A phenomenon of contrasting explicit learning effects in verbal statistical learning

Authors: Lu Wang, Tianlin Wang, Wenbo Yu, Dandan Liang, Dandan Liang

Date: 2024-09-02T00:00:00+00:00

Abstract

When accessing the learning effect in statistical learning, participants are requested to distinguish target words from partwords or nonwords in a two-alternative forced-choice task. However, this task did not answer how individuals represent target words and foils, and thus may not be sufficient in providing an independent learning effect on the items. The current study examined the explicit learning effect for each word type with a familiarity rating task. Participants were randomly assigned to learn a continuous artificial speech in one of three conditions: a baseline, short-exposure time, or long-exposure time condition. The ratings scores and correlations across three types of words between the baseline condition and the other two learning conditions were examined. Results revealed a significant contrasting learning effect: familiarity ratings for target words were significantly higher than baseline, whereas foils' ratings were significantly lower, reflecting the explicit knowledge during the exposure phase and the metacognition during the testing process. Furthermore, the distribution of target words' rating scores tends to be more centralized in the long-exposure time condition, suggesting a new type of SL effect. This study is the first to explore explicit learning effects across word types, and provides insight regarding how to measure SL more exactly.

Full Text

Preamble

A Phenomenon of Contrasting Explicit Learning Effects in Verbal Statistical Learning

Lu Wanga, Tianlin Wangb, Wenbo Yua, Dandan Lianga,c*

a School of Chinese Language and Culture, Nanjing Normal University, Jiangsu Province, P.R. China
b University at Albany, State University of New York, New York, USA

c Interdisciplinary Research Centre for Linguistic Science, University of Science and Technology of China, Anhui Province, P.R. China

Corresponding author
Telephone: (+86) 13815866738
E-mail: ldd233@sina.com

Abstract

When assessing learning effects in statistical learning, participants are typically asked to distinguish target words from partwords or nonwords in a two-alternative forced-choice (2AFC) task. However, this task does not reveal how individuals represent target words and foils, and thus may not provide an independent measure of learning for each item type. The current study examined explicit learning effects for each word type using a familiarity rating task. Participants were randomly assigned to learn a continuous artificial speech stream under one of three conditions: a baseline condition, a short-exposure condition, or a long-exposure condition. We analyzed rating scores and correlations across the three word types, comparing the baseline condition with the two learning conditions. Results revealed a significant contrasting learning effect: familiarity ratings for target words were significantly higher than baseline, whereas ratings for foils were significantly lower, reflecting explicit knowledge acquired during the exposure phase and metacognitive processes during testing. Furthermore, the distribution of target word ratings tended to be more centralized in the long-exposure condition, suggesting a novel type of SL effect. This study is the first to explore explicit learning effects across word types and provides insight into how SL might be measured more precisely.

Keywords: statistical learning, learning effect, explicit knowledge, familiarity rating task, two-alternative forced-choice task

1. Introduction

Statistical learning (SL) refers to the capacity to recognize statistical patterns and identify cognitive units, which are fundamental aspects of cognition (Saffran et al., 1996). In this context, verbal SL has traditionally been recognized as playing a key role in segmenting words from continuous speech. While substantial research has linked SL to the development of language skills and reading abilities (e.g., Shoaib et al., 2018; von Koss Torkildsen et al., 2019; Qi et al., 2019; Frost et al., 2019; Isbilen et al., 2022; Lukács et al., 2023), some fundamental yet controversial issues remain, drawing significant attention from researchers: Is statistical learning an implicit or explicit cognitive mechanism? Is there a better measurement method than the two-alternative forced-choice (2AFC) task for adults? Does the length of exposure affect learning effects?

1.1 Is statistical learning an implicit or explicit mechanism?

The most prevalent paradigm in SL research is the "learning-testing" paradigm. In the learning/exposure phase, participants are typically asked to learn an artificial language with statistical regularities without explicit instruction. For example, in the study by Mirman et al. (2008), participants were asked to listen to a "made-up language" and "be ready to answer questions about the language," while in the study by Wang and Saffran (2014), they were told to "pay as much attention to the language pattern as possible." In the testing phase, researchers have typically used a 2AFC task for both child and adult groups, where participants are told to choose the option that sounds more familiar. Since neither the instructions in the exposure phase nor those in the test phase made it clear that there were units to be segmented from the continuous speech, the entire design is considered to tap into an implicit learning mechanism.

