**What is learned during statistical learning? Investigating online and offline implicit learning**

Contents

[**Abstract** 1](#_Toc47536752)

[**Introduction** 1](#_Toc47536753)

[**Experiment 1** 2](#_Toc47536754)

[**Method** 2](#_Toc47536755)

[**Results** 4](#_Toc47536756)

[**Discussion** 8](#_Toc47536757)

[**Experiment 2** 9](#_Toc47536758)

[**Method** 9](#_Toc47536759)

[**Results** 9](#_Toc47536760)

[**Discussion** 11](#_Toc47536761)

[**Coding of Structural Features in Online Target Detection** 11](#_Toc47536762)

[**General Discussion** 12](#_Toc47536763)

[**Conclusion** 13](#_Toc47536764)

[**Acknowledgements** 13](#_Toc47536765)

[**References** 13](#_Toc47536766)

[**Supplementary Materials** 14](#_Toc47536767)

[**Stimuli** 14](#_Toc47536768)

[**Supplementary Figures, Tables** 14](#_Toc47536769)

## **Abstract**

## **Introduction**

Statistical learning (SL) is a powerful mechanism by which the brain is able to detect implicit regularities in the sensory environment. (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015) Regardless of input modality, humans are sensitive to subtle distributional properties of incoming streams of information, and able to employ this information to complete subsequent tasks. (Armstrong, Frost, & Christiansen, 2017; Frost, Armstrong, Siegelman, & Christiansen, 2015)

Numerous experiments have shown individuals are capable of extracting transitional probabilities embedded in streams of auditory (Endress & Mehler, 2009; Pelucchi, Hay, & Saffran, 2009; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) as well as visual sequences of stimuli. (Turk-Browne, Jungé, & Scholl, 2005). The ability to track regularities in sensory streams not only leads to improvement in performance on behavioral tasks, but can also be observed in neural responses to the sequence. This neural tracking of statistics occurs whether the items are within (Turk-Browne, Scholl, Johnson, & Chun, 2010) or outside (Batterink & Paller, 2019; Turk-Browne, Scholl, Chun, & Johnson, 2008) the focus of attention (for a review, see Batterink, Paller, & Reber, 2019). Complex statistics, such as non-adjacent relationships, can also be detected alongside adjacent relationships (Bonatti, Peña, Nespor, & Mehler, 2005).

Yet, despite the considerable amount of evidence in favor of SL as a robust ability, the behavioral measures used to evaluate the degree of learning are psychometrically weak and have yielded highly variable results among participants (both within and across studies). (Siegelman, Bogaerts, & Frost, 2017) Since the original study using continuous speech to test for SL in infants (Saffran, Aslin, et al., 1996), this ability has been canonically measured using an explicit recognition task, often with little or no modification from the original, after participants have been exposed to the stimuli sequence in a learning phase. (Siegelman et al., 2017) The task typically consists of pairing each of the units that make up the sequence (e.g. tri-syllabic pseudowords or triplets of visual shapes) with a foil, a test item of equal length but which either violates the transitional probability structure of the learned units, or is made up of random elements from the stimulus set (syllables, shapes). Participants then report which of the pair is more familiar, more similar to the exposure stream, or explicitly asked which appeared in the previous part of the experiment.

Meanwhile, several other measures have been developed that purportedly measure statistical learning “online.” These tasks aim to capture the dynamic aspects of SL (such as its onset, stability, and change over the course of exposure) and provide more insight into the contribution of individual stimulus items in generating the ultimate effect. One such measure is the target detection task, which entails presenting participants with a target item (syllable, shape, etc.) and asking them to respond as fast as possible whenever the target is seen or heard in the ensuing sequence. In both visual (Bertels, Franco, & Destrebecqz, 2012; Franco, Eberlen, Destrebecqz, Cleeremans, & Bertels, 2015; Turk-Browne et al., 2005) and auditory (Batterink & Paller, 2017; Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015) SL studies, this task reveals graded reaction times (RT), such that responses to items in less predictable positions (e.g. with TPs of 0.33) are slower than responses to items in more predictable positions (e.g. with TPs of 1). The task can also be employed as a cover task during the exposure phrase, with similar results. (Batterink, 2017)

Variations of this theme consist of having participants detect click sounds embedded in the exposure stream, where RT to clicks within pseudowords are slower than clicks placed between pseudowords. (Gómez, Bion, & Mehler, 2010; but see (Franco, Gaillard, Cleeremans, & Destrebecqz, 2014 for a non-replication)

This finding, that embedded transitional probabilities modulate RT, is often interpreted as evidence that participants learned the higher-order units. In other words, that they formed a representation of the tri-syllabic pseudowords or shapes that made up the exposure sequence. (e.g. Bertels, Franco, & Destrebecqz, 2012) This claim appears unsupported or, at best, controversial, given the inconclusive data from correlation analyses between the two measures.

Some evidence exists that performance in the online task predicts (or minimally, correlates with) performance in the offline task. In a previous study, an analysis comparing a familiarity task (of words vs. part-words as well as non-words) versus change in median RT to word-initial versus word-final syllables, the correlation coefficient was moderate and statistically significant, at . (Batterink & Paller, 2017) Another study that measured RT to occasionally-repeated shapes in a visual SL paradigm also found a significant correlation between an online measure of SL, here defined as:

)

and compared with an offline recognition score (out of 32 2AFC tasks with triplet and phantom triplet foils) (). (Siegelman, Bogaerts, Kronenfeld, & Frost, 2018)

However, in another paper by Batterink and colleagues, the same analysis as above yielded no significant correlation (). (Batterink, Reber, Neville, et al., 2015) A second experiment reported in the same paper similarly reports a correlation of for a different cohort of participants. This null finding is echoed by (Franco et al., 2014) using auditory stimuli, as well as (Siegelman & Frost, 2015) using a similar serial reaction time task in battery of visual and auditory SL sub-experiments.

This lack of correlation between SL measures has received little direct attention, save for general comments noting the potential different theoretical mechanisms underlying each task (i.e. one being an implicit measure of sensitivity to transitional probabilities, the other being an explicit task requiring recall). In two experiments, we addressed the question of why these two measures might both consistently points towards learning, while remaining uncorrelated or merely intermittently correlated.

Here, we replicated the online target detection task and found further evidence for transitional probabilities modulating RT. We also found significant word recognition performance at the group level in the classic 2AFC task. We failed to uncover a correlation between these two measures. We extended this replication by running a second experiment in which participants performed the online target detection task for both structured and randomly ordered sequences of continuous speech syllables. Data from this task in both experiments was combined to address the question of why the target detection task fails to correlate with word recognition performance.

# **Experiment 1**

## **Method**

#### **Stimuli**

Speech stimuli consisted of 12 consonant-vowel (CV) pairs. We selected 5 unique vowels that are maximally separated in their manner and place of articulation. We ensured that none of these vowels typically occurred in unstressed syllables in spoken German. We then selected 12 unique consonants, in order to render each syllable phonetically distinct from the others. We used the CELEX database to calculate the frequency of occurrence of each of our syllables in spoken German, as well as the frequency of co-occurrence between each pair of syllables. We eliminated high-frequency CV pairings from our list of possible syllables and formed the final words by combining three syllables (each with distinct vowels) for which no transitions were frequent in spoken German. Final syllabes were: be, di, ga, ki, la, mi, nu, po, ro, se, tu, za.

A male native speaker of German was recorded pronouncing each syllable in our set separately and with a flat intonation. Each syllable was repeated several times to ensure we obtained a quality token. The final selection of syllables were then high-pass filtered at 50 Hz and silences before and after syllable were removed using a custom script in Matlab 2017b. The 12 syllables were normalized for pitch and intensity using Praat to ensure relative homogeneity between tokens. Finally, syllables were temporally compressed to 240 ms in duration and a 10 ms silence was added at the end of each syllable, for a total duration of 250 ms.

Syllables were combined into 4 tri-syllabic pseudowords such that each word featured no repeating consonants or vowels and similarity between any possible succeeding pairs of syllables was minimized. We also ensured that no pairs were phototactically illegal or shared a resemblance with existing words in German. Pseudowords for our study were: nugadi, rokise, mipola, zabetu. Part-words, used in the word recognition task, were of the form C’AB (word-final syllable from one word followed by word-initial and word-medial syllables from another): dizabe, semipo, lanuga, turoki.

Continuous speech sequences (24) of ~1 minute in length were created in Matlab by concatenating the pseudowords such that none repeated consecutively. As per the design in (Saffran, Aslin, et al., 1996), standard in SL studies, the only cue to segmenting the sequence lay in the transitional probabilities between syllables. The transitional probability of word-medial and word-final syllables (relative to the preceding syllable) was 1, while the transitional probability of word-initial syllables was 0.33. Speech streams were ramped up and down in amplitude over a period of 1.5 seconds so that onset and offset syllables were not clearly distinguishable and could not serve as cues to word boundaries.

#### **Procedure**

41 individuals participated in the study (**X** female, mean age, **X** ± **X** sd). Two participants were removed from the data pool due to technical failure. Of the 39 remaining datasets, 33 were used in analyzing the target detection task (one participant failed to follow instructions, and technical issues caused partial data loss for the other five). Since the design of our experiment was modular, technical failure in one task did not necessarily affect data in another. Thus, of the 39 datasets, we were able to use 38 for analyzing the word recognition task (data from one participant in this task was overwritten).

Participants were seated in a dimly-lit, sound-attenuated booth, approximately 52 cm from the monitor and listened to the stimuli via headphones connected to a headphone amplifier (Beyerdynamics-DT-770 80 Ohm; Lakepeople G103P1262). Stimulus intensity level was approximately 57 dB (LAF: min 44 dB, max 76 dB), as measured by a NTi Audio device connected to an artificial ear on which the experiment headphones were mounted.

The experiment was designed using Presentation® (Version 20.1 Build 12.04.17) and delivered on two versions of the software (Version 20.0 Build 07.26.17 and Version 21.1 Build 09.05.19, Neurobehavioral Systems, Inc., Berkeley, CA, [www.neurobs.com](http://www.neurobs.com)). The experiment was conducted on a 64-bit Windows machine (Fujitsu Celsius M740B) running Windows 10.

During the exposure phase, participants listened to a total of approximately 24 minutes of continuous speech. Participants were told they would hear brief sequences of sounds from an alien language. Audio was presented binaurally. Before the start of each stream, one of the 12 syllables was displayed orthographically on the screen and played aurally twice. Participants were instructed to press the spacebar as fast as they could during the subsequent stream whenever they heard this target syllable. Each of the 12 syllables served as a target syllable twice. The presentation order of syllables was pseudo-randomly shuffled for each participant so that a syllable from each ordinal position in the pseudoword (1st/word-initial, 2nd/word-medial, or 3rd/word-final) was tested before any were repeated. The 24 streams were organized into 8 blocks, where each block consisted of 3 streams with one target syllable from each ordinal position tested. Within each stream, target syllables appeared approximately 18 times. Participants could take self-paced breaks between blocks.

In the word recognition task, participants completed 16 trials of a two-alternative forced-choice task. In each trial, a pseudoword and a part-word were presented (counterbalanced across trials), and participants were prompted to determine which of the pair was a word in the alien language they had just heard in the previous section. The inter-stimulus-interval between words was 400 ms, while inter-trial-interval was 1.2 seconds. Each pseudoword was paired with each part-word once (4 x 4 trials).

Our experiment also included an additional task, designed to measure perceived speed of the speech stream before versus after the exposure phase. Results from this task will not be discussed here.

