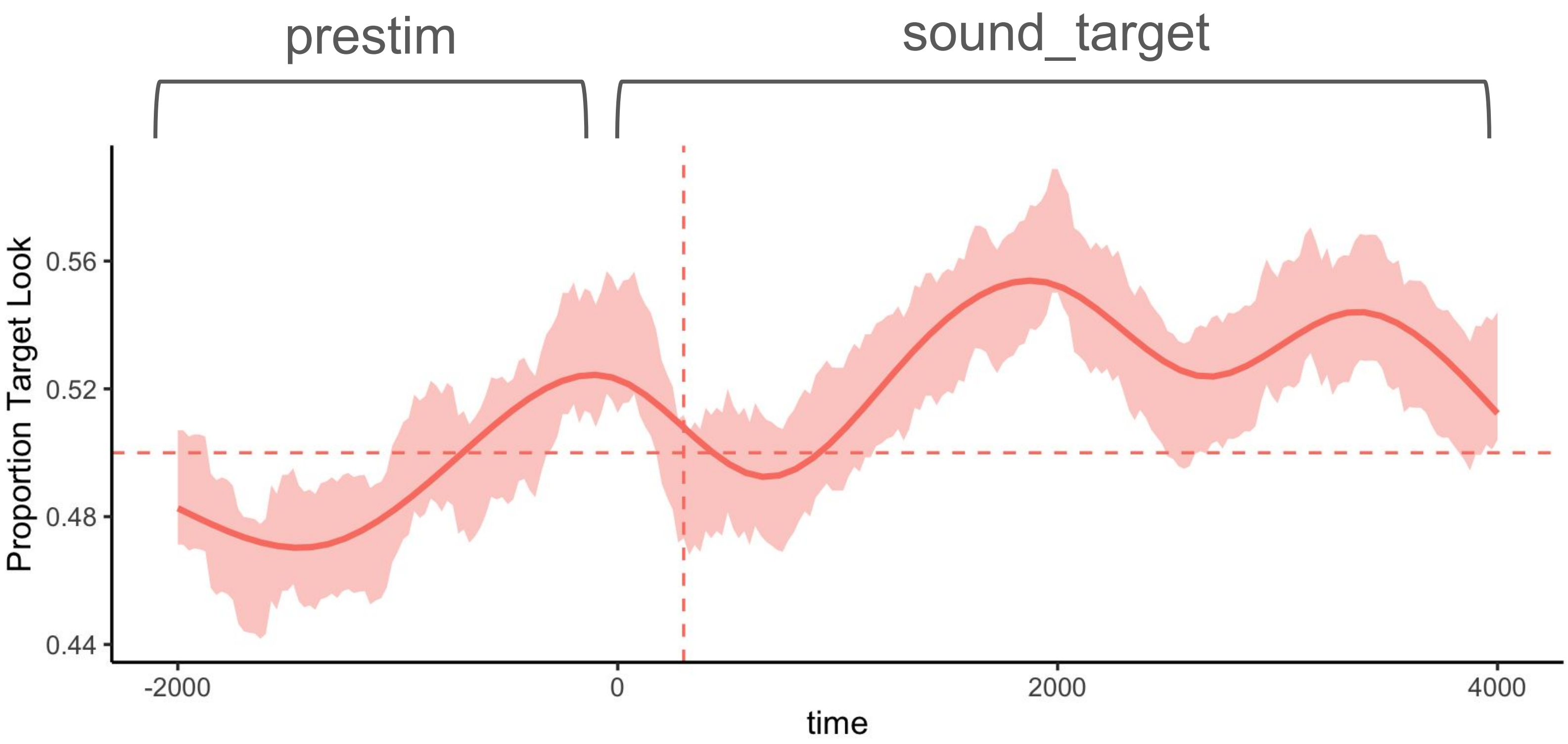




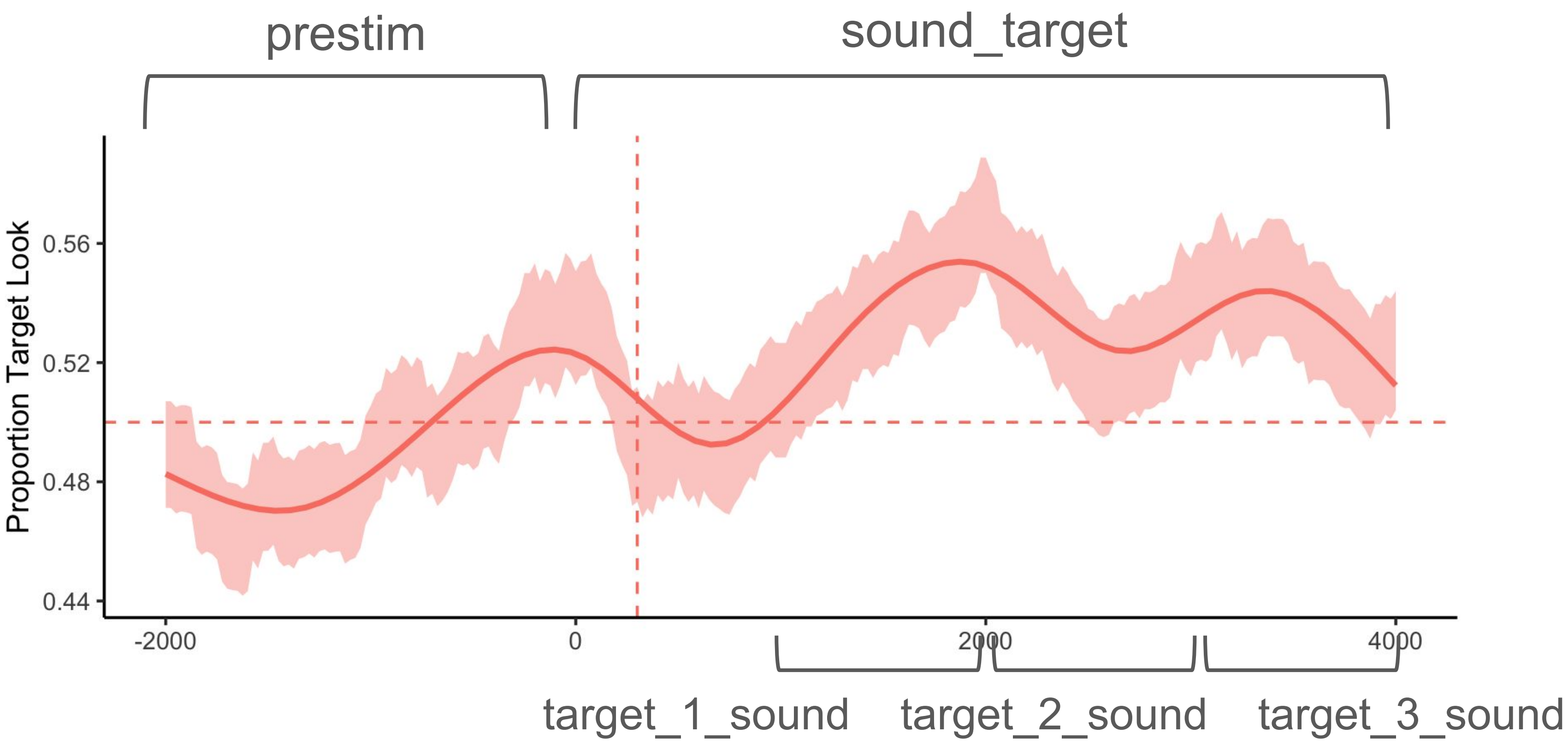
Feature Selections



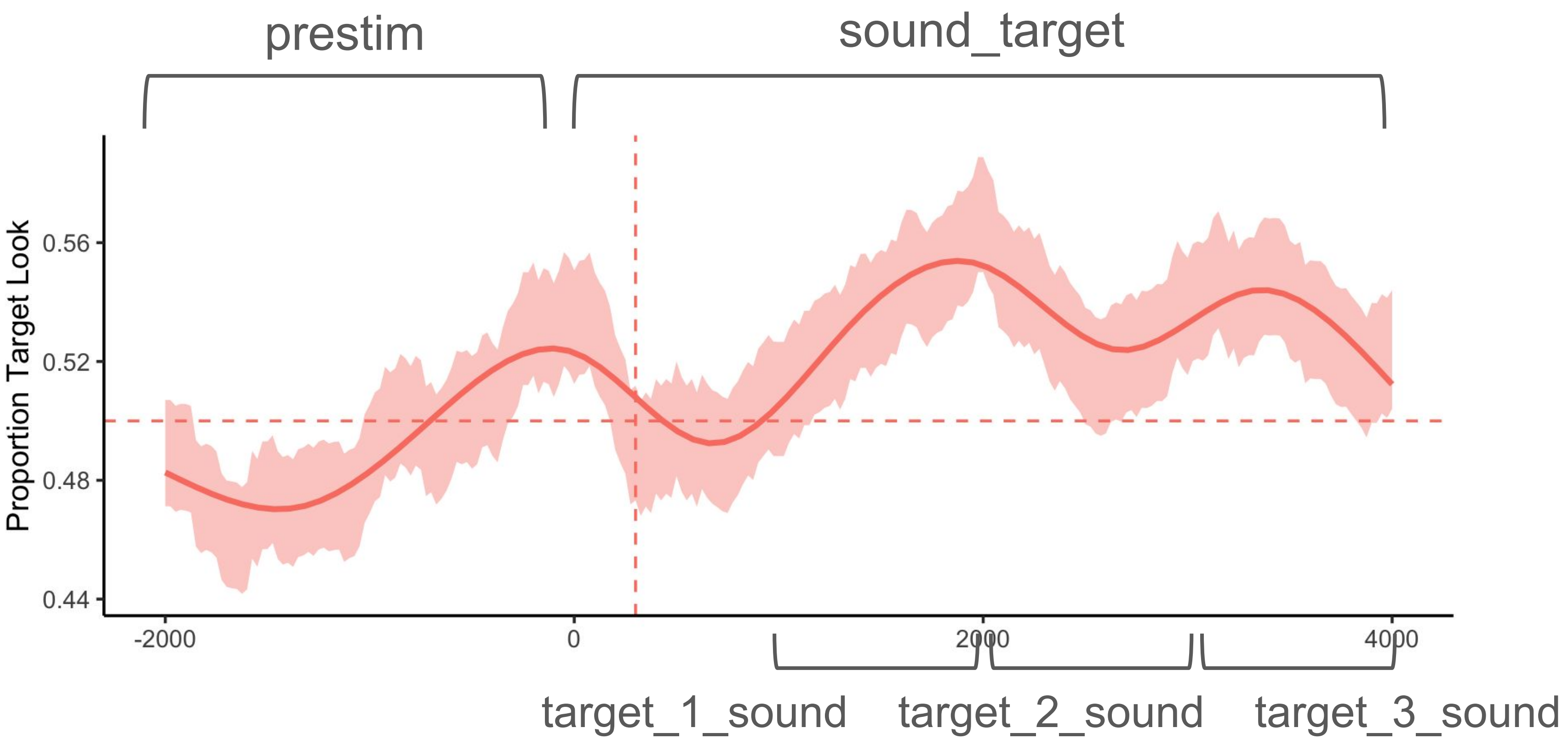
Feature Selections



Feature Selections