Recently, the view that statistical learning is not a purely implicit mechanism has become increasingly popular. For example, Batterink et al. (2015a) examined learning effects in explicit and implicit groups using a target-detection task, where participants were first exposed to a continuous stream of repeating nonsense words and were then instructed to detect a specific syllable within the speech stream. The results showed that participants in the explicit group, who received supplementary explicit training on the nonsense words, responded faster and more accurately when the target syllable was in the third position than the implicit group.

In another study, Batterink et al. (2015b) told participants in the explicit group to figure out the words' boundaries in a nonsense language and test their knowledge, whereas participants in the implicit group were simply instructed to listen to the auditory stimuli. Despite significant learning effects observed in both recognition and target-detection tasks, these effects failed to correlate with each other, underscoring the complexity of SL effects.

The confidence rating task, where participants indicate how sure they are of their response on a 4-point or 7-point scale, provides further evidence that explicit knowledge is involved in SL. One study explored how adults abstract different types of statistical regularities simultaneously and revealed that participants were more aware of their knowledge of non-adjacent regularities than adjacent regularities (Romberg & Saffran, 2013). Ordin and Polyanskaya (2021) designed a 2AFC task with three types of pairs: (1) target words against phantoms (triples that are not embedded as whole units in the familiarization input but are statistically congruent with words), (2) target words against nonwords (created by utilizing non-adjacent syllables within the artificial language), and (3) phantoms against nonwords. They then estimated participants' metacognition scores based on their 2AFC performance and confidence ratings. Not surprisingly, participants' discrimination between words and nonwords was similar to that between phantoms and nonwords, while metacognitive sensitivity was higher in trials where target words were paired with nonwords than in trials where phantoms were paired with nonwords. This indicates that participants could consciously identify the target words. These empirical studies provide direct and compelling evidence that learning effects in SL indeed encompass explicit knowledge.

1.2 Is there a better measurement method than the 2AFC task for SL effects?

Related to the discussion about the learning mechanism of SL, many researchers have claimed that different test tasks reflect different learning outcomes. Generally, performance on recognition measures like 2AFC and statistically induced chunking recall requires the actual extraction and memorization of word units during the learning phase—that is, explicit statistical knowledge. In contrast, outcomes from syllable target-detection tasks reflect the extent to which participants are sensitive to statistical information—that is, implicit statistical knowledge.

A growing body of research has confirmed low and non-significant correlations between 2AFC and target-detection tasks (e.g., Batterink, 2017), but high and significant relationships between 2AFC and statistically induced chunking recall in both adult and child groups (Isbilen et al., 2020; Isbilen et al., 2022). However, comparatively little attention has been given to the drawbacks of the 2AFC task itself.

When utilizing the 2AFC task, three distinct word types are involved: target words, partwords (formed by combining consecutive syllables from two target words), and nonwords (formed by syllabic sequences that never occurred during the exposure phase). As participants engage in the 2AFC task, they are confronted with pairs consisting of target words paired with either partwords or nonwords. Consequently, the scores derived from this task reflect a compound measure, where learning outcomes for both target words and foils (partwords and nonwords) can potentially contribute. This highlights that while the 2AFC task is used to detect SL learning effects, it only provides information about the ability to differentiate between target words and foils, lacking the ability to independently assess learning for each item type. In addition, each artificial language in an SL study is typically created with four to six target words. Thus, to increase the number of forced-choice trials, each target word must be repeated to pair with multiple foils—sometimes as many as six times, as in Batterink et al. (2015a)—introducing a secondary learning effect produced during the test phase (Siegelman et al., 2017). Especially when participants must distinguish between a word and a partword that differs by only a single syllable, repeated exposure to the same target word may increase participants' confidence in making choices, thereby exaggerating the learning effect observed during the exposure phase. In sum, when measuring explicit knowledge of SL, the 2AFC task not only has limited ability to assess the learning effects of individual word types but also comes with inherent methodological limitations.

The familiarity rating task emerges as a viable alternative to the conventional 2AFC task. This approach has already been explored in prior studies. For instance, Batterink and Paller (2017) identified a linear decline in participants' learning across three distinct word types, reflecting participants' ability to abstract statistical regularities. Additionally, Erickson et al. (2016) found that participants' performance on both the 2AFC task and the familiarity rating task was correlated in specific versions of artificial languages. Crucially, unlike the 2AFC task, the familiarity rating task allows participants to independently form explicit memory representations for different word types, enabling them to assess their familiarity with each individual item separately. Finally, the familiarity rating task requires each word to be presented only once, thereby avoiding secondary learning effects. This study adopts this approach to evaluate independent explicit knowledge of each word type.