#### **Analyses**

All analyses were performed in RStudio (version 1.2.1335; RStudio Team 2018) using R version 3.6.1 (R Core Team 2019). Raw data from Presentation® was transformed into appropriate formats using Matlab R2017b (version 9.3.0.713579). Presentation® scenario files, data wrangling scripts, and R scripts with used packages are available publicly on Github.

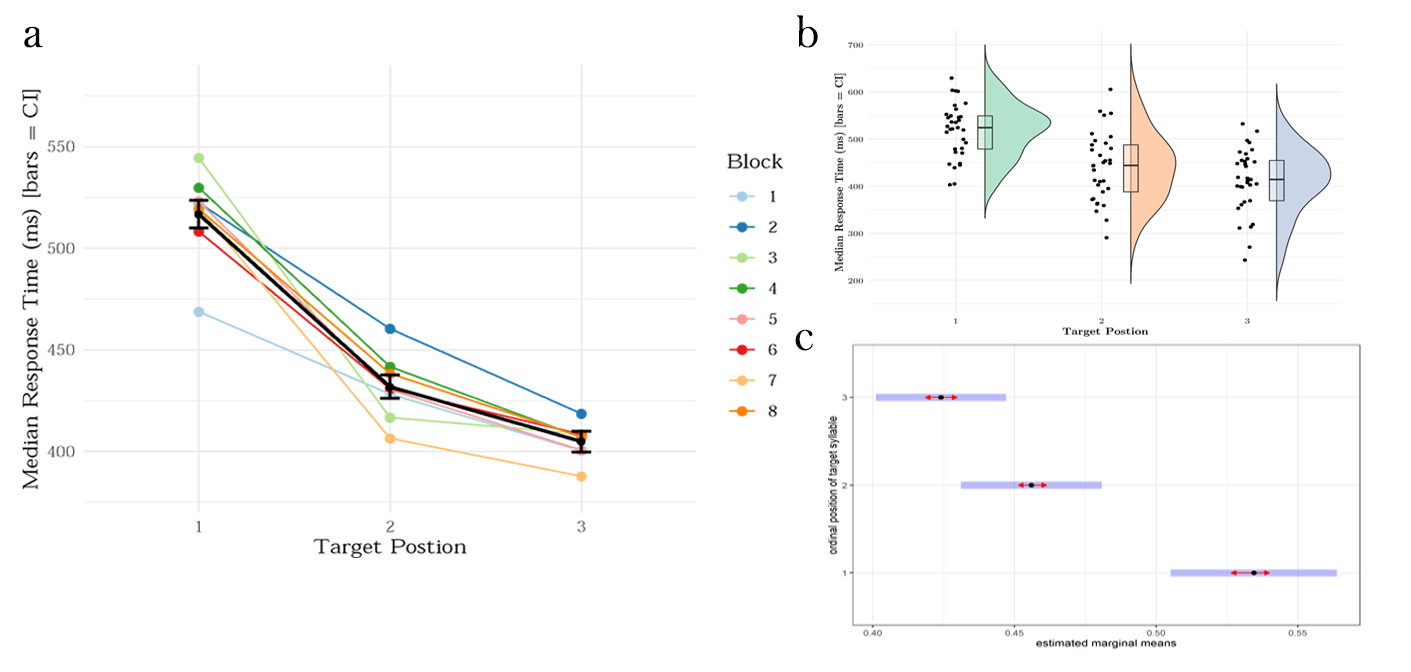
For the target detection task analyses, we considered only those responses that occurred within a boundary of ± 3 times the median absolute deviation. This procedure ensures that RT cutoffs would be based on the distribution of the raw data and not arbitrary limits. At the same time, the use of the median as the centrality metric is arguably more appropriate, given that the mean can be a biased estimator of RT data, which typically follows a gamma, lognormal, or ex-Gaussian distribution. This procedure eliminated only 0.034% of the data and resulted in RT that ranged from 0 to 943 ms (versus the original 0 to 1298 ms).

## **Results**

***Ordinal Position Modulates Reaction Time***

To replicate findings that showed graded reaction times in response to syllables in different ordinal positions, we ran a generalized linear model with reaction time (in seconds) as outcome variable, fitted with a gamma function and log link function. Our full model included both ordinal position and block as fixed effects factors, and subject as a random intercept-random effects factor. This model was compared with a lesser model in which only ordinal position was used as a fixed effect. The lesser model provided a better fit of the data, with a lower AIC () value and significantly lower deviance (). (See **Table S1** for regression results.) We also compared both the fuller and the lesser models with random slopes for levels of ordinal position in the random effects term, but the lesser model with only varying random intercepts in the random effects term still proved a better fit for observed data (see **Table 1.** Model output for the fuller random slopes model not shown; lesser vs. fuller random slopes models deviance was and , respectively; ). Thus, we conducted further analysis on results of the lesser model.

We found that reaction times are modulated by ordinal position (main effect of ordinal position, ). (**Fig. 1**) We conducted pairwise comparisons on estimated marginal means for levels of the factor position with Tukey adjustment, to explore the drop in reaction times between each ordinal position. (**Table 1**. Note that estimates represent differences in estimated marginal means on the response scale in seconds.) Specifically, RTs to word-initial syllables () are notably slower than those to word-medial () and word-final syllables (). The estimated drop in mean RT between positions 1 and 2 was 78 ms, while the drop between positions 1 and 3 was roughly 110 ms. The difference in mean RT between positions 2 and 3 was smaller, at about 32 ms. All comparisons reached statistical significance at the 5% alpha level, suggesting both the transitional probability and ordinal position contributed to the graded RT effect. This is because TP is the same between word-medial and word-final syllables, yet word-final syllables yielded significantly faster RTs than word-medial syllables.



**Figure 1. A. Reaction times to target syllables are modulated by ordinal position in pseudowords. Participants responded more slowly to syllables in the word-initial (1st) position (TP = 0.33) than to syllables in the word-medial (2nd) or word-final (3rd) position. This effect was present in block 1 and remained stable throughout subsequent blocks. B. Median RTs for each position collapsed over blocks. Dots are individual participant medians. Box plots indicate group median and CI’s, half-violins describe distribution. C. Estimated marginal means for each ordinal position, from generalized linear model. Red arrows indicate 95% CI.**

To improve interpretability of the results, we also computed s-values for each p-value in our contrasts. The s-value is a metric of self-information, or surprisal information measure. It can assist in the interpretation of p-values by providing an intuitive transformation of the p-value, taking the negative log of the p-value: . (Greenland et al., 2016) A value of *p = 1* is perfectly unsurprising under the null hypothesis. Surprisal increases exponentially as *s* approaches zero. Thus, *s* becomes a measure of information in bits against the null hypothesis. The difference in means between positions 1 and 3 can be quantified in roughly 139.3 bits. If the null hypothesis of no difference in means is true, this result is as surprising as getting all heads in 140 fair coin tosses (*s* rounded to the nearest integer).

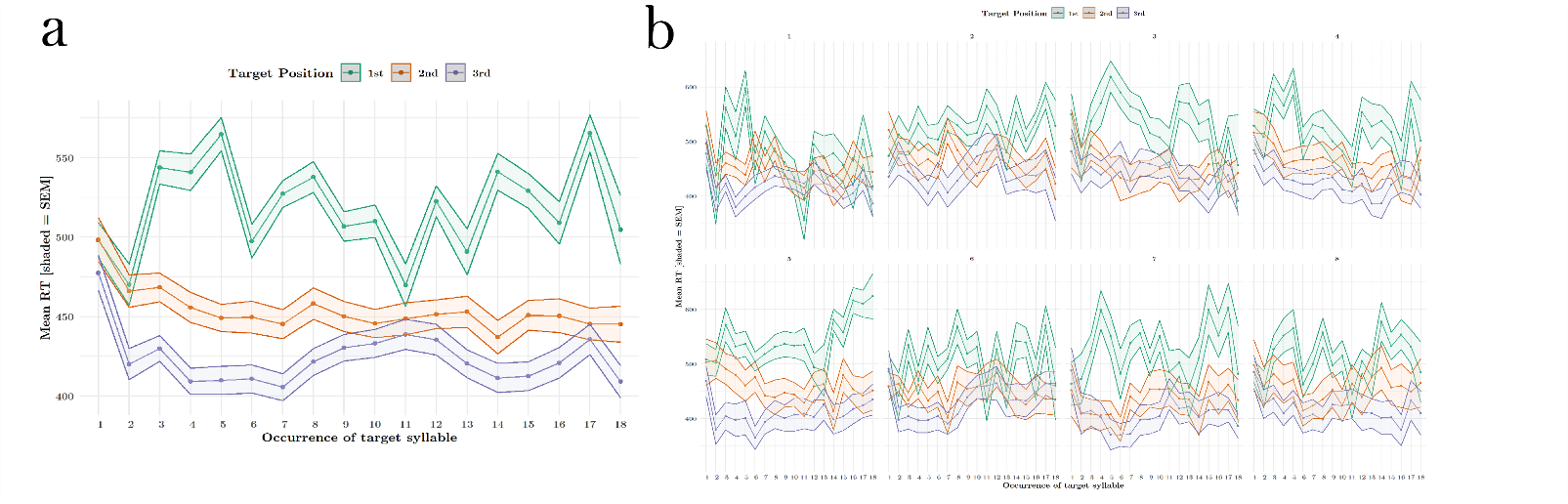
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | |
| ordinal position | estimate | SE | lower CI | upper CI | z ratio | p value | s value |
|  | | | | | | | |
| 1 - 2 | 0.078 | 0.006 | 0.065 | 0.092 | 13.770 | < .001 | 139.307 |
| 1 - 3 | 0.110 | 0.006 | 0.097 | 0.124 | 18.964 | < .001 | 262.419 |
| 2 - 3 | 0.032 | 0.004 | 0.022 | 0.042 | 7.285 | < .001 | 39.917 |
|  | | | | | | | |

***Table 1. Estimated marginal mean contrasts for reaction times to targets in each ordinal position. All contrasts reached statistical significance at the 5% alpha threshold. Contrasts were performed on model output from the generalized linear model described above.***

To ensure that participants were able to perform the task, we computed mean detection accuracy across participants and for each target syllable. We observed that certain syllables were detected less often than others, suggesting some unwanted variability in the stimuli (notably for syllables *ro, za, be, and mi).* (See **Fig. S2** and following section for discussion.) We therefore sought to validate the results reported above by regressing out the effect of individual syllable as a function of ordinal position. We ran a generalized mixed model with ordinal position and target syllable as fixed effects factors and subject as random effect factor, with the RT (in ms) as outcome variable. We then subtracted the resulting residual values for each data point from the raw RT, and re-ran the lesser model as specified above using the adjusted RT values. We still observed the main effect of ordinal position (*)* and therefore concluded that slight variations in the acoustics of our stimuli did not significantly affect our results*.*