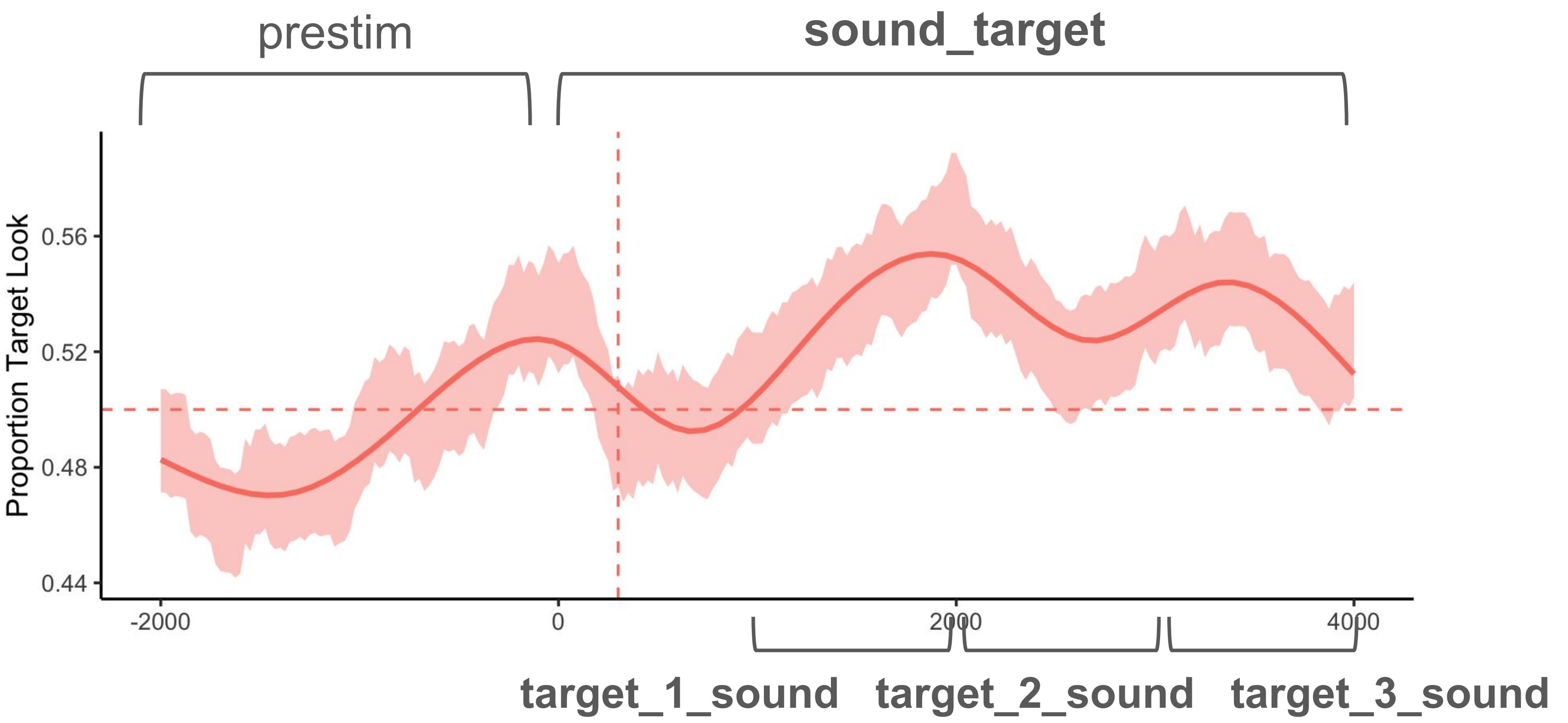


Feature Selections



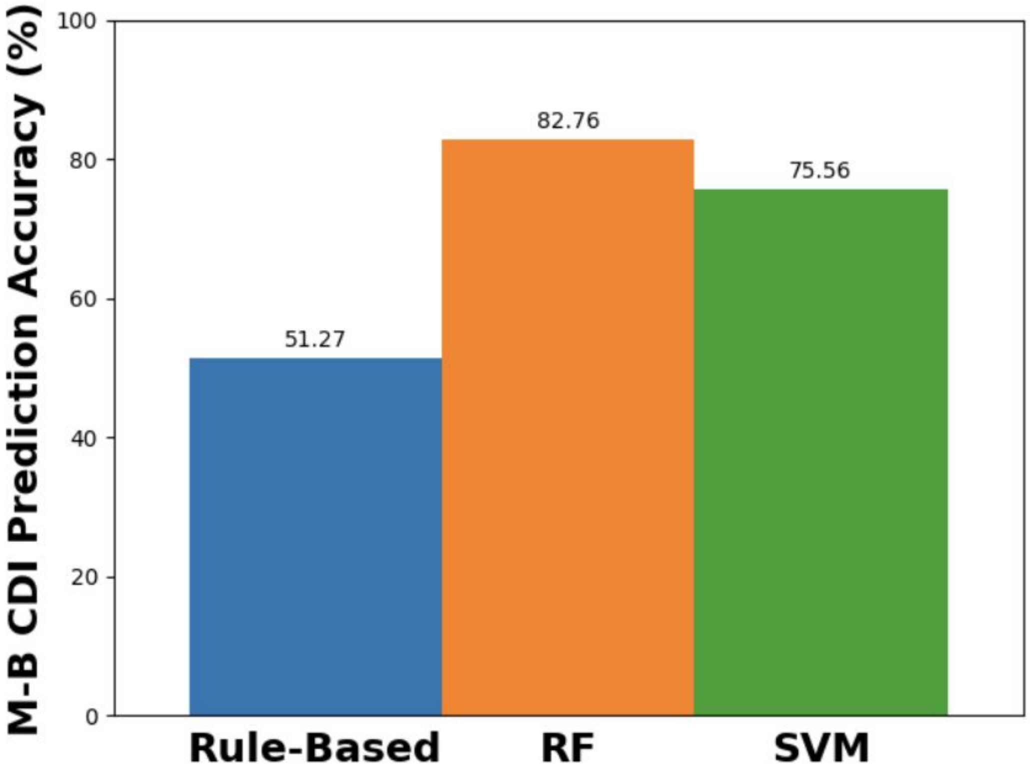
- targetdiff_sound_prestim = sound_target - prestim
- targetdiff_1_sound_prestim = target_1_sound - prestim