To accurately assess these independent learning effects, a direct comparison of score differences among the three word types is insufficient. Instead, a novel approach that can more effectively isolate and analyze these effects should be devised. This study introduces a baseline condition inspired by the work of Toro et al. (2011). In this baseline condition, the artificial language was composed of nonsensical syllables from the same pool used in the experimental condition. Because none of the three word types appeared in the exposure phase of this condition, memory representations for these items are anticipated to remain at baseline levels, allowing a clean comparison of learning effects for each word type. Another rationale for incorporating a baseline condition arises from the need to address potential experiment-related effects within the artificial language learning paradigm. This concern is often addressed by employing two counterbalanced groups of participants using different learning materials. By comparing effects between these groups, researchers can attribute the experiment's impact to manipulated variables rather than to preferences for arbitrary unit combinations. In this study, the baseline condition serves to eliminate this alternative explanation; the absence of significant rating differences across the three word types would suggest that the design of the artificial language did not influence experimental outcomes. This will be initially examined.

1.3 Does the length of exposure affect learning effects?

How to design the duration of the exposure phase is also an important issue in the SL field. To ensure detectability of learning effects, many studies have employed extended exposure phases. For instance, Toro et al. (2005) repeated each nonsense word 150 times, and Wang and Saffran (2014) repeated each nonsense word 130 times in a tonal artificial language. Contradictory findings have emerged regarding the timing of SL learning effects. Recent studies have challenged the assumption that long exposure times are requisite for SL to occur, using relatively shorter exposure phases and still observing pronounced SL effects (Qi et al., 2019; Arnon, 2020). For example, adults and children as young as 7 to 9 years old displayed SL effects after only 32 repetitions, and older children aged 8 to 16 showed SL effects after 48 repetitions. In summary, the timing of when learning effects occur during the exposure phase remains an open question, and few studies have explored the trajectory of learning effects in familiarity rating tasks in the verbal SL field. The existing research landscape offers varying viewpoints on whether learning effects emerge early in the exposure phase or require extended exposure times. Further investigation is needed to clarify this aspect of SL processes, particularly in the context of verbal SL tasks.

1.4 The current study

The primary goal of the present study is to explore the independent explicit learning patterns associated with three distinct word types and to examine how explicit knowledge changes along with exposure time. To do so, the study manipulated three verbal SL conditions:

1. **Baseline Condition**: This served as a reference point, involving randomly synthesized syllables without any occurrence of target words or partwords.

2. **Short-Exposure Learning Condition (SEL)**: Each word was repeated 45 times during the artificial language exposure.

3. **Long-Exposure Learning Condition (LEL)**: Each nonsense word was repeated 90 times within the artificial language.

The study design employed a mixed-method approach, incorporating both within-subject (word type) and between-subject (learning condition) variables. As an exploratory study, we did not formulate a priori hypotheses. However, based on the discussions above, our planned analyses were as follows: (1) examine the non-significant learning effect in the baseline condition, (2) compare learning effects for the three word types between the SEL condition and the baseline condition to assess explicit effects in verbal SL, and (3) compare learning effects for the three word types between the SEL and LEL conditions to observe changes in the explicit effect over time.

2. Method

2.1 Participants

One hundred forty-four native Mandarin speakers (age range: 18-28; 123 females) were recruited from a university in Southeast China. Participants were randomly assigned to the long-exposure learning condition (49 participants), short-exposure learning condition (49 participants), or baseline condition (46 participants). All participants were right-handed, had no formal musical training, and were not majoring in foreign languages. The experiment was approved by the Institutional Review Board of the institution, and all participants signed informed consent forms before beginning the experiment.

2.2 Materials

Research on statistical learning in tone languages remains relatively scarce. The Mandarin phonological system includes 413 syllables and four lexical tones (Tone 1, Tone 2, Tone 3, and Tone 4), which combine into approximately 1,522 tonal syllables. For our experiment, we followed the design of Gómez et al. (2017) and selected twelve syllables with Tone 1 to create nonsense tonal syllables, thereby minimizing the influence of different statistical information carried by tone and segmental structure. The syllables were recorded in a sound-attenuating room at 44,100 Hz with 16-bit precision. All twelve syllables were then normalized for duration (350 ms), mean pitch (266 Hz), and intensity (70 dB) using Praat software.