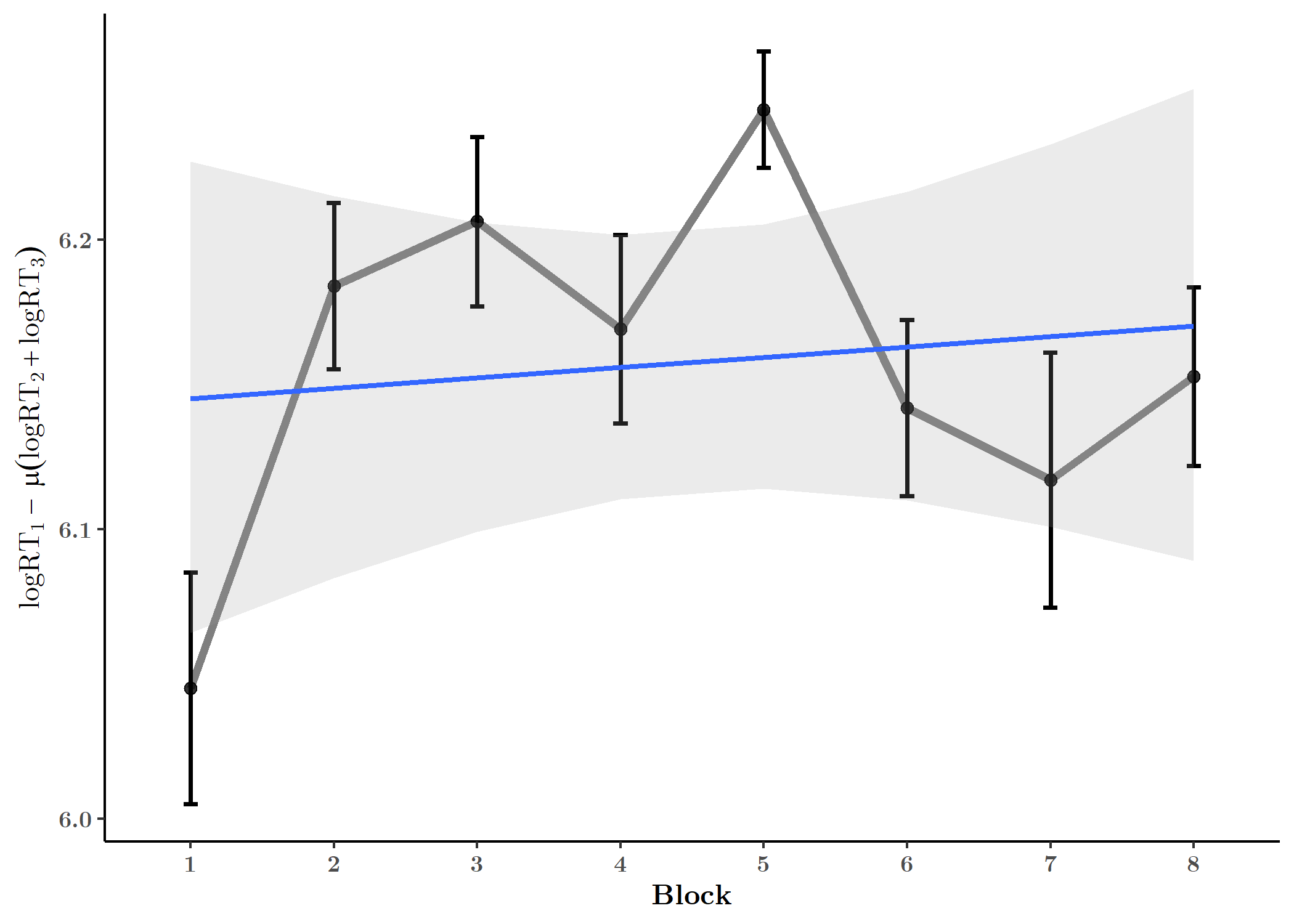
***Rapid Onset of Graded RT to Predictable Syllables***

Furthermore, this pattern of reaction times emerged within the first few presentations of the target syllable, and remained stable throughout the remainder of the experimental blocks. (**Fig. 2**) This observation accounts for why the factor of block did not contribute significantly to model fit (); reaction times differentiate early on and persist across the whole data set, (**Fig. 2a**) despite some variation observable within individual blocks (**Fig. 2b**).



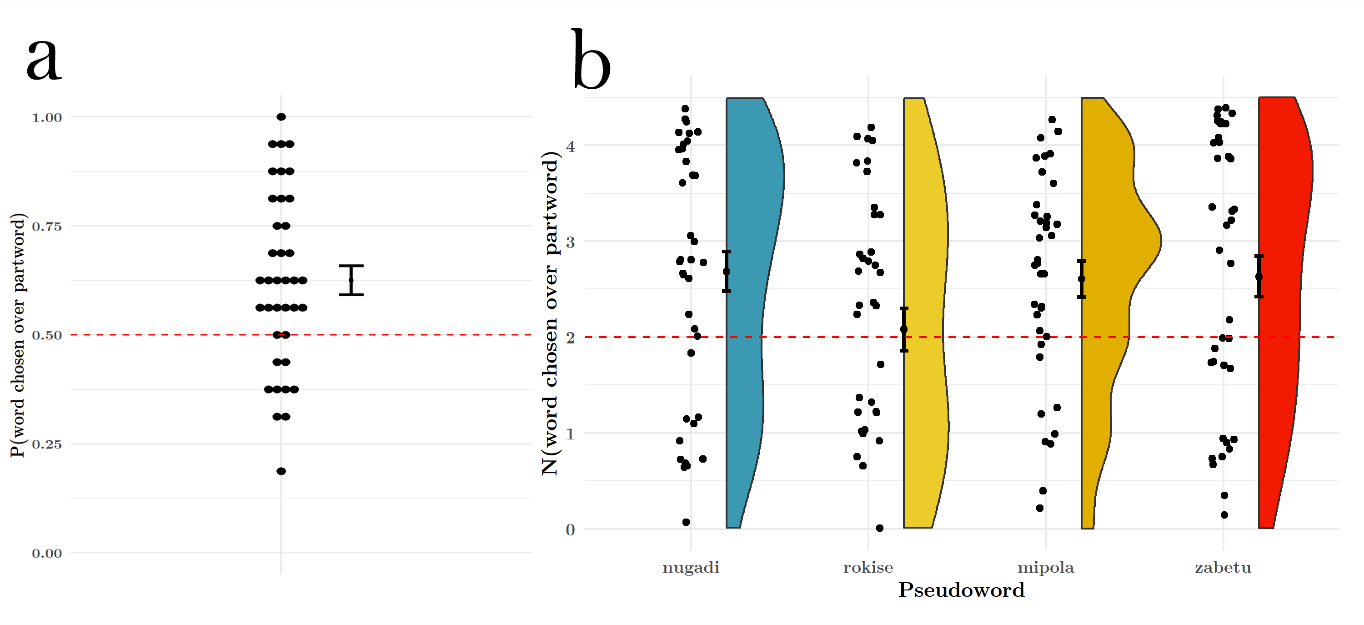
**Figure 2. Moving Average of RTs to target syllables, by ordinal position. Moving averages were calculated by taking the average of every three trials (each block contained ~18 occurrences of each syllable).**

Finally, we computed a measure of online statistical learning from the target detection task, borrowed from Siegelman and colleagues, following the formula stated in the Introduction. (Siegelman et al., 2018) This online measure of SL, when observed for each block, revealed rapid “learning”, as measured by differences in log RT to the three target positions, between blocks 1 and 2, that remained stable albeit with some variability thereafter. A linear model fit to predict this online measure from factor block failed to reveal a main effect ().



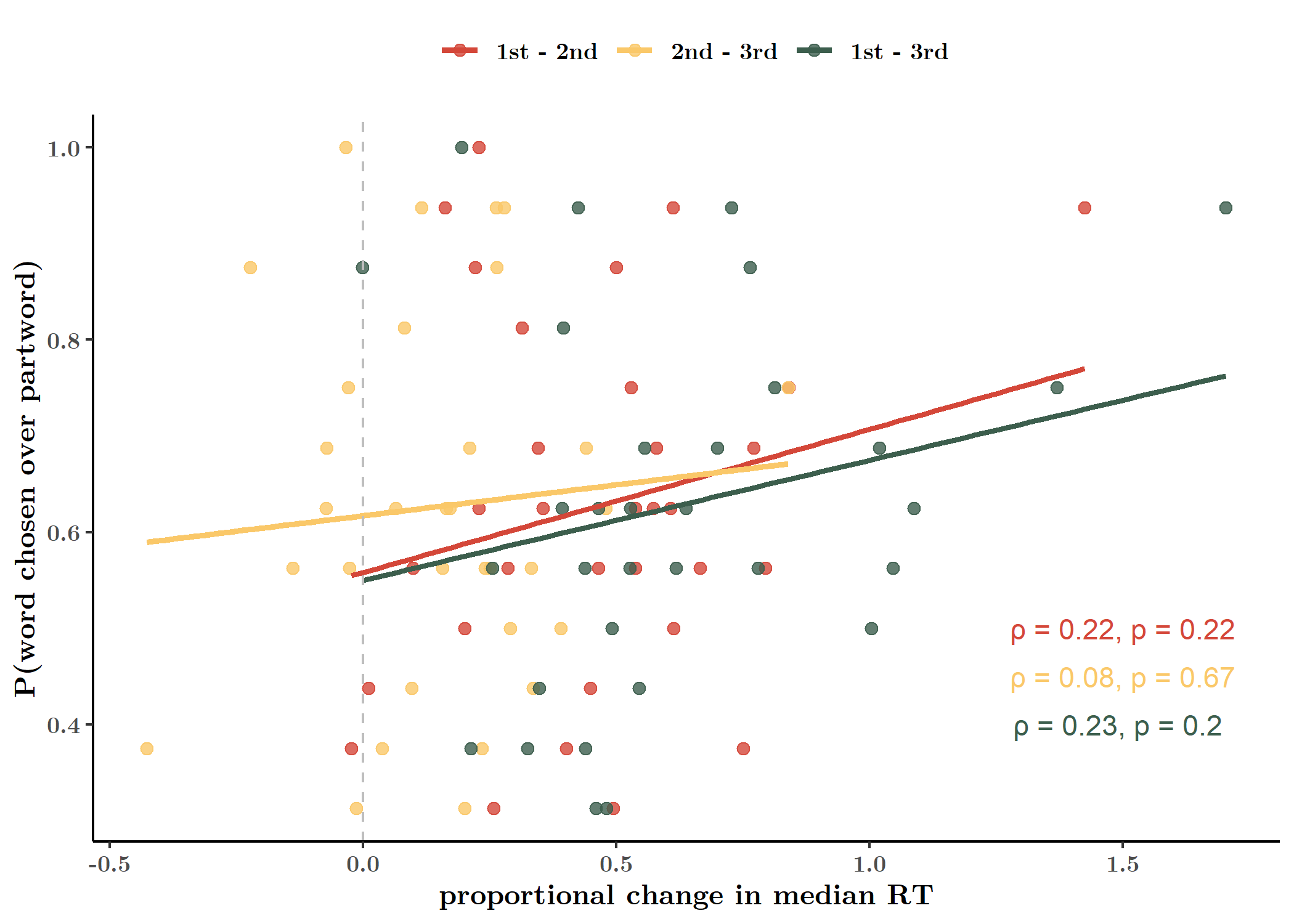
***Psuedowords Can Be Distinguished From Part-words***

In the explicit word recognition task, 71% of participants performed above a 50% chance level (27 out of 38). The proportion of trials on which participants correctly distinguished the pseudoword from the part-word was significantly above a chance level of 0.5 (), indicating that participants were sensitive to the implicit regularities of the syllable stream. **(Fig. 3a)** In an exploratory analysis, we also calculated the proportion correct responses for each psuedoword individually, in order to determine if any particular word was driving the overall word recognition effect. This is because distinguishing only a single word from its paired part-word, would be sufficient to push a participant’s performance above chance level. We found that across participants, 3 out of 4 words were discriminated above chance level of 2 correct discriminations per word. () (**Fig. 3b**)