- targetdiff_2_sound_prestim = target_2_sound - prestim
- targetdiff_3_sound_prestim = target_3_sound - prestim

Feature Selections

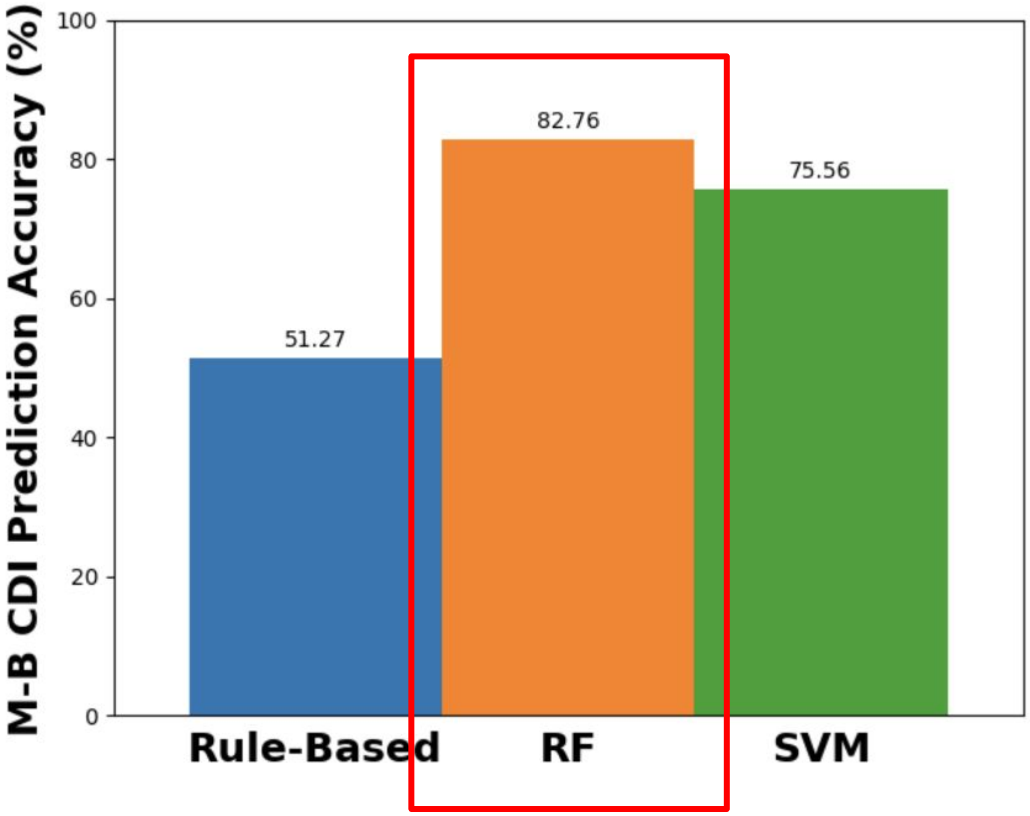


- **targetdiff_sound_prestim** = sound_target - prestim
- **targetdiff_1_sound_prestim** = target_1_sound - prestim
- **targetdiff_2_sound_prestim** = target_2_sound - prestim
- **targetdiff_3_sound_prestim** = target_3_sound - prestim