The 12 target syllables were randomly combined to form six disyllabic nonsense words, as this type of word is most common in Mandarin Chinese. We then created partwords and nonwords based on the target words. The within-word transitional probability (TP) at the syllabic level for target words was 1.0, while the TP of syllables spanning word boundaries was 0.2. Nonwords consisted of two syllables that never co-occurred during exposure, giving them a within-word TP of 0. The three types of nonsense words are shown in Table 1 .

The artificial languages in the two learning conditions were created from the same pool of target words. In the LEL condition, six target words were used to create an artificial language stream containing 90 tokens of each target word and 18 tokens of each partword. The same six words were used to create the artificial language in the SEL condition, which contained 45 tokens of each target word and nine tokens of each partword. The LEL and SEL streams were concatenated using a Praat script into a pseudorandom sequence that ensured the same word could not occur twice in succession. In the baseline condition, the artificial language was created from the same syllables used in the other two conditions but did not contain any disyllabic words. The artificial languages lasted approximately six minutes in the LEL condition and three minutes in the baseline and SEL conditions. The test items were identical across all three conditions, with a total of 18 items across the three word types.

2.3 Procedure

All participants were told they would hear an artificial language through headphones and would later be tested on their knowledge of that language. They then listened to the artificial language for either six or three minutes in a soundproof booth. Following this exposure phase, participants completed a 6-point Likert scale familiarity rating task (1 = not familiar at all, 6 = very familiar). Participants first completed two practice trials, then 18 test trials. On each trial, participants rated the familiarity of an item based on the artificial language they had just heard (see Fig. 1 [Figure 1: see original paper]). All three word types (six target words, six partwords, and six nonwords) appeared only once, with presentation order randomized across participants. No additional constraints were imposed on participants' responses, and all trials were included in the analyses. The entire experiment was conducted using E-Prime 3.0 and lasted approximately 10 minutes across all three conditions.

3. Results

3.1 Learning Effect in Baseline Condition

We first investigated whether a learning effect existed among the three word types in the baseline condition. To assess this, we used a Linear Mixed Model (LMM) with the lmer function in R (version 4.3.1), with word type as a fixed effect and both subject and item as random intercepts1. ANOVA results showed no significant main effect of word type ($F(2,15) = 0.37$, $p = 0.70$), suggesting no substantial differences in learning among the three word types. Furthermore, the analysis revealed no significant fixed effects when comparing familiarity ratings between target words and partwords (β = -0.16, t = 0.62, p = 0.55) or between partwords and nonwords (β = -0.05, t = 0.19, p = 0.85). These findings indicate that under the baseline condition, participants showed no discernible learning effect or preference for any of the three word types based on their familiarity ratings.

3.2 SEL Learning Effect Compared with Baseline Condition

Another LMM was established with condition and word type as fixed effects, item as a random slope, and both item and subject as random intercepts2. One nonword in the SEL condition was incorrectly designed for 11 participants, resulting in only 17 trials for these participants, while others completed 18 trials. ANOVA results indicated a significant main effect of condition ($F(1,47.86) = 6.50$, $p = 0.014$) and a significant interaction between condition and word type ($F(2,15.00) = 17.53$, $p < 0.001$). Unstandardized coefficients for fixed effects are presented in Table 2 . We then conducted post-hoc analyses using the emmeans function, with p-values Bonferroni-adjusted for pairwise comparisons involving more than two levels.

For target words, rating scores in the SEL condition (M = 4.57) were significantly higher than those in the baseline condition (M = 4.11), t = 2.29, β = 0.46, p = 0.03. In contrast, participants rated partwords as more familiar in the baseline condition (M = 4.27) than in the SEL condition (M = 3.49), t = 3.86, $\beta = 0.78$, $p < 0.001$. A similar pattern emerged for nonwords (baseline: M = 4.32, SEL: M = 3.53), t = 3.93, $\beta = 0.79$, $p < 0.001$. See Fig. 2 [Figure 2: see original paper] for a visualization of rating patterns in the baseline and SEL conditions. These results suggest that participants began to demonstrate a learning effect after the exposure phase, but the strength of this effect varied by word type.