**Figure 3**

***Online and Offline Measures Are Weakly Correlated***



**Fig. 4. Pearson correlation of online and offline learning measures (median).**

Finally, also asked whether the online and offline measures of statistical learning were correlated, i.e. whether sensitivity to transitional probability online predicts explicit word recognition. Since we had an unequal number of data sets for the two tasks, we used data only from participants with complete data from both tasks (N = 30). For this analysis, we computed the difference in median RTs for each participant and for each ordinal position pair (1-2, 2-3, and 1-3). These values were correlated against the participant’s proportion correct word recognition performance. Surprisingly, word recognition performance (prop. correct responses) and response time change was weakly correlated (Test for association between paired samples of Pearson's product moment correlation coefficient). (**Fig. 4**)

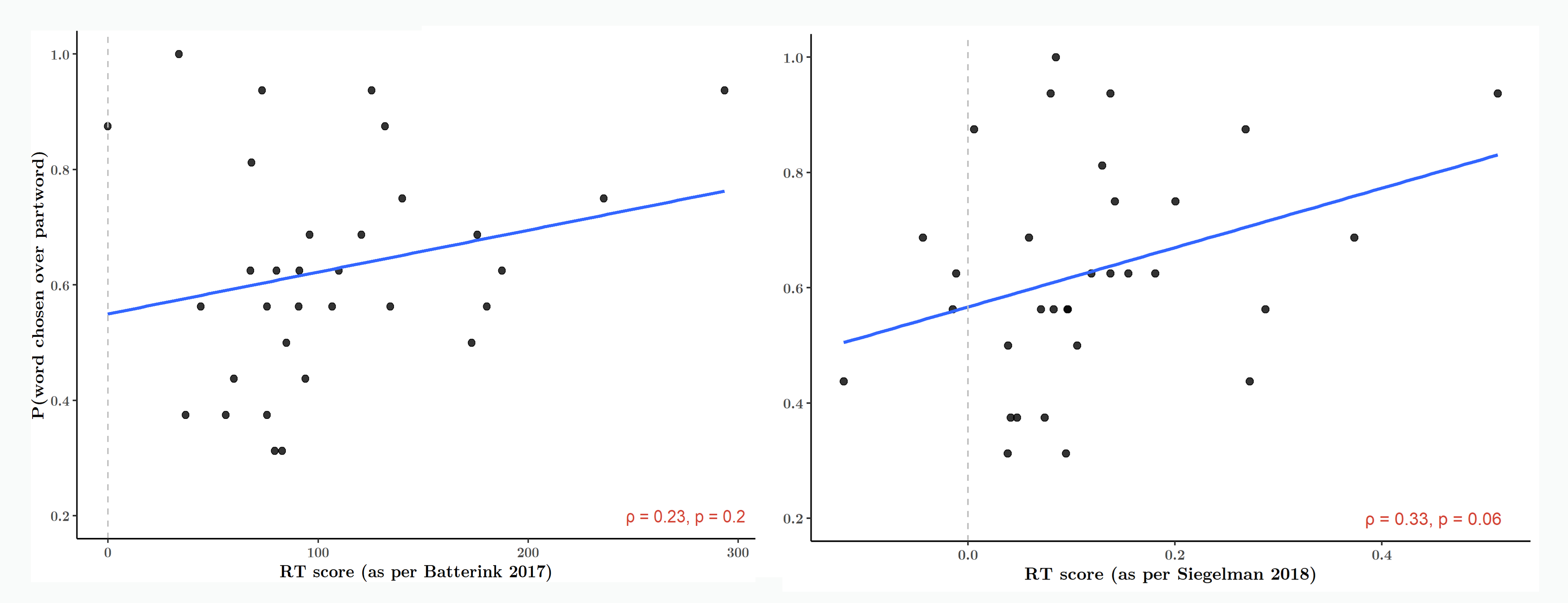


Fig. 5.

## **Discussion**

Our study replicated two tasks that measure statistical learning in distinct ways. Our offline word recognition task revealed a well-established effect of statistical learning, which is the ability to explicitly discriminate a properly formed pseudoword from a sequence of syllables that was heard but which span a word boundary (a part-word). This pseudoword vs. part-word test is more conservative than its sister version, the word vs. non-word test, in which words are tested against random combinations of syllables which never occurred in that particular order during the learning phase. (Batterink, Reber, Neville, et al., 2015)

The online target detection task asks participants to response as soon as they heard a target syllable. Reaction times appeared to be modulated primarily by their transitional probability, which is equivalent to their predictability, such that highly predictable syllables occurring in word-medial and word-final positions elicited much faster responses than less predictable syllables occurring in word-initial positions. The rapid differentiation of response times, in the first block, confirms previous findings that statistical learning is a very fast and robust mechanism.

However, it is puzzling why these two measures of statistical learning are poorly correlated. In a previous studying comparing a similar explicit task of familiarity (of words vs. part-words as well as non-words) versus change in median RT to word-initial versus word-final syllables, the correlation coefficient was higher, at r = 0.42 (p = 0.044). (Batterink & Paller, 2017) This correlation coefficient is more similar to what we observed if we compared word recognition performance with the differences in mean RT to each position (r = 0.34, **Fig. S6**), but this relationship remained weak for our data (p = 0.05). In another paper published by the same group, the same analysis yielded no significant correlation. Despite finding a similar reaction time effect, no correlation was observed between the median RT to word-initial versus word-final syllables and accuracy on the familiarity task (r = 0.07, n.s.) (Batterink, Reber, Neville, et al., 2015) A second experiment reported in the same paper reports a correlation of r = 0.26 (n.s.) for a different cohort of participants. A study that measured RT to occasionally-repeated shapes in a visual SL paradigm, however, found a significant correlation between an online measure of SL (logRT(1st position) – meanlogRT (2 & 3rd positions) when compared with an offline recognition score (out of 32 2AFC tasks with triplet and phantom triplet foils) (r = 0.46, p < 0.001). (Siegelman et al., 2018)

Multiple explanations may account for this finding. First, the two tasks vary in their psychometric sensitivity. The target detection task in our experiment comprised a total of roughly 432 “trials” (occurrences of a target to be detected; ~18 per stream x 24 streams) and 35 trials per test item (syllable), while the word recognition task consisted of only 16 trials and 8 test items (4 pseudowords and 4 part-word foils). The correlation therefore may be weak (or the computed coefficient unstable) due to a lack of power stemming from the experiment design. Nonetheless, a vast majority of statistical learning studies have used this task with little or no modification. Test trials in the 2AFC recognition task are rarely greater than 36, typically testing only 4-8 pseudowords or visual triplets (in the case of visual statistical learning). (Siegelman et al., 2017)

Second, it had been widely noted that explicit tasks such as the word recognition task are different in nature than implicit detection tasks (Batterink, Reber, Neville, et al., 2015), which may more closely resemble series response time tasks while yielding similar behavioral results. (Karuza et al., 2016; Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013) Online tasks have the potential to capture the dynamic progress of statistical learning and therefore reveal more subtle learning effects. (Siegelman et al., 2018). Theoretically, it is unclear whether the tracking of statistical regularities or transitional probabilities necessarily entails the chunking or explicit representation of larger units within a sensory stream. (Dehaene et al., 2015) These regularities may result in the generation of an event boundary (Schapiro et al., 2013; Zacks & Swallow, 2007), without enabling explicit recall of the event itself.

We wished to investigate why graded response times in the online task failed to adequately predict offline pseudo-word recognition. The syllables in the stream are each characterized by several features: ordinal position, transitional probability, within-word duplet pairing, and within-word triplet pairing. Success on the online tasks requires only a tracking of transitional probability, while success in the offline task requires a representation of the within-word triplet pairing (in other words, the representation of the tri-syllablic pseudoword as a single unit).

To do this, we first aimed to replicate our online target detection task findings. By comparing the current (structured stream) task with a condition in which participants are exposed to the same stimuli with the same detection task, but without any embedded regularities, we could additionally demonstrate that the reported effect is primarily driven by statistical regularities and not uncontrolled variation in the stimuli acoustics. We then aim to run a correlation analysis on the combined data set to further explore what features of the stream are being captured by the online task.

# **Experiment 2**

## **Method**

***Stimuli***

The stimuli used in Experiment 2 were identical to those used in Experiment 1.

***Procedure***

For this experiment we synthetized 12 structured streams (as per the method above) and 12 random streams (wherein syllables are pseudorandomly ordered).

## **Results**

***Ordinal Position in Structured Stream Modulates Reaction Time***

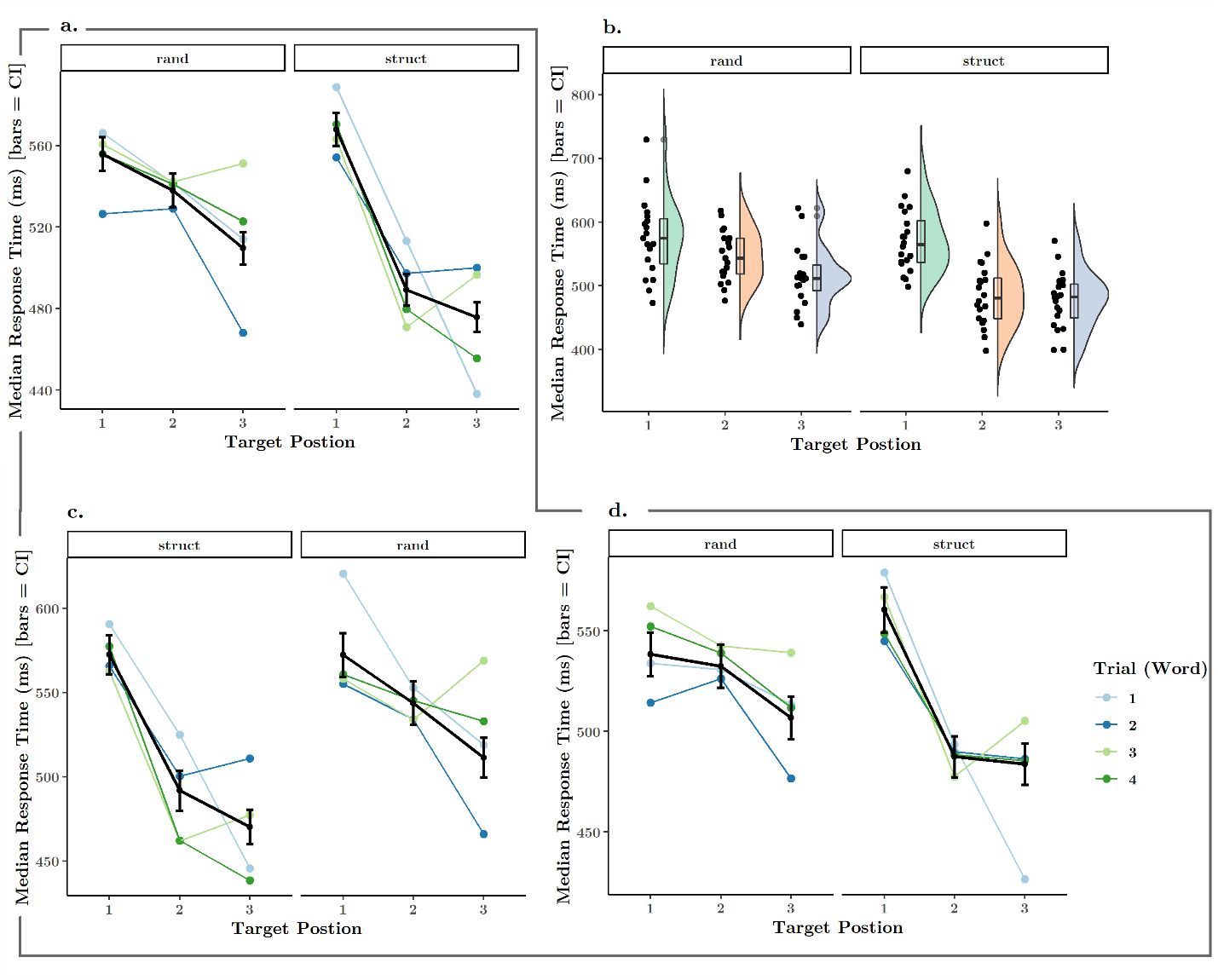


Figure 6

We performed a modelling procedure similar to that from Experiment 1. Our hypothesis entailed response times (in seconds) would be predicted by a combination of ordinal position and session (structures or random). We included subject as a nested effect within session order (whether participants completed the structured session before the random session, and vice versa), as the latter was our between-subjects variable. We further specified the random effects term by allowing random intercepts and uncorrelated to this, random effects for each level of session. This structure allows the graded RT curve for each participant to vary between sessions, as well as their baseline RT (intercept). (See **Table S2** for regression results.) We observed main effects of both session and ordinal position, as well as an interaction between them ().

