# Experimental Results

# Measuring statistical learning by eye-tracking

**Authors:** Tamás Zolnai, Dominika Réka Dávid, Orsolya Pesthy, Marton Nemeth, Mariann Kiss, Márton Nagy, Dezso Nemeth

**Supplementary Materials**

# Supplementary methods

## Tasks

### ASRT stimuli

**Stimuli** - First, we displayed four empty circles as stimuli placeholders. These placeholders remained on the screen during the whole task (during the Response-to-next Stimulus Interval (RSI), and when the stimuli were on the screen). The background color was set to the “Ivory” color (#FFFFF0). The stimuli were “Dark Blue” (#00008B) circles at one of the placeholder positions. The diameter of the circles was 3 cm (≈ 2.64° visual angle). The four stimuli positions were at equal distances from each other (15 cm ≈ 13.16° visual angle) and the center of the screen (10.6 cm ≈ 9.32° visual angle), as shown in Figure 2.

### More details on the ASRT sequence-structure

There were 64 possible triplets, 16 of which were high-probability, and 48 were low-probability. High-probability triplets occur as pattern-ending triplets (pattern-random-pattern) in 50%, or as random-ending triplets (random-pattern-random) in 12.5% of all trials. Therefore, high-probability triplets occurred with 62.5%, while low-probability triplets with 37.5% overall probability. On the unique triplet level, high-probability triplets occur with a 4% probability (62.5%/16), while low-probability ones with 0.8% (37.5%/48).

### Explicit questionnaire

We administered a short questionnaire at the end of the Testing phase to probe whether participants gained explicit, that is, conscious knowledge about the statistical regularities underlying the ASRT task, which could have influenced both learning and consolidation processes. The questionnaire consisted of two increasingly specific questions: “Have you noticed anything special regarding the task?”, and “Have you noticed some regularity in the sequence of stimuli?”.

### Inclusion-Exclusion test

To further investigate the level of explicitness participants gained, we administered the inclusion-exclusion task (Destrebecqz et al., 2005; Destrebecqz & Cleeremans, 2001; Fu et al., 2010; Horváth et al., 2020; Jiménez & Méndez, 1999), which is based on the well-established Jacoby process dissociation procedure (PDP) test (Jacoby, 1991). In this task, we asked participants to freely generate sequences of elements in two conditions using different instructions (see later). The participants saw the same four placeholders on the screen as they saw during the ASRT task and could mark one of them by fixating and then looking away from the given position. If the subject fixated on one of the placeholders, it turned active (i.e., blue), indicating that the response was registered.

Participants performed the task under two different conditions. Both conditions were repeated four times, and each run consisted of 24 fixations. In the first condition, we first instructed participants to generate a sequence of responses that resembled the structure of the main ASRT task as much as possible, including both the random and pattern trials (inclusion condition). In this condition, implicit knowledge is sufficient to perform successfully, i.e., to produce high-probability triplets above chance level. In the second condition, we asked the participants to generate a novel sequence of responses, that is, to consciously exclude the patterns they could recognize during the ASRT task (exclusion condition). Being able to produce high-probability triplets at or below chance level requires explicit knowledge to exert control over the responses. Taken together, a high ratio of high-probability triplets in both the inclusion and exclusion conditions indicates a high-level, but implicit knowledge of the statistical structure (Horváth et al., 2020; Kiss et al., 2019; Kobor et al., 2017; Kóbor et al., 2019).

To test whether this is the case, we calculated the percentage of produced high-probability triplets in the inclusion and exclusion conditions separately. We then compared these percentages to chance level (25%: out of the 64 possible triplets, 16 are high-probability). Performing above chance in the inclusion condition can indicate either implicit or explicit knowledge about the statistical regularities. In contrast, under the exclusion condition, a below-chance ratio of high-probability triplets can be achieved solely by explicit knowledge. To access this, we ran a one-sample t-test on both conditions, comparing the mean of our sample to chance level. We also performed the same analysis excluding trills and repetitions. Trills and repetitions of three elements are by nature low-probability (they can be formed only by two random and one pattern element – r-P-r). Thus, not producing them in the inclusion and producing them in the exclusion condition could be a successful strategy. But this strategy does not reflect explicit knowledge of the complete pattern structure or triplet probabilities, excluding them can provide useful information about the strategy participants generated the sequences with. In this case, the number of all possible triplets changed to 48, while the number of high-probability triplets remains the same, meaning a chance level of 33.33%.