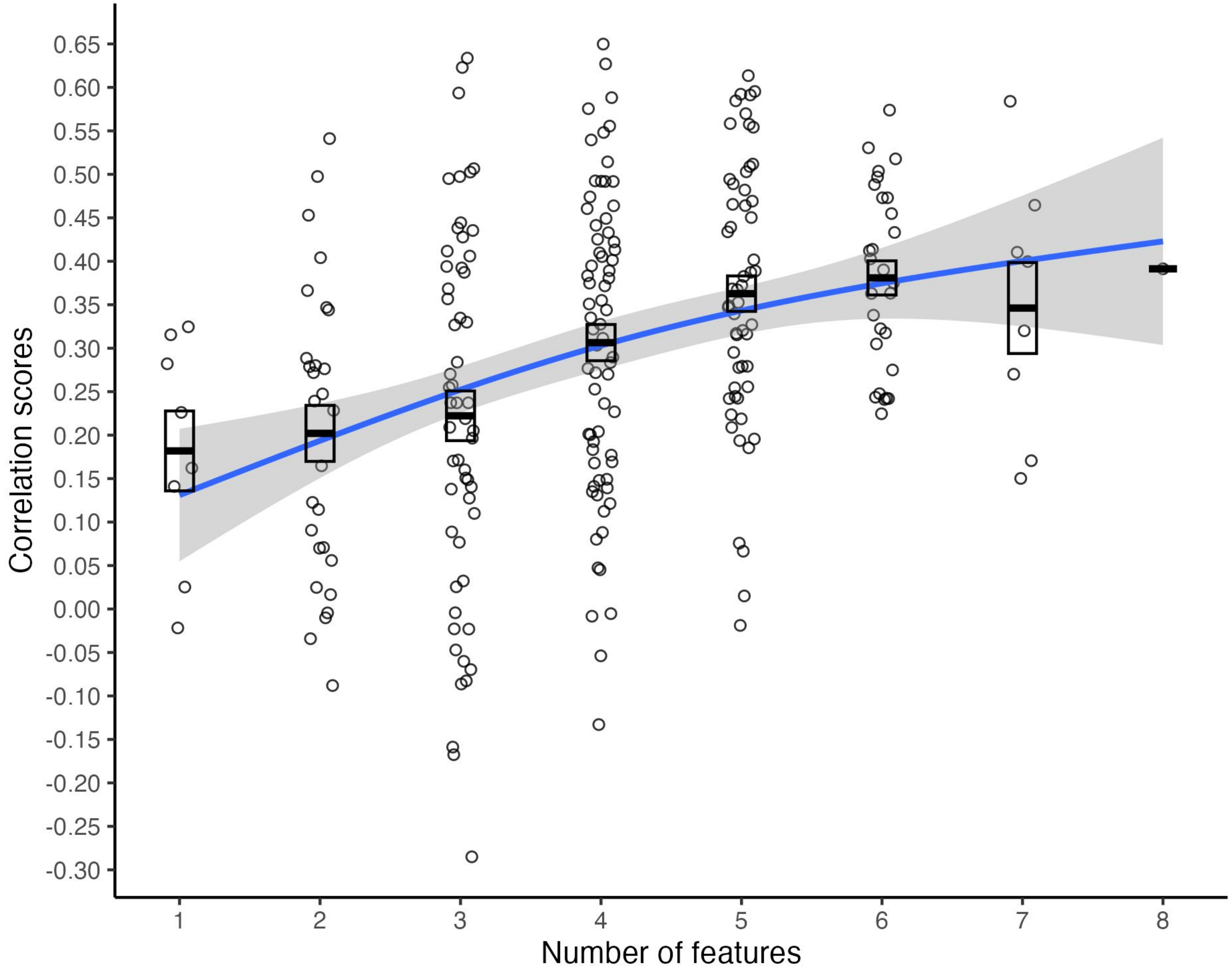
Model validation



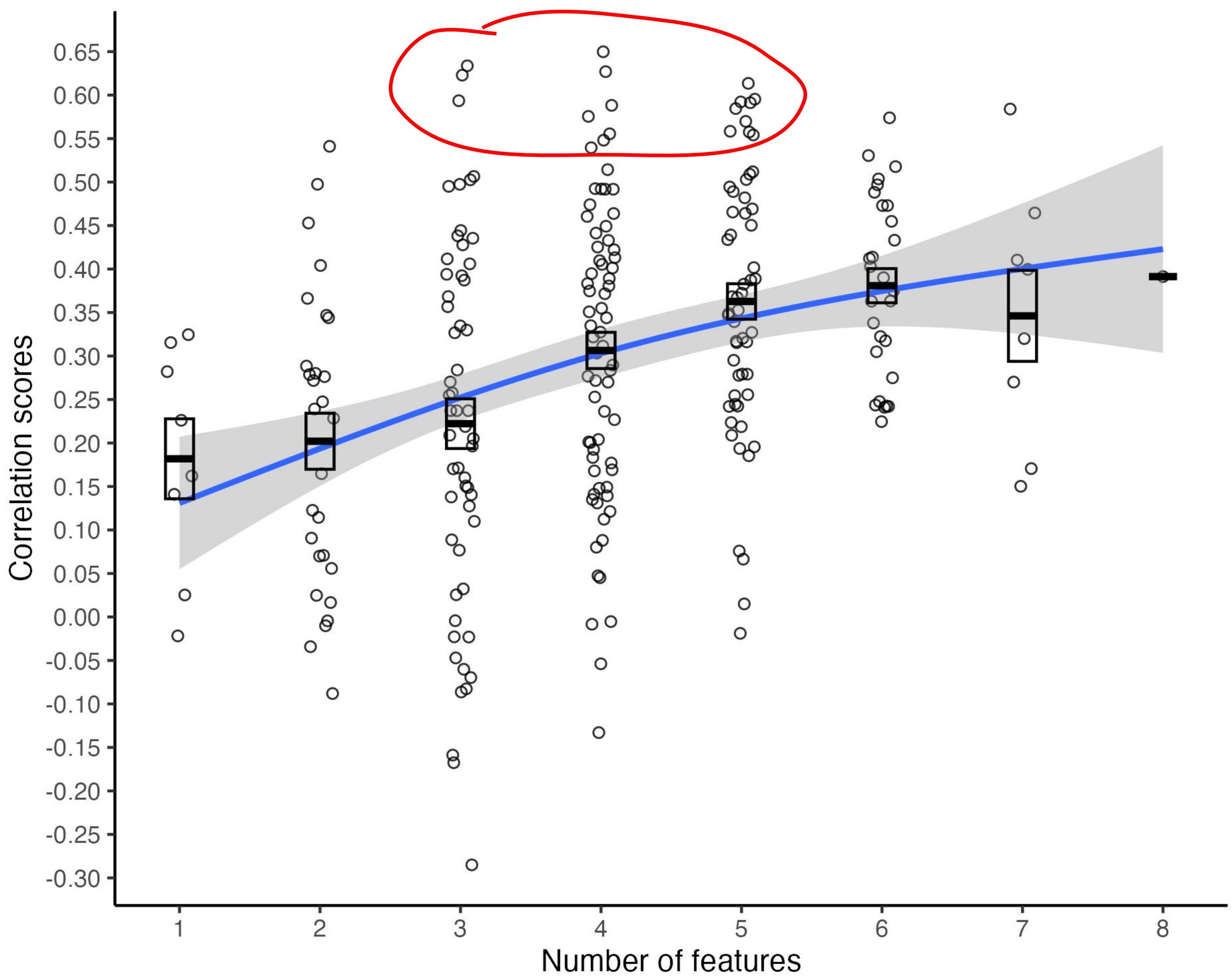
Model validation



Optimal number of features



Optimal number of features



Optimal feature combinations

Best ten combinations:

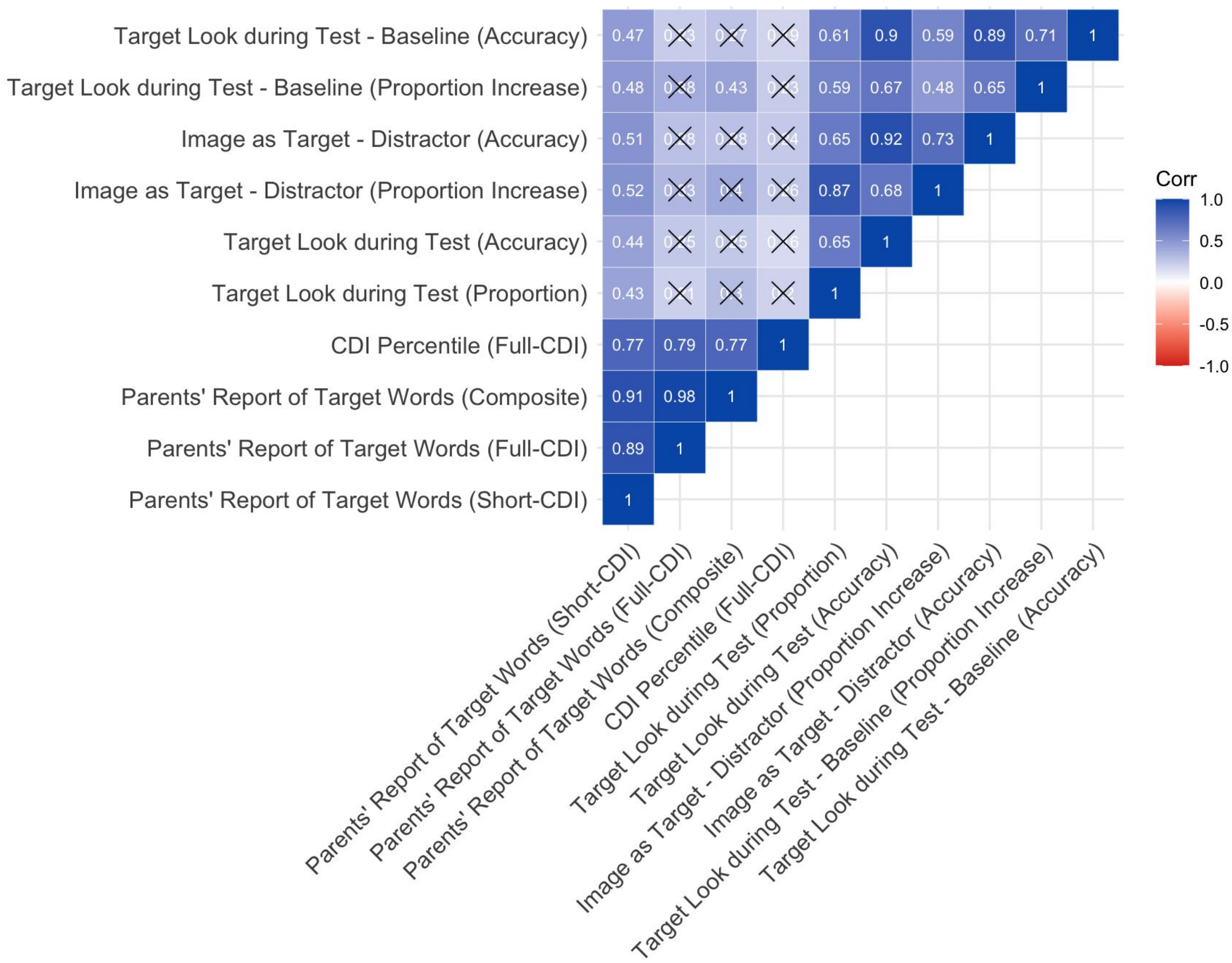
Feature combination	Number of features	Correlation
1_sound:2_sound:diff_1:diff_3	4	.650
2_sound:diff_1:diff_3	3	.634
1_sound:3_sound:diff_1:diff_3	4	.627
1_sound:diff_1:diff_3	3	.623
1_sound:2_sound:diff_1:diff_2:diff_3	5	.614
2_sound:diff:diff_1:diff_2:diff_3	5	.595
1_sound:2_sound:diff_3	3	.594
sound:2_sound:diff_1:diff_2:diff_3	5	.592
1_sound:3_sound:diff_1:diff_2:diff_3	5	.591
2_sound:diff_1:diff_2:diff_3	4	.588