Two performed two planned contrasts, the first to evaluate the effect of ordinal position within each level of session (i.e. to determine the modulation of reaction times for each condition). The second evaluated the effect of session for each level of ordinal position (i.e. how much session affected reaction times to targets in each ordinal position). When examining the effect of position within session, we observed a significant drop in estimated means between positions 1 (M = 0.581, SE = 0.011) and 2 (M = 0.565, SE = 0.011), as well as between 1 and 3 (M = 0.534, SE = 0.01) in the structured condition. There was a smaller, but also statistically significant decrease in means between positions 2 - 3 (**Table 2**.). In the random condition, we were surprised to observe a similar pattern, where differences in estimated means for each pair of positions reached still significance, with the smallest change occurring between positions 1 (M = 0.582, SE = 0.011) and 2 (M = 0.510, SE = 0.01); and position 3 syllables having shorter RTs (M = 0.489, SE = 0.009). (**Table 2,** **Fig. 7a**) Given that there were no regularities in the random stream that could bias reaction times to certain tokens more than others, we hypothesized that the modulation observed here is due to variations in the acoustic features of the stimuli, as noted above.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
| ordinal position | session | estimate | SE | lower CI | upper CI | z ratio | p value | s value |
|  | | | | | | | | |
| 1 - 2 | rand | 0.015 | 0.006 | 0.001 | 0.030 | 2.453 | 0.038 | 4.729 |
| 1 - 3 | rand | 0.047 | 0.006 | 0.032 | 0.061 | 7.731 | < 0.001 | 44.190 |
| 2 - 3 | rand | 0.031 | 0.006 | 0.017 | 0.045 | 5.324 | < 0.001 | 21.650 |
| 1 - 2 | struct | 0.072 | 0.006 | 0.059 | 0.086 | 12.690 | < 0.001 | Inf |
| 1 - 3 | struct | 0.093 | 0.006 | 0.079 | 0.106 | 16.370 | < 0.001 | Inf |
| 2 - 3 | struct | 0.020 | 0.005 | 0.008 | 0.032 | 4.006 | < 0.001 | 12.420 |
|  | | | | | | | | |

**Table 2.**

However, this observation alone only suggests that variability in stimuli caused some variability in response. When examining the effect of session (presence or absence of statistical structure) for each level of ordinal position, we observed that the presence of structure significantly decreased mean reaction times for 2nd (z(Inf) = 5.28, p < 0.0001) and 3rd position targets (z(Inf) = 4.51, p < 0.0001). While RTs to 1st position targets remained largely the same (reaction times for 1st position targets in fact increased, but this change was not statistically significant; z(Inf) = -0.12, p = 0.90). (**Fig. 6**)

We additionally tested a model that included session order as a fixed effect, to ensure that there were no carry-over effects of session. Previous studies have indicated that statistical learning can be easily disrupted by exposure to streams with differing or no statistics. [citation needed] We did not detect a three-way interaction between session order, ordinal position, and session in modulating response times (). (**Fig. 5 c-d**, **Fig. S5**)

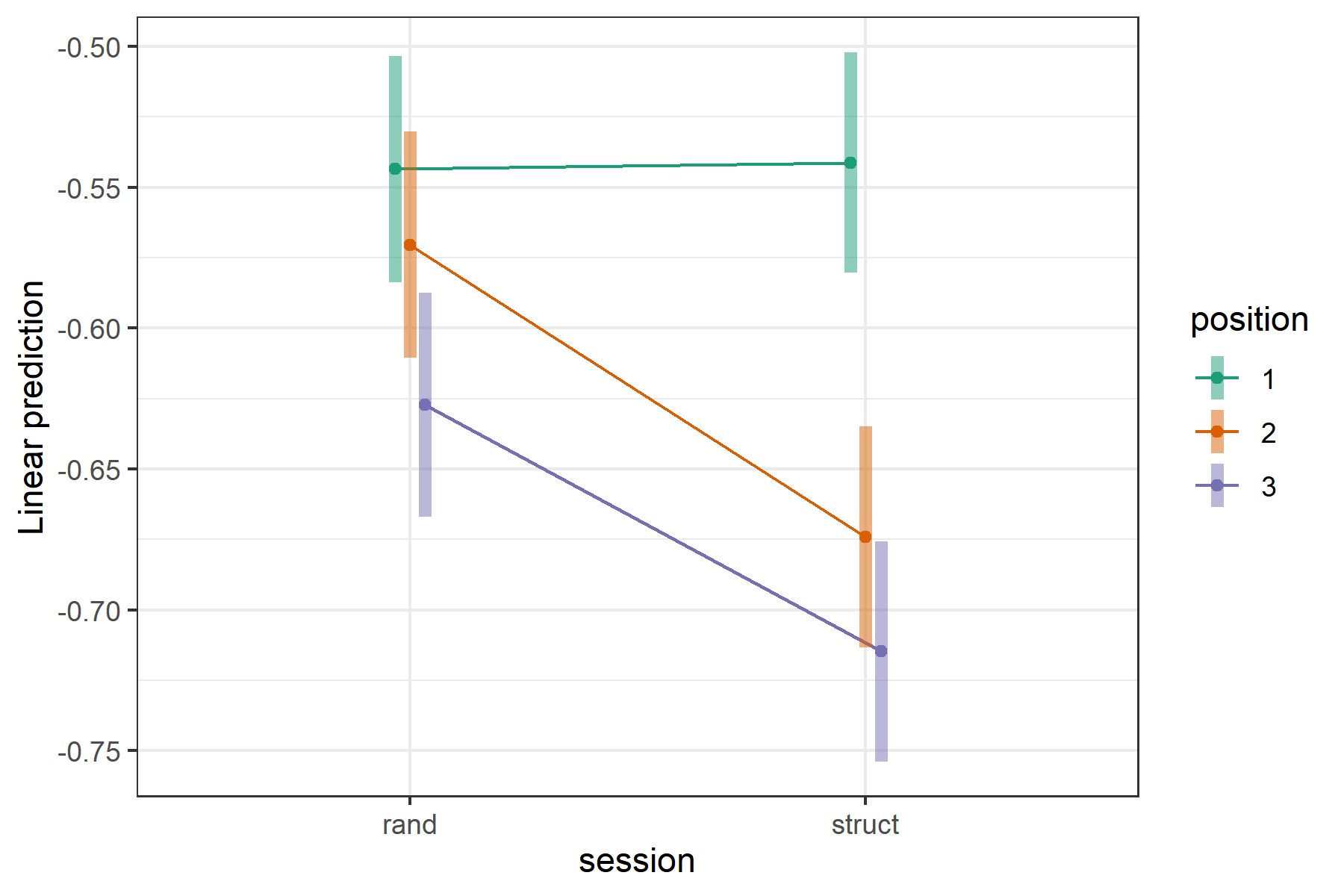


Fig. 7

## **Discussion**

Here we were able to replicate our findings from Experiment 1, in which the presence of implicit statistical structure modulates reaction times to items embedded in the stream. Notably, we could establish that this modulation is most extreme for those items which become strictly predictable (those with a transitional probability of 1, in word-medial and word-final positions), as compared with the same items tested in a stream of randomly ordered syllables.

# **Coding of Structural Features in Online Target Detection**

In a final analysis, we aimed to explore the reaction time data from the online target detection task further, in order to determine if patterns of RTs could reveal any particular coding strategies which might lend clues to what features of the implicit structure participants picked up.

Specifically, we hypothesizes that one of four possible features could be encoded in reaction time data: transitional probability, ordinal position, word identity, and duplets.

We first combined the data from Experiment 1 (N = 33) with the data from structured sessions in Experiment 2 (N = 20) for a combined data set with greater power (N = 53). For each participant, we computed the Pearson correlation on RTs between each pair of syllables, thus generating a 12-x-12 matrix of correlation values for each participant. For each of the four analyses mentioned above, we identified those cells that would constitute the two data arrays for each analysis. Finally, we performed a random sampling of values from those pre-specified cells from all participants 200 times (with replacement, the N for each sample was equal to 4/5 times the length of the shortest of the two arrays being compared). We then computed a Wilcoxon’s rank sum test on the complete set of sampled values.

For the test of transitional probability, within values included the correlation between all pairs of word-initial syllables (TP = 0.33) versus the correlation between all pairs of word-medial and word-final syllables (TP = 1). Here we observed a significant shift in the true location difference of the two arrays. (W(53) = 1303113221, p-value < 0.0001, Cis: 0.0175, 0.0269; estimates of difference in location 0.022).

For the test of ordinal position, within values included the correlation between all pairs of word-initial syllables (e.g. nu-ro), all pairs of word-medial syllables (e.g. ga-ki), and all pairs of word-final syllables (e.g. di-se) versus correlations between syllables within each word (e.g. nu-ga, ga-di, nu-di). Here we also observed a significant shift in means between the two groups. (W(53) = 4995263642, p-value < 0.0001, Cis: 0.007771155, 0.014582123; estimates of difference in location 0.011).

For the test of word identity, within values included the correlation between syllables within each word (e.g. nu-ga, ga-di, nu-di) versus “phantom” word pairs where each item in the pair is drawn from two different words (e.g. nu-ki, nu-se). Here we observed no significant difference. (W(53) = 4891836441, p-value = 0.4957, Cis: -0.0024, 0.0002; estimates of difference in location: -1.26268e-05).

Finally, to test duplet identity, we compared values from all pairs of duplets within words (e.g. nu-ga, ga-di) versus pairs of word-initial and word-final syllables within words (e.g. nu-di). Our Wilcoxon test suggested the two groups differ in their means. (W(53) = 557892245, p-value = 0.006, Cis: 0.0007, 0.0118; estimates of difference in location: -0.0058).