## Eye-tracking

### Calibration

We used the Tobii Pro Eye Tracker Manager (TODO version) for calibration. Participants saw dots appearing in a five-point grid: four points near the corners and one in the center of the screen, and we asked them to look at the dots appearing on the screen while we measured their gaze position. We validated the accuracy of calibration before both phases using a mini-block of 20 random trials. If we found extreme RTs (>1000 ms), we started the calibration process again. If we failed to reach zero extreme RTs through six recalibrations, we stopped the process and excluded the given participant.

### Calculation of dispersion value

We used the following formula to determine the dispersion value: D = [max(x) – min(x)] + [max(y) – min(y)], where D is the dispersion value, x is the horizontal, and y is the vertical coordinate of the eye position on the screen.

### Parameter selection

The algorithms described above have four parameters: DT, DuT, AOI size, and the maximum allowed missing data (MAM). Our goal in parameter selection was to be able to record accurate RTs, which requires the software to register fixations on the active stimuli as responsively as possible while keeping the noise low. To achieve this, we used a DT value of 2.8 cm (~2.5°), which is considered relatively large compared to the DT suggested in previous studies (0.5-1° in Salvucci & Goldberg, (2000), and 1.4° – 3.12° in Blignaut & Beelders, 2008). The DuT is commonly set based on the minimum duration of fixations. This allows us to separate saccades from fixations. To keep the software responsive, we set the DuT to 100 ms, which is the shortest DuT recommended in the literature (Manor & Gordon, 2003; Salvucci & Goldberg, 2000; Vakil et al., 2021). Given our 120 Hz tracking rate, this meant 12 samples for each 100 ms. We used four, 4x4 cm large squares as AOIs, the center of each square was one of the four stimuli. The smallest distance between the edge of the stimuli and AOIs was 0.5 cm (≈ 0.44°), which is a rather strict AOI. Regarding the MAM, we had two important considerations: to have more real data than interpolated and to avoid interpolating data within blinks, since during blinks no fixation can occur, thus, it is preferable to wait for new incoming data. Thus, the suggested maximum gap length is shorter than 75 ms (Olsen, 2012). We allowed a maximum of four missing samples of the 12 samples within each 100 ms fixation window, which means a maximum of 33.33% interpolated data. Consequently, our maximum gap length within a fixation window was 33.33 ms. As it is shorter than the recommended 75 ms, we ensured not to interpolate blinks.

### Outlier filtering for eye-tracking data quality

To increase the eye-tracker data quality, we filtered outliers on several metrics. First, we pooled all epochs of all participants (34 participants x 8 epochs = 272 epochs in total), and defined outlier epochs using boxplots (i.e., the lower bound of the included range was 1.5 inter-quartile distance (IQD) from the first quartile, the upper bound was 1.5 IQD from the third quartile). For expected and observed values and outlier bounds see Table 1.

Table 1

*Data quality measures*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Measurement | Expected/optimal | Observed mean | Observed SD | [lower bound, upper bound] | Number of outlier epochs |
| Precision - RMS(S2S) | M=0.23-0.52°,SD=0.06-0.13° | 0.37° | 0.08° | [0.098, 0.630] | 0 epochs |
| Precision - RMS(E2E) | - | 1.05° | 0.26° | [0.52°, 1.51°] | 17 epochs |
| Data loss | 0% | 8.3214 | 5.2616 | [0, 15.81%] | 19 epochs |
| Distance from the screen | 50-90 cm | 65.34 cm | 43.97 cm | [55.19 cm, 76.85 cm] | 8 epochs |

*Note* Expected RMS(S2S) mean, and SD are based on (Tobii AB, 2015). In the number of outlier epochs column, we represent how many of the total 272 epochs were excluded based on the filtering.