3.3 LEL Learning Effect Compared with SEL Condition

We employed a similar method to examine explicit learning effects between SEL and LEL conditions. In this LMM, word type and learning condition were treated as fixed effects, subjects as random intercepts, and items as both random intercepts and random slopes3. Results showed that only the main effect of word type was significant ($F(2,15.03) = 17.63$, $p < 0.001$), while the main effect of condition ($F(1,56.17) = 1.22$, $p = 0.28$) and the interaction ($F(1,14.99) = 0.11$, $p = 0.89$) were not significant. Similarly, only two fixed effects reached significance: partwords versus target words ($t = -4.48$, $\beta = -1.08$, $p < 0.001$) and nonwords versus target words ($t = -4.33$, $\beta = -1.04$, $p < 0.001$). Unstandardized coefficients for fixed effects are shown in Table 3 . See Fig. 3 [Figure 3: see original paper] for a visualization of rating patterns in the SEL and LEL conditions.

3.4 Correlations Across Word Types in Three Conditions

We further conducted a series of correlation analyses to characterize learning effects under different conditions. In the baseline condition, participants' rating scores for target words correlated significantly with those for partwords ($r = 0.33$, $p = 0.02$) and nonwords ($r = 0.64$, $p < 0.001$), and the correlation between partwords and nonwords also reached significance ($r = 0.49$, $p < 0.001$). This correlation pattern changed dramatically in the SEL condition: only partword and nonword ratings correlated significantly ($r = 0.67$, $p < 0.001$), while the other two correlations were no longer significant (target words and partwords: $r = 0.08$, $p = 0.59$; target words and nonwords: $r = 0.22$, $p = 0.13$). Finally, we replicated this analysis in the LEL condition and found that target word ratings again correlated significantly with partword and nonword ratings (target words and partwords: $r = 0.43$, $p < 0.01$; target words and nonwords: $r = 0.46$, $p < 0.01$), while partword and nonword ratings maintained a strong correlation ($r = 0.71$, $p < 0.001$). The shifting correlation patterns across the three conditions indicate that significant changes in learning effects emerged from the SEL to LEL condition, even though mean rating scores did not change. All p-values in these correlation analyses were Bonferroni-adjusted for multiple comparisons. See Fig. 4 [Figure 4: see original paper] for a visualization of correlation patterns across the three learning conditions.

4. Discussion

While previous research has used the 2AFC task to establish explicit learning effects in verbal SL tasks across various participant groups, few studies have explored whether other tasks, such as familiarity rating tasks, could provide new insights into the components of learning effects in statistical learning. The current study included a baseline condition to investigate explicit learning effects for different word types in verbal SL tasks. The findings revealed an intriguing contrasting pattern: participants showed higher familiarity with target words and reduced familiarity with foils (both partwords and nonwords) in the short-exposure condition compared to baseline. Furthermore, this explicit effect did not change with exposure time, indicating that the learning pattern was independent of exposure duration.

4.1 The Contrasting Learning Effect of Targets Versus Foils

Our initial assumption was that target words and partwords would receive higher familiarity ratings due to repeated exposure during the learning phase, similar to the linear decline across three word types reported by Batterink and Paller (2017), while nonwords would maintain consistent familiarity ratings across both learning and baseline conditions due to their absence from exposure. However, the results differed significantly from these expectations. Target words received significantly higher familiarity ratings in the short learning condition, whereas partwords and nonwords showed substantial decreases in familiarity ratings from baseline to the SEL condition.

To our knowledge, previous studies have primarily focused on whether participants can recognize target words in 2AFC tasks, with little attention given to the learning effects for nonwords and partwords. As nonwords were deliberately excluded and partwords occurred infrequently during the artificial language exposure, their lower familiarity ratings suggest that participants either recognized these items as not being stored in memory or noticed that they deviated from the statistical regularities of the artificial language. We propose that a self-awareness cognitive mechanism may have contributed to these results during the test phase. The artificial languages used in the current study were arguably easier than those in previous studies, as the TPs between target words were 0.2, which is lower than in other studies (e.g., Batterink, 2017). The disyllabic nonsense units also likely placed less pressure on working memory, which is an important factor in SL task difficulty (Palmer & Mattys, 2016). Given the relative simplicity of the task for adults and the instruction that learning would be assessed after exposure, participants likely consciously memorized the target words and ignored other constructions during exposure. Consequently, when partwords and nonwords were presented during the test phase without time constraints, participants could recall target words and were therefore consciously aware that these items had not been learned.