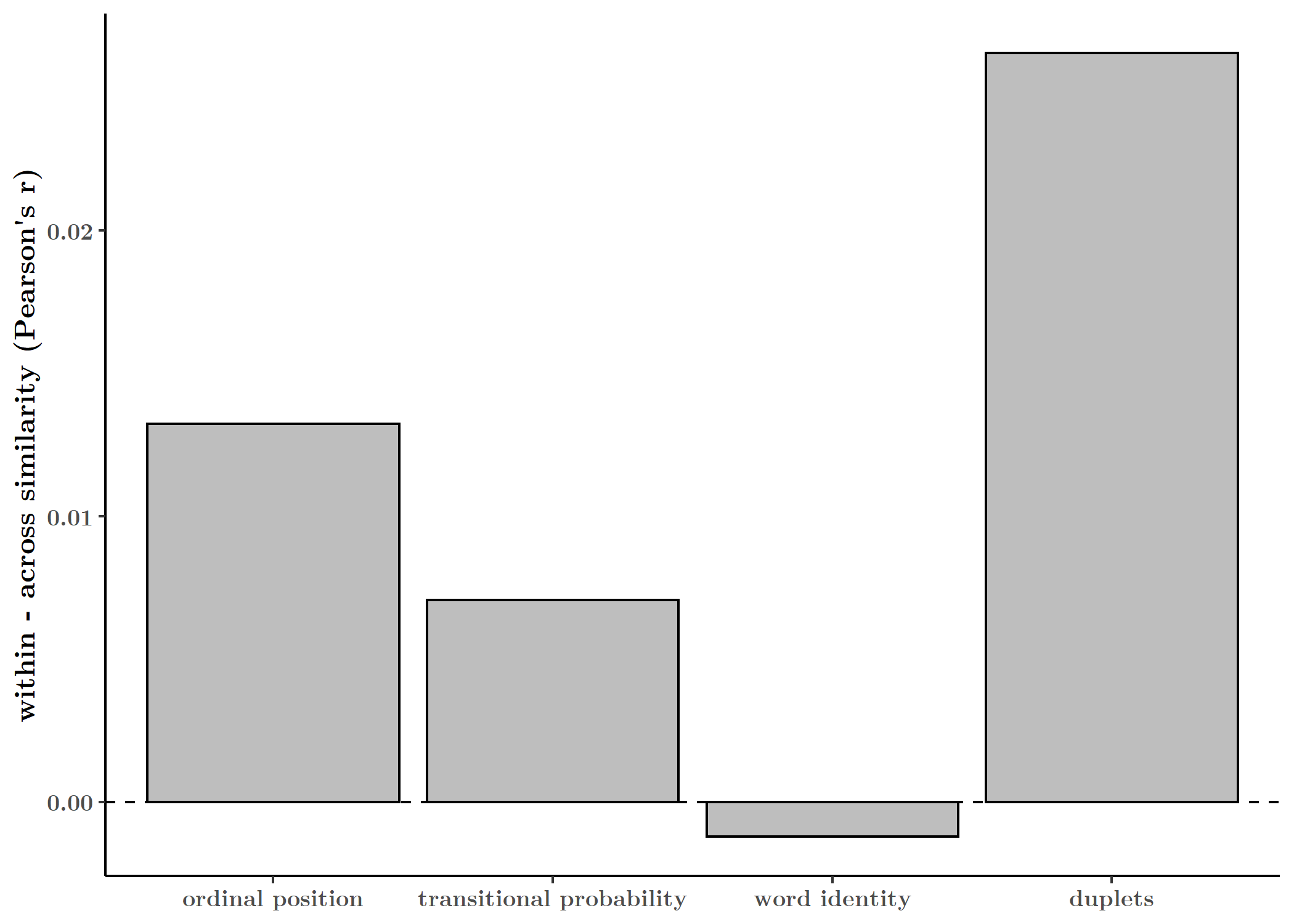


Fig. 8

# **General Discussion**

Online detection task reveals sensitivity to transitional probability, ordinal position, and duplet pairs, but not the pseudo-word units. We see a weak correlation not only because the tasks potential tap into different representations (implicit vs. explicit), but also because the target detection task doesn’t have the power to reflect word-level chunking, as it reflects only sensitivity to pairwise relationships.

# **References**

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: past, present and future. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711), 20160047. https://doi.org/10.1098/rstb.2016.0047

Batterink, L. J. (2017). Rapid Statistical Learning Supporting Word Extraction From Continuous Speech. *Psychological Science*, 1–8. https://doi.org/10.1177/0956797617698226

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/J.CORTEX.2017.02.004

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, *115*, 56–71. https://doi.org/10.1016/j.cortex.2019.01.013

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the Neural Bases of Implicit and Statistical Learning. *Topics in Cognitive Science*, 1–22. https://doi.org/10.1111/tops.12420

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning and Memory*, *22*, 544–555. https://doi.org/10.1101/lm.037986.114

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1425–1431. https://doi.org/10.1037/a0027210

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic Constraints on Statistical Computations The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science*, *16*(6). Retrieved from https://journals.sagepub.com/doi/pdf/10.1111/j.0956-7976.2005.01556.x

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, *88*(1), 2–19. https://doi.org/10.1016/j.neuron.2015.09.019

Endress, A. D., & Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, *60*. https://doi.org/10.1016/j.jml.2008.10.003

Franco, A., Eberlen, J., Destrebecqz, A., Cleeremans, A., & Bertels, J. (2015). Rapid Serial Auditory Presentation. *Experimental Psychology*, *62*(5), 346–351. https://doi.org/10.1027/1618-3169/a000295

Franco, A., Gaillard, V., Cleeremans, A., & Destrebecqz, A. (2014). Assessing segmentation processes by click detection: online measure of statistical learning, or simple interference? https://doi.org/10.3758/s13428-014-0548-x

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/J.TICS.2014.12.010

Gómez, D. M., Bion, R. A. H., & Mehler, J. (2010). The word segmentation process as revealed by click detection. *Language and Cognitive Processes*, *26*(2), 212–223. https://doi.org/10.1080/01690965.2010.482451

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Karuza, E. A., Li, P., Weiss, D. J., Bulgarelli, F., Zinszer, B. D., & Aslin, R. N. (2016). Sampling over nonuniform distributions: A neural efficiency account of the primacy effect in statistical learning. *Journal of Cognitive Neuroscience*, *28*(10), 1484–1500. https://doi.org/10.1162/jocn\_a\_00990

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247. https://doi.org/10.1016/J.COGNITION.2009.07.011

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(December), 1926–1928. https://doi.org/10.1126/science.918660

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/JMLA.1996.0032

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining “Learning” in Statistical Learning: What Does an Online Measure Reveal About the Assimilation of Visual Regularities? *Cognitive Science*, *42*, 692–727. https://doi.org/10.1111/cogs.12556

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120. https://doi.org/10.1016/J.JML.2015.02.001

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2008). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. Retrieved from www.omniglot.com.

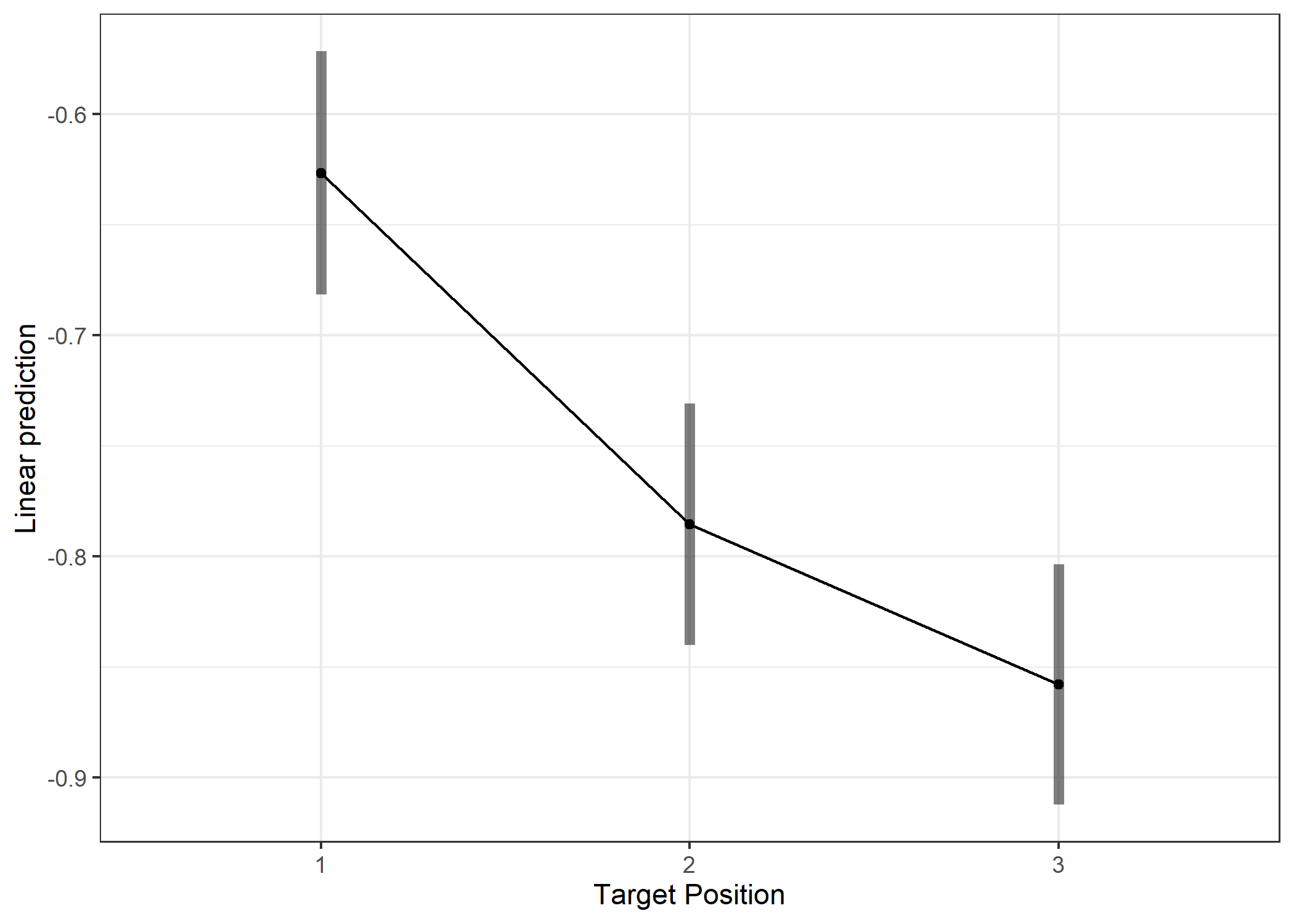
Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit Perceptual Anticipation Triggered by Statistical Learning. *The Journal of Neuroscience*, *30*(33), 11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010

Zacks, J. M., & Swallow, K. M. (2007). Event Segmentation. *Current Directions in Psychological Science*, 80–84. Retrieved from https://journals.sagepub.com/doi/pdf/10.1111/j.1467-8721.2007.00480.x