There are three commonly used measures to describe eye-tracker data quality: accuracy, precision, and data loss (Holmqvist et al., 2012). Accuracy, i.e., the metric of differences between the true eye position and the position recorded by the eye-tracker, was not directly assessed in this study. Precision refers to the consistency of the recording, that is, how close the recorded eye positions are to each other when the participant is looking at a reference point (in our case, the active stimulus). First, we calculated the commonly used RMS(S2S) score to measure precision during the experiment (see Holmqvist et al., 2012). The RMS(S2S) score shows the differences between successive samples within a fixation (meaning, when the participant is assumably gazing at the same spot). We calculated this root mean square statistic for the final fixation of each stimulus within each epoch. We did not find any outliers using this measure. However, the dispersion-based fixation algorithm used here imposes an upper bound on the RMS(S2S) score, as the DT sets the maximum difference between all samples within a fixation (12 samples), which limits the possible difference between two consecutive samples the RMS(S2S) score is based on. For this reason, we calculated another precision measure as well. Our hybrid eye selection method enabled us to use the differences in the position of the two eyes. Thus, we calculated another precision metric, the root-mean-square of the eye-to-eye distance (RMS(E2E)). We calculated the differences in the two eye’s positions for each sample within a fixation window (except for the samples that did not provide valid data for both eyes) and calculated the RMS value of these differences. This metric is more beneficial, because it is also a valid measure of precision (correlation with the RMS(S2S): *r* = 0.63, *p* < 0.001), and it is not limited by the dispersion threshold. Based on the RMS(E2E) value, we found 17 outlier epochs of five different participants (for details, see Table 1).

Data loss refers to the measurement of traceability: how many of the samples were marked invalid by the eye-tracker for any reasons such as blinking, the position of the participant, or fast head movements (Holmqvist et al., 2012). We calculated the ratio of invalid samples within each epoch. Based on this, we found 19 outlier epochs. Besides the above-mentioned factors, the participant’s distance from the screen can also affect the data quality. Eight epochs were detected as outliers based on this screen-to-eye metric and were excluded.

Taken together, to keep the RTs reliable, 43 (15.80%) of the 272 epochs were excluded based on poor eye-tracking data quality. Additionally, we could not run the experiment with 4 participants due to the failure of calibration. This ratio is not unique in the eye-tracking literature, the ratio of excluded data varies significantly in the literature, up to as high as 20-60% (Holmqvist et al., 2012; Schnipke & Todd, 2000).

# Supplementary results

### Explicitness – was the acquired statistical knowledge implicit?

According to the questionnaire we took after the Testing phase, none of the participants were able to report the exact statistical structure of the ASRT task, neither the alternating sequence nor the exact triplet structure. Most of the participants (N = 21) did not notice anything particular/any pattern in the ASRT task. 3 participants reported that there might be some pattern but were not able to explicitly phrase what. One person reported that he realized that trills were occurring less frequently than non-trill triplets.

To further investigate the level of explicitness, we analyzed the Inclusion-exclusion task. According to the standard evaluation method, we compared the generated high-probability triplets’ ratio to chance level (25%). Participants generated 4.83% more high-probability triplets in the inclusion task than the given chance level (Mhigh = 29.83%, SD = 4.61), *t*(23) = 5.13, *p* < .001, *d* = 1.048, *BF10* = 722.69), which means the subjects acquired the statistical structure of the ASRT. In contrast, in the exclusion condition, participants generated high-probability triplets on chance level: there is no significant difference compared to 25%, which is also supported by substantial evidence for H0 indicated by the Bayes Factor (Mhigh = 24.24%, SD = 8.45, *t*(23) = -0.44, *p* = .665, *d* = -0.09, *BF10* = 0.23). We also compared the two conditions to each other, where we found a significant difference (*t*(33) = -3.94, *p* < .001, *d* = 0.68, *BF10* = 72.08). To gain more insight into the explicitness level, we also analyzed the generated sequences excluding the repetitions and trills. In the inclusion condition, we found that the difference from the 33.33% chance level was nonsignificant, however, this null result was not supported by the Bayes factor (*t*(23) = 1.206, *p* = .240, *d* = 0.25 *BF10*= 0.410, Mhigh = 34.35%, SD = 4.12). In the exclusion condition, we did not find a significant difference compared to chance level (*t*(23) = 0.30, *p* = .763, *d* = -0.06, *BF10* = 0.22, Mhigh = 33.86%, SD = 8.56). The Bayes factor again indicates moderate evidence for equality (H0). Not performing above chance level in the exclusion condition raised the question of whether the learning was fully implicit. Comparing the data of the two conditions we did not find a significant difference (*t*(23) = 0.27, *p* = .788, *d* = 0.05, *BF10* = 0.22).

**

*SM Figure 1.* The results of the ASRT task without epoch-level filtering based on the data quality (see Methods section). Unfiltered RTs are presented as a function of high-probability (blue line with triangle symbols) and low-probability (orange line with square symbol) triplets throughout the epochs of the Learning phase (1-5) and the Testing phase (6-8). Note, that in the first epoch stimuli were presented randomly, and in the seventh epoch, participants performed on an interference sequence instead of the original sequence used in the 2-4th, sixth and eighth epochs. The difference between high- and low-probability triplets represents statistical learning. Error bars represent the SEM.