Optimal feature combinations

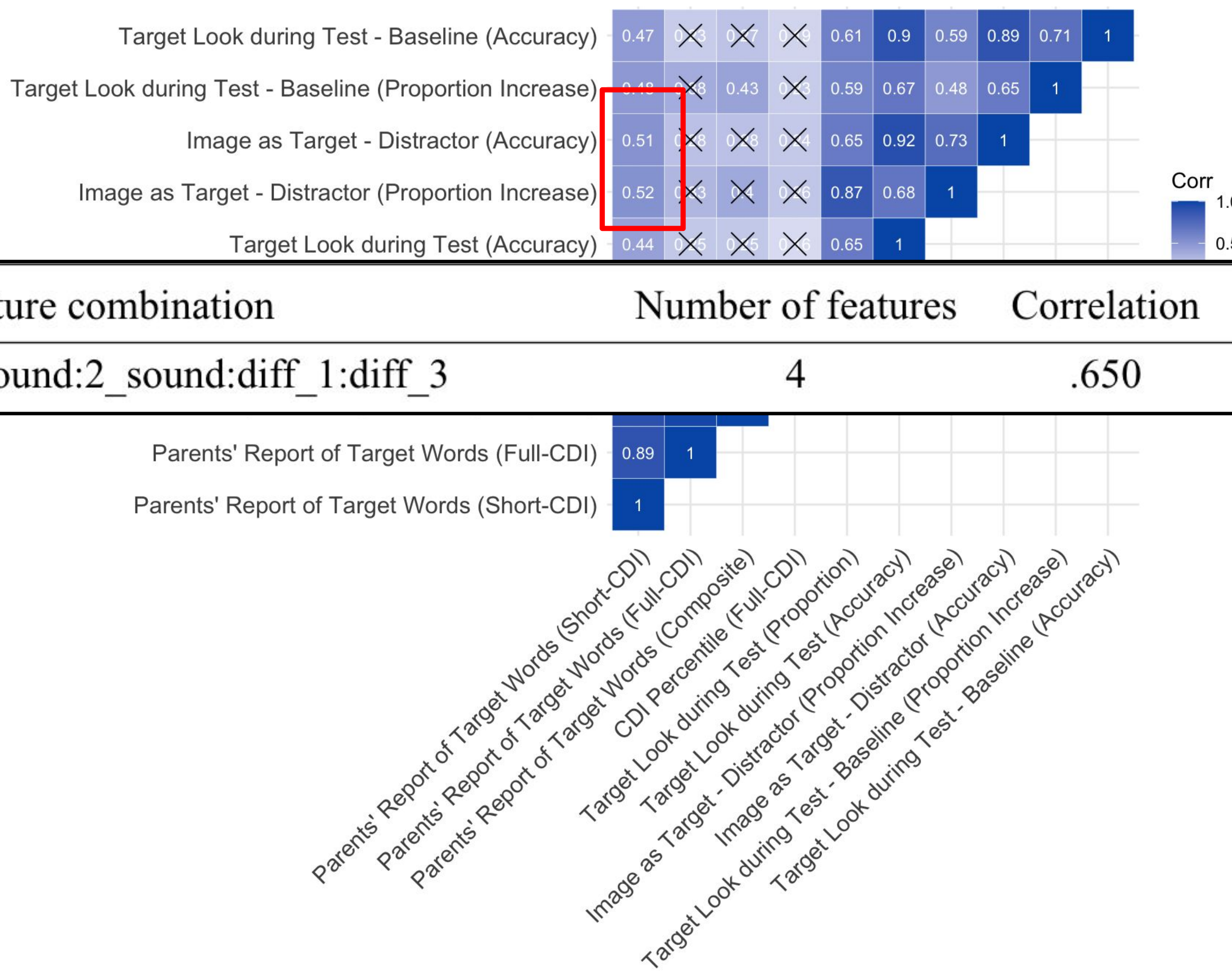
Best ten combinations:

Feature combination	Number of features	Correlation
1_sound:2_sound:diff_1:diff_3	4	.650
2_sound:diff_1:diff_3	3	.634
1_sound:3_sound:diff_1:diff_3	4	.627
1_sound:diff_1:diff_3	3	.623
1_sound:2_sound:diff_1:diff_2:diff_3	5	.614
2_sound:diff:diff_1:diff_2:diff_3	5	.595
1_sound:2_sound:diff_3	3	.594
sound:2_sound:diff_1:diff_2:diff_3	5	.592
1_sound:3_sound:diff_1:diff_2:diff_3	5	.591
2_sound:diff_1:diff_2:diff_3	4	.588