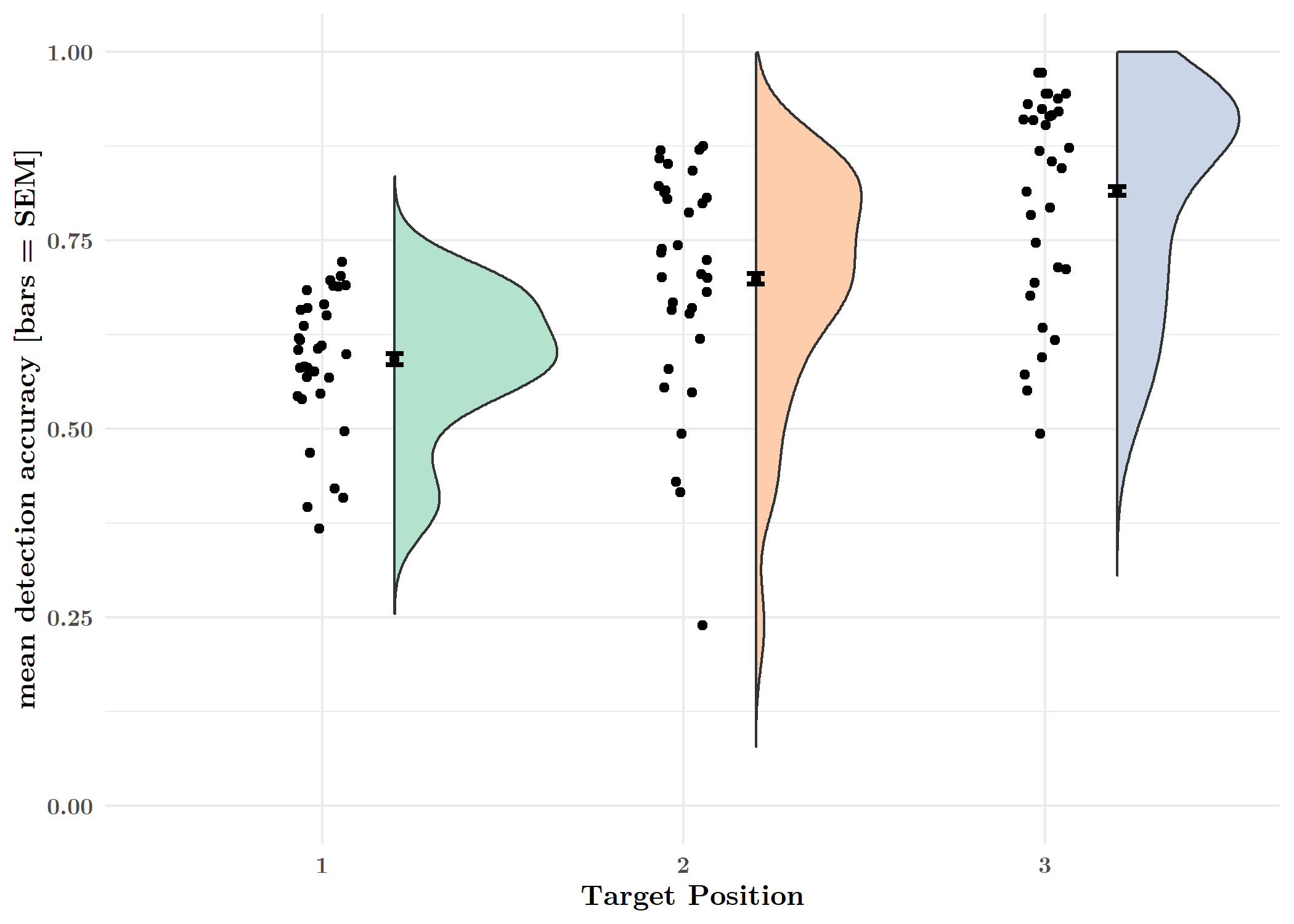
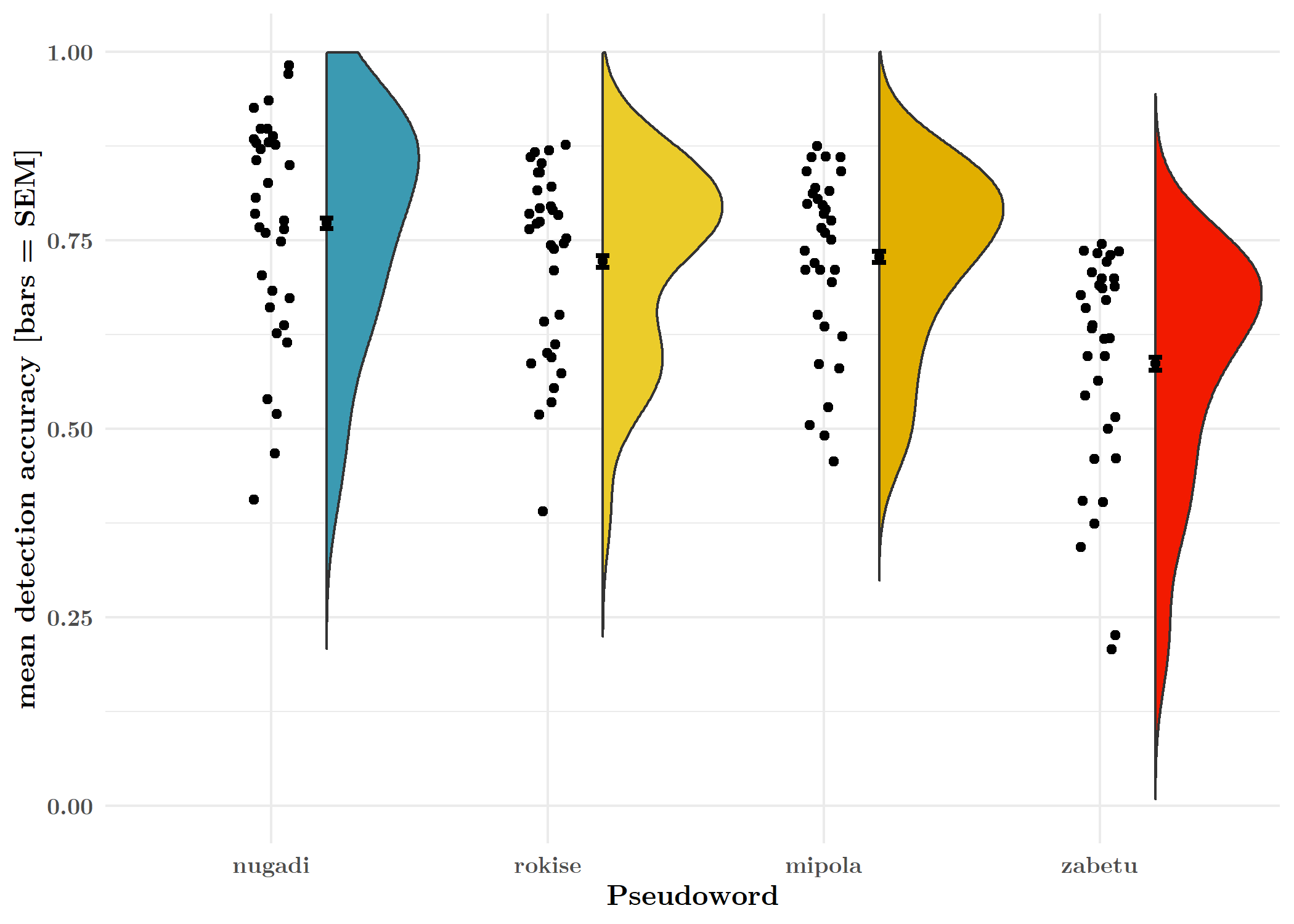
# **Supplementary Materials**