Table 1

*Detailed results of the ANOVA on the RT*

|  |
| --- |
| **With outlier filtering** |
| **Learning phase** |
| **Effect** | **df1, df2** | **F** | **p** | **η²p** |
| EPOCH | 2.204, 50.682 | 3.007 | 0.054 | 0.116 |
| TRIPLET | 1, 23 | 11.588 | 0.002 | 0.335 |
| EPOCH x TRIPLET | 4, 92 | 5.253 | < .001 | 0.186 |
| **Training phase** |
| **Effect** | **df1, df2** | **F** | **p** | **η²p** |
| EPOCH | 1.605, 36.913 | 6.007 | 0.009 | 0.207 |
| TRIPLET | 1, 46 | 22.313 | < .001 | 0.492 |
| EPOCH x TRIPLET | 1.388, 31.93 | 5.804 | 0.014 | 0.202 |
| **Without outlier filtering** |
| **Learning phase** |
| **Effect** | **df1, df2** | **F** | **p** | **η²p** |
| EPOCH | 1.91, 63.21 | 1.290 | 0.282 | 0.038 |
| TRIPLET | 1, 33 | 22.558 | < .001 | 0.406 |
| EPOCH x TRIPLET | 4, 13 | 4.291 | 0.003 | 0.115 |
| **Training phase** |
| **Effect** | **df1, df2** | **F** | **p** | **η²p** |
| EPOCH | 1.65, 54.60 | 3.673 | 0.040 | 0.100 |
| TRIPLET | 1, 33 | 3.856 | 0.058 | 0.105 |
| EPOCH x TRIPLET | 1.15, 66 | 4.400 | 0.038 | 0.118 |
| *Note* For means and SEM, see SM Figure 2. |

Table 2

*Detailed results of the ANOVA on the ratio of learned anticipations in the epochs of the Learning and Testing phases*

|  |
| --- |
| **With outlier filtering** |
| **Effect** | **df1, df2** | **F** | **p** | **η²p** |
| **Learning phase** |  |  |  |  |
| EPOCH | 4, 92 | 14.76 | < .001 | 0.391 |
| **Training phase** |
| EPOCH | 2, 46 | 14.47 | < .001 | 0.386 |
| **Without outlier filtering** |
| **Effect** | **df1, df2** | **F** | **p** | **η²p** |
| **Learning phase** |
| EPOCH | 4, 132 | 21.39 | < .001 | 0.39 |
| **Training phase** |
| EPOCH | 2, 66 | 18.683 | < .001 | 0.36 |
| *Note* For means and SEM, see SM Figure 2. |

**

*SM Figure 2.* The results of the ASRT task without epoch-level filtering based on data quality (see Methods section). Percentage of learned anticipation compared to all anticipatory eye movements during the ASRT task. Error bars represent the SEM. The dashed line indicates the chance level.

# Supplementary discussion

In our study, we aimed to adapt another task, the Inclusion-exclusion task (Destrebecqz & Cleeremans, 2001) to eye-tracking, which is based on the Process Dissociation Procedure (PDP, Jacoby, 1991), to the oculomotor version of the ASRT task. The Inclusion-exclusion task following the manual version of the ASRT task typically reveals a lack of explicit knowledge (Horváth et al., 2020; Kiss et al., 2019; Kobor et al., 2017; Sævland & Norman, 2016; Vékony et al., 2020). In our study, we found that participants were able to produce high-probability triplets above chance level, supporting our results of the ASRT task: participants acquired statistical knowledge of the task structure. On the other hand, participants produced high-probability triplets on chance level even when they were asked to generate a sequence that is different from the sequence they saw in the ASRT task. This result does not directly indicate explicit knowledge of the sequence structure, on the other hand, previous studies have found an above-chance ratio of high-probability triplets on the exclusion condition of this task, and no difference between the inclusion and exclusion conditions (Horváth et al., 2020; Vékony et al., 2020), unlike our findings. One possible explanation is that these differences emerged due to a slightly more explicit knowledge of the oculomotor ASRT task compared to the manual version. However, no participants were able to report neither the sequence nor any pattern-specific regularity, which questions the validity of the oculomotor version of the Inclusion-exclusion task. Moreover, considering that the difference between the conditions disappeared once we filtered for repetitions and trills, maybe our participants relied more on producing these (low-probability) triplets in the exclusion condition than participants in previous studies that used the Inclusion-exclusion task. Thus, further studies are needed to test whether the Inclusion-Exclusion task can access the explicitness of the acquired knowledge gained on the oculomotor version of the ASRT task.