In other words, our familiarity rating results demonstrated that these two foil types underwent a similar cognitive process: a suppression effect from baseline to the SEL condition. This self-awareness mechanism may be considered a form of metacognition: participants knew what they had or had not learned, allowing them to explicitly reject foils and resulting in lower familiarity ratings compared to baseline. This explanation aligns with recent evidence showing that nonwords accepted by participants corresponded with lower confidence ratings, whereas rejected nonwords corresponded with higher confidence ratings (Polyanskaya, 2022). The metacognition defined in this study is not exactly the same as the explicit mechanism proposed in previous studies, which provided explicit instructions or supplemental training to participants (Batterink et al., 2015a; 2015b), making the explicit mechanism run through both learning and testing phases. In contrast, this study used implicit instructions—participants were only told that their knowledge would be tested—so this type of metacognition is more likely to appear at the testing stage.

Unlike foil ratings, target word ratings in the SEL condition reached approximately 4.8 on a 6-point scale, which is significantly higher than in the baseline condition. This suggests that participants successfully recognized target words by tracking statistical regularities during exposure. Combined with the foil results, it appears that the dual nature of target word explicit memory leads to the contrasting rating patterns: recognition of target words alongside rejection of partwords and nonwords.

Beyond group-level analyses, the correlations between word types across baseline and SEL conditions deserve attention. The high correlations between each pair of word types in the baseline condition indicate that participants showed no clear preference for any specific type but instead based their ratings on syllable familiarity, supporting the validity of the baseline condition. In the SEL condition, the sharp changes in correlations between target words and partwords, as well as between target words and nonwords, suggest that participants consciously categorized items into two groups: targets and foils. This resulted in high correlations within the same category and low correlations between categories. In sum, our results demonstrate a distinction in familiarity ratings between target words and foils in verbal SL, reflecting a combination of explicit knowledge of target words obtained during the exposure phase and metacognitive processes during the test phase.

Previous research has suggested that learning effects observed in recall-based tasks like the 2AFC reflect a composite of different abilities rather than isolating the targeted cognitive process (Frost et al., 2015; Isbilen et al., 2020), and are also susceptible to individual decision-making strategies (Christiansen, 2019; Isbilen & Christiansen, 2022). Because familiarity tasks allow participants to reflect on what they learned during exposure, our findings provide direct evidence for this notion: when participants face a forced-choice trial between a partword and a target word, they may make decisions by eliminating foils rather than actively recognizing target words, or by employing both mechanisms. This contrasting pattern of familiarity ratings across word types also adds depth to our understanding of the components underlying learning effects traditionally measured by the 2AFC task. Based on our results, the correctness of each 2AFC trial can be deconstructed into two parts: explicit memory of target words and metacognitive knowledge about what has been learned. This suggests that learning effects measured by the 2AFC task might overestimate the actual statistical knowledge gained during the exposure phase.

4.2 Changes in Learning Effects with Exposure Time

Another key aim of the present study was to determine whether an extended exposure phase could enhance learning effects across word types. Notably, familiarity ratings for target words in the SEL condition, where each word was repeated 45 times, were already significantly higher than baseline. Intriguingly, when exposure time was doubled in the LEL condition (90 repetitions per word), no additional learning effect was observed compared to the SEL condition.

This rapid learning effect aligns with findings from other online studies. For instance, Batterink (2017) demonstrated that after just one exposure to words within continuous nonsense speech, participants showed faster reaction times to final syllables compared to initial syllables. Similarly, Siegelman and colleagues (2018) used a self-paced SL paradigm in the visual modality and found that learning effects followed a logarithmic function, with improved learning rates after as few as seven repetitions of each triplet. By situating these outcomes within the broader context, the present study adds to the growing body of evidence supporting rapid learning in both visual and verbal SL tasks. The consistency between these findings further emphasizes the intriguing nature of rapid learning effects within SL processes.

Interestingly, the correlation between target word ratings and part-word/nonword ratings returned to significance in the LEL condition, similar to the baseline pattern. However, we do not believe this reflects the same cognitive mechanism. This outcome appears directly related to the reduced range of target word rating scores in the LEL condition, as shown in the third row of Figure 3 [Figure 3: see original paper]. Mean rating scores remained unchanged, but participants rated target words with greater consistency, leading to a steeper distribution. Although we suspect that the significance of this correlation may be an artifact, its presence also hints at a distinct learning effect stemming from extended exposure time. In conclusion, these data reveal a previously unobserved learning effect: extended learning time does not alter the average familiarity ratings but does increase their consistency. Future research should investigate this unique effect, addressing key questions such as: When does this type of learning effect occur? Is it implicit or explicit knowledge?