Comparisons with rule-based approach



Comparisons with rule-based approach



Feature combination	Number of features	Correlation
1_sound:2_sound:diff_1:diff_3	4	.650

Next step

- Collect more data
- Construct a better model – further refining the alignment between extracted gaze features and word knowledge by:
  - incorporating more gaze metrics
    - adjustment by image (require more repeated test with yoked pairs)
    - repeated testing of items
    - quantifying effect of word difficulty, bias, semantic relatedness, etc, but require framework adjustments
  - improving the reliability and validity of the ground truth, by conduct multiple test for convergence, such as
    - more test-retest short CDIs,
    - behavioral and naturalistic observation of child interacting with target objects

Next step

- Collect more data
- Construct a better model – further refining the alignment between extracted gaze features and word knowledge by:
  - incorporating more gaze metrics
    - adjustment by image (require more repeated test with yoked pairs)
    - repeated testing of items
    - quantifying effect of word difficulty, bias, semantic relatedness, etc, but require framework adjustments
  - improving the reliability and validity of the ground truth, by conduct multiple test for convergence, such as
    - more test-retest short CDIs,
    - behavioral and naturalistic observation of child interacting with target objects

Next step

- Collect more data
- Construct a better model – further refining the alignment between extracted gaze features and word knowledge by:
  - incorporating more gaze metrics
    - adjustment by image (require more repeated test with yoked pairs)
    - repeated testing of items
    - quantifying effect of word difficulty, bias, semantic relatedness, etc, but require framework adjustments
  - improving the reliability and validity of the ground truth, by conduct multiple test for convergence, such as
    - more test-retest short CDIs,
    - behavioral and naturalistic observation of child interacting with target objects

Next step

- Collect more data
- Construct a better model – further refining the alignment between extracted gaze features and word knowledge by:
  - incorporating more gaze metrics
    - adjustment by image (require more repeated test with yoked pairs)
    - repeated testing of items
    - quantifying effect of word difficulty, bias, semantic relatedness, etc, but require framework adjustments
  - improving the reliability and validity of the ground truth, by conduct multiple test for convergence, such as
    - more test-retest short CDIs,
    - behavioral and naturalistic observation of child interacting with target objects

Next step

- Collect more data
- Construct a better model – further refining the alignment between extracted gaze features and word knowledge by:
  - incorporating more gaze metrics
    - adjustment by image (require more repeated test with yoked pairs)
    - repeated testing of items
    - quantifying effect of word difficulty, bias, semantic relatedness, etc, but require framework adjustments
  - improving the reliability and validity of the ground truth, by conduct multiple test for convergence, such as
    - more test-retest short CDIs,
    - behavioral and naturalistic observation of child interacting with target objects

Next step

- Collect more data
- Construct a better model – further refining the alignment between extracted gaze features and word knowledge by:
  - incorporating more gaze metrics
    - adjustment by image (require more repeated test with yoked pairs)
    - repeated testing of items
    - quantifying effect of word difficulty, bias, semantic relatedness, etc, but require framework adjustments
  - improving the reliability and validity of the ground truth, by conduct multiple test for convergence, such as
    - more test-retest short CDIs,
    - behavioral and naturalistic observation of child interacting with target objects

Next step

- Collect more data
- Construct a better model – further refining the alignment between extracted gaze features and word knowledge by:
  - incorporating more gaze metrics
    - adjustment by image (require more repeated test with yoked pairs)
    - repeated testing of items
    - quantifying effect of word difficulty, bias, semantic relatedness, etc, but require framework adjustments
  - improving the reliability and validity of the ground truth, by conduct multiple test for convergence, such as
    - more test-retest short CDIs,
    - behavioral and naturalistic observation of child interacting with target objects

Thank you!

Infant eye-tracking Studies

- Eye-tracking studies more accessible
- Once finished setup, the word recognition task can be conducted anywhere with an internet connection
- Diversifying access to research tool

e-Babylab

e-Babylab (Lo et al., 2023) is an open source authoring tool that allows users or researchers to easily create, host, run, and manage online experiments, without writing a single line of code. Using this tool, experiments can be programmed to include any combinations of image, audio, and/or video contents as stimuli and record key presses, clicks, screen touches, audio, video, and eye gaze[1]. Short-form versions of the MacArthur–Bates Communicative Development Inventories (CDIs; Chai et al., 2020; Mayor & Mani, 2019) can additionally be included in experiments, allowing users or researchers to collect CDI data online.

Contents

https://github.com/lochhh/e-Babylab

WebGazer.js

WebGazer.js

Democratizing Webcam Eye Tracking on the Browser

WebGazer.js is an eye tracking library that uses common webcams to infer the eye-gaze locations of web visitors on a page in real time. The eye tracking model it contains self-calibrates by watching web visitors interact with the web page and trains a mapping between the features of the eye and positions on the screen. WebGazer.js is written entirely in JavaScript and with only a few lines of code can be integrated in any website that wishes to better understand their visitors and transform their user experience. WebGazer.js runs entirely in the client browser, so no video data needs to be sent to a server, and it requires the user's consent to access their webcam.

https://webgazer.cs.brown.edu/