# Supplementary references

Blignaut, P., & Beelders, T. (2008). The effect of fixational eye movements on fixation identification with a dispersion-based fixation detection algorithm. *Journal of Eye Movement Research*, *2*(5). https://doi.org/10.16910/JEMR.2.5.4

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*(2), 343–350. https://doi.org/10.3758/BF03196171

Destrebecqz, A., Peigneux, P., Laureys, S., Degueldre, C., Fiore, G. del, Aerts, J., Luxen, A., van der Linden, M., Cleeremans, A., & Maquet, P. (2005). The neural correlates of implicit and explicit sequence learning: Interacting networks revealed by the process dissociation procedure. *Learning & Memory*, *12*(5), 480–490. https://doi.org/10.1101/LM.95605

Fu, Q., Dienes, Z., & Fu, X. (2010). Can unconscious knowledge allow control in sequence learning? *Consciousness and Cognition*, *19*(1), 462–474. https://doi.org/10.1016/J.CONCOG.2009.10.001

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. In S. N. Spencer (Ed.), *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA ’12*. ACM Press. https://doi.org/10.1145/2168556

Horváth, K., Török, C., Pesthy, O., Nemeth, D., & Janacsek, K. (2020). Divided attention does not affect the acquisition and consolidation of transitional probabilities. *Scientific Reports 2020 10:1*, *10*(1), 1–14. https://doi.org/10.1038/s41598-020-79232-y

Jacoby, L. L. (1991). A Process Dissociation Framework: Separating Automatic from Intentional Uses of Memory. *JOURNAL OF MEMORY AND LANGUAGE*, *30*, 513–541.

Jiménez, L., & Méndez, C. (1999). Which Attention Is Needed for Implicit Sequence Learning? *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(1), 236–259. https://doi.org/10.1037/0278-7393.25.1.236

Kiss, M., Nemeth, D., & Janacsek, K. (2019). Stimulus presentation rates affect performance but not the acquired knowledge – Evidence from procedural learning. *BioRxiv*, 650598. https://doi.org/10.1101/650598

Kóbor, A., Horváth, K., Kardos, Z., Takács, Á., Janacsek, K., Csépe, V., & Nemeth, D. (2019). Tracking the implicit acquisition of nonadjacent transitional probabilities by ERPs. *Memory & Cognition 2019 47:8*, *47*(8), 1546–1566. https://doi.org/10.3758/S13421-019-00949-X

Kobor, A., Janacsek, K., Takacs, A., & Nemeth, D. (2017). Statistical learning leads to persistent memory: Evidence for one-year consolidation. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-00807-3

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, *128*(1–2), 85–93. https://doi.org/10.1016/S0165-0270(03)00151-1

Olsen, A. (2012). The Tobii I-VT Fixation Filter Algorithm description. *Tobii Technology*, *21*. www.tobii.com

Sævland, W., & Norman, E. (2016). Studying different tasks of implicit learning across multiple test sessions conducted on the web. *Frontiers in Psychology*, *7*(JUN), 808. https://doi.org/10.3389/FPSYG.2016.00808/BIBTEX

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying Fixations and Saccades in Eye-Tracking Protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications  - ETRA ’00*. https://doi.org/10.1145/355017

Schnipke, S. K., & Todd, M. W. (2000). Trials and tribulations of using an eye-tracking system. In M. M. Tremaine (Ed.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 273–274). Association for Computing Machinery. https://doi.org/10.1145/633292.633452

Tobii AB. (2015). *Accuracy and precision Test report Tobii Pro X3-120 fw 1.7.1*. https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-pro-x3-120-accuracy-and-precision-test-report.pdf

Vakil, E., Hayout, M., Maler, M., & Schwizer Ashkenazi, S. (2021). Day versus night consolidation of implicit sequence learning using manual and oculomotor activation versions of the serial reaction time task: reaction time and anticipation measures. *Psychological Research 2021*, 1–18. https://doi.org/10.1007/S00426-021-01534-1

Vékony, T., Marossy, H., Must, A., Vécsei, L., Janacsek, K., & Nemeth, D. (2020). Speed or Accuracy Instructions During Skill Learning do not Affect the Acquired Knowledge. *Cerebral Cortex Communications*, *1*(1), 1–13. https://doi.org/10.1093/TEXCOM/TGAA041