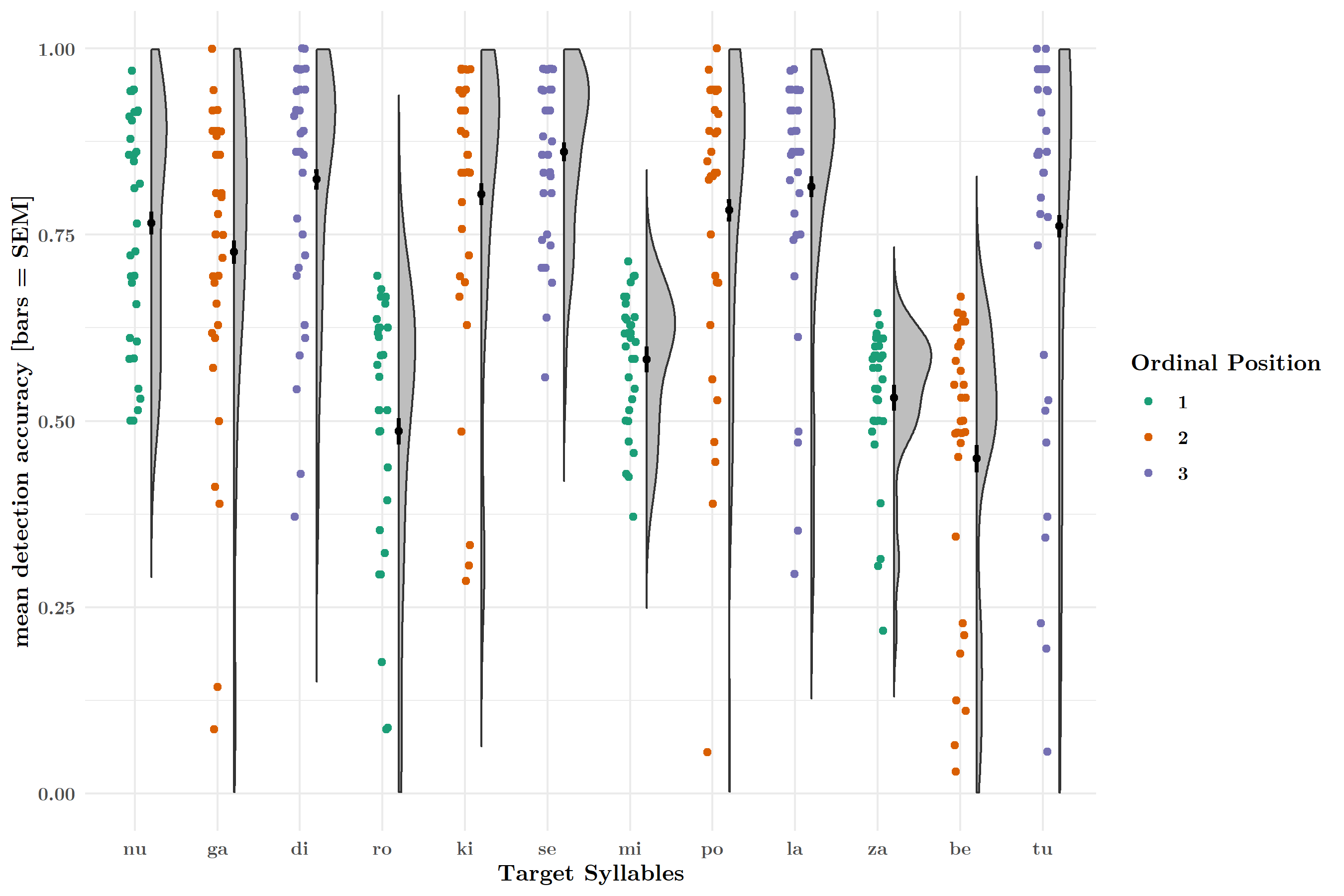
## **Stimuli**

## **Supplementary Figures, Tables**



**Fig. S1. Predicted coefficients from GLM for each ordinal position.**

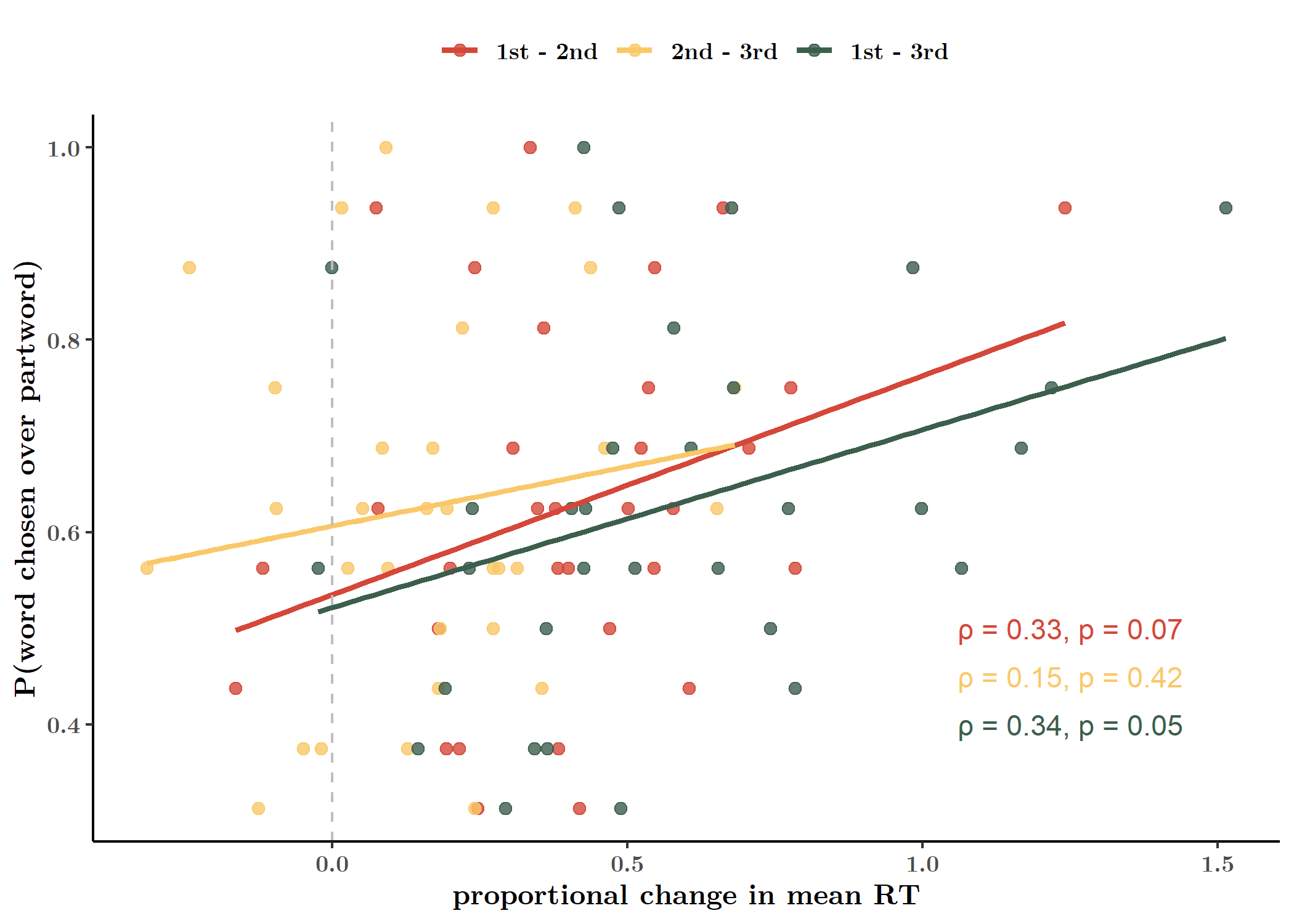
****



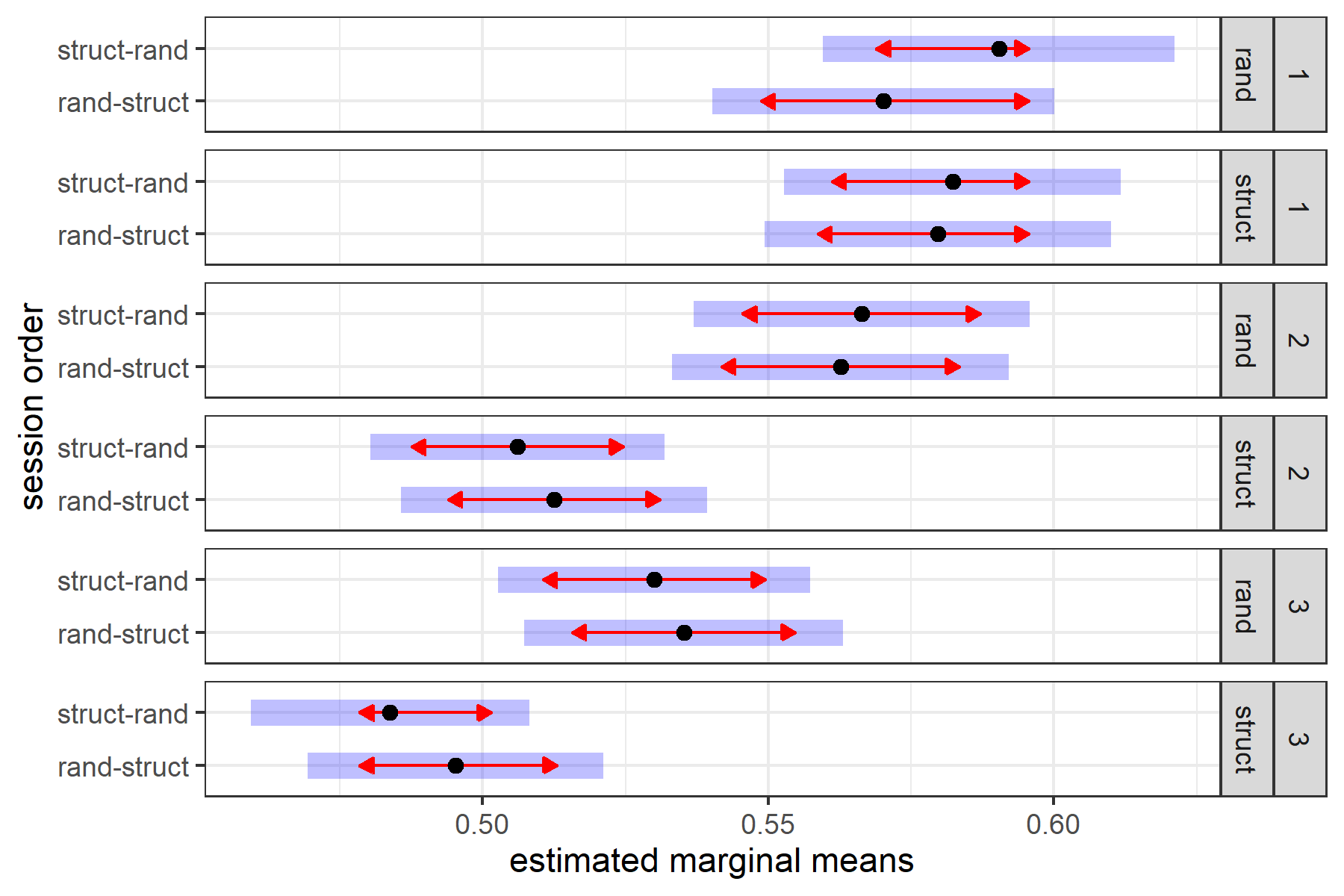
**Fig. S2. Mean detection accuracy for each ordinal position, syllable and pseudoword.**

We quantified accuracy in the target detection task to ensure participants complied with task instructions. The hit rate in Experiment 1 was 0.70 (sd = 0.45). The hit rate was also modulated by ordinal position, with each successive position having a higher mean accuracy (MPos 1 = 0.59, sdPos 1 = 0.49; MPos 2 = 0.70, sdPos 2 = 0.46; MPos 3 = 0.82, sdPos 3 = 0.39). (Fig. S2a) When averaging across all syllables in a pseudoword, accuracy varied between the four words. (Fig. S2b) This effect appears to be driven by differences in recognizability of individual syllables; certain CV syllable pairs may have been easier to detect than others, due to minor variations in stimuli acoustics. (Fig. S2c)

|  |  |  |  |
| --- | --- | --- | --- |
| **Table S1. GLM Results** | | | |
|  | | | |
|  | *Dependent variable:* | | |
|  |  | | |
|  | reaction time (s) | | |
|  | lesser | lesser (random slopes) | fuller |
|  | (1) | (2) | (3) |
|  | | | |
| Intercept(Pos 1) | -0.63\*\*\* (-0.68, -0.57) | -0.64\*\*\* (-0.68, -0.61) | -0.69\*\*\* (-0.76, -0.62) |
| Pos 2 | -0.16\*\*\* (-0.18, -0.14) | -0.14\*\*\* (-0.19, -0.10) | -0.11\*\*\* (-0.17, -0.05) |
| Pos 3 | -0.23\*\*\* (-0.25, -0.21) | -0.22\*\*\* (-0.28, -0.15) | -0.18\*\*\* (-0.24, -0.12) |
| Block 2 |  |  | 0.07\*\* (0.01, 0.13) |
| Block 3 |  |  | 0.08\*\*\* (0.02, 0.15) |
| Block 4 |  |  | 0.08\*\* (0.02, 0.14) |
| Block 5 |  |  | 0.11\*\*\* (0.05, 0.17) |
| Block 6 |  |  | 0.03 (-0.03, 0.09) |
| Block 7 |  |  | 0.05 (-0.01, 0.11) |
| Block 8 |  |  | 0.06\* (-0.005, 0.12) |
| Pos 2:Block 2 |  |  | -0.04 (-0.12, 0.04) |
| Pos 3:Block 2 |  |  | -0.03 (-0.11, 0.05) |
| Pos 2:Block 3 |  |  | -0.07\* (-0.16, 0.01) |
| Pos 3:Block 3 |  |  | -0.03 (-0.11, 0.05) |
| Pos 2:Block 4 |  |  | -0.05 (-0.13, 0.04) |
| Pos 3:Block 4 |  |  | -0.07 (-0.15, 0.01) |
| Pos 2:Block 5 |  |  | -0.06 (-0.14, 0.03) |
| Pos 3:Block 5 |  |  | -0.14\*\*\* (-0.22, -0.06) |
| Pos 2:Block 6 |  |  | -0.01 (-0.10, 0.07) |
| Pos 3:Block 6 |  |  | 0.01 (-0.06, 0.09) |
| Pos 2:Block 7 |  |  | -0.06 (-0.15, 0.02) |
| Pos 3:Block 7 |  |  | -0.09\*\* (-0.17, -0.004) |
| Pos 2:Block 8 |  |  | -0.06 (-0.14, 0.03) |
| Pos 3:Block 8 |  |  | -0.06 (-0.14, 0.02) |
|  | | | |
| Fixed Effects | Subject | Position | Subject | Subject |
| Fixed Effects Struct. | Rand. Int. | Rand. Int., Slope | Rand Int. |
| Observations | 9,531 | 9,531 | 9,531 |
| Log Likelihood | 3,199.61 | 3,312.13 | 3,224.15 |
| Akaike Inf. Crit. | -6,389.23 | -6,604.26 | -6,396.29 |
| Bayesian Inf. Crit. | -6,353.42 | -6,532.64 | -6,210.07 |
|  | | | |
| *Note:* | \*p\*\*p\*\*\*p<0.01 | | |
|  | Fitted using Gamma distribution and log link function. | | |



**Fig. S3. Pearson correlation of online and offline learning measures (mean).**

****

**Fig. S4. No interaction of session order with session and ordinal position**

|  |  |
| --- | --- |
| **Table S2. GLM Results** | |
|  | |
|  | *Dependent variable:* |
|  |  |
|  | reaction time (s) |
|  | |
| Intercept(Pos 1/Rand) | -0.54\*\*\* (-0.58, -0.50) |
| Struct | 0.002 (-0.04, 0.04) |
| Pos 2 | -0.03\*\* (-0.05, -0.01) |
| Pos 3 | -0.08\*\*\* (-0.10, -0.06) |
| Struct:Pos 2 | -0.11\*\*\* (-0.14, -0.08) |
| Struct:Pos 3 | -0.09\*\*\* (-0.12, -0.06) |
|  | |
| Fixed Effects | Session Order/Subject + Session |
| Fixed Effects Struct. | Rand. Int. + Rand. Slope |
| Observations | 6,524 |
| Log Likelihood | 4,140.28 |
| Akaike Inf. Crit. | -8,250.56 |
| Bayesian Inf. Crit. | -8,148.81 |
|  | |
| *Note:* | \*p\*\*p\*\*\*p<0.01 |
|  | Fitted using Gamma distribution and log link function. |

**Fig. S5.**