5. Conclusions

Statistical learning is widely regarded as a process that incorporates both implicit and explicit mechanisms. The results of our familiarity rating task provide valuable insights into the nature of independent learning effects, which vary by word type and highlight the influence of explicit knowledge acquired during the exposure phase and metacognition during the test phase. At the same time, we have discovered a new pattern in how SL effects vary with learning time. This study is the first to explore explicit knowledge of SL using a familiarity rating task and emphasizes the need to reassess the components of learning effects as measured in previous studies.

Notes

1 We started model trimming from a full model (i.e., stimulus item set as both random slope and intercept), but found that model fit was only achieved when item was set as random intercept only. All models from the current study can be found in the Rmd file on the OSF website.

2 The model was established with treatment contrast, with both target words and baseline condition set as baseline levels in the LMM. We constructed four models total; only the model with subjects as random intercept and items as random slope and intercept, and the simplest model without any random slopes, exhibited no singular values and were successfully fitted. Furthermore, the difference between these two models was not statistically significant ($^2 = 3.99$, $p = 0.14$). To more accurately assess experimental effects, we retained the model with random slopes as our final model.

3 The model was established with treatment contrast, with both target words and SEL condition set as baseline levels in the LMM. We constructed four models total; only the model with participants as random intercept and items as random slopes and intercept, and the simplest model without any random slopes, exhibited no singular values and were successfully fitted. Furthermore, the difference between these two models was not statistically significant ($^2 = 4.32$, $p = 0.11$). To more accurately assess experimental effects, we retained the model with random slopes as our final model.

4 To compare differences in rating score variability, we conducted a test for homogeneity of variance. For target words, variance in the LEL condition was marginally significantly higher than in the SEL condition ($F(1,96) = 2.86$, $p = 0.09$). For partwords and nonwords, results were not significant (partwords: $F(1,96) = 1.38$, $p = 0.24$; nonwords: $F(1,96) = 1.54$, $p = 0.22$).

Acknowledgments

This work was supported by the Social Science Foundation of Jiangsu Province Higher Education Institutions [2022SJYB2051] and the Initial Scientific Research Fund of Nanjing Normal University [184080H202A121].

Declaration of Interest Statement: The authors report no competing interests to declare.

Data Availability Statement: The data supporting the findings of this study are openly available in OSF at [https://osf.io/xh6ju/?view_{only}=6f1659f166934a47b4f5494aa4025dd1].

References

Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Be-*

havior Research Methods, 52, 68-81. https://doi.org/10.3758/s13428-019-01205-5

Batterink, L. J. (2017). Rapid statistical learning supporting word extraction from continuous speech. *Psychological Science, 28*(7), 921-928. 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., Reber, P. J., & Paller, K. A. (2015a). Functional differences between statistical learning with and without explicit training. *Learning & Memory, 22*(11), 544. https://doi.org/10.1101/lm.037986.114

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015b). 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

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science, 11*(3), 468-481. https://doi.org/10.1111/tops.12332

Erickson, L. C., Kaschak, M. P., Thiessen, E. D., & Berry, C. (2016). Individual differences in statistical learning: Conceptual and measurement issues. *Collabra, 2*(1). https://doi.org/10.1525/collabra.41

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

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 145*(12), 1128-1153. https://doi.org/10.1037/bul0000210

Gómez, D. M., Mok, P., Ordin, M., Mehler, J., & Nespor, M. (2017). Statistical speech segmentation in tone languages: The role of lexical tones. *Language & Speech, 61*(1), 84-96. https://doi.org/10.1177/0023830917706529

Isbilen, E. S., & Christiansen, M. H. (2022). Statistical learning of language: A meta-analysis of years of research. *Cognitive Science, 46*(9), e13198. https://doi.org/10.1111/cogs.13198

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition.* https://doi.org/10.1016/j.cognition.2022.105123

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science, 44*(7). https://doi.org/10.1111/cogs.12848

Lukács, Á., Dobó, D., Szőllősi, Á., Németh, K., & Lukics, K. S. (2023). Reading fluency and statistical learning across modalities and domains: Online and offline measures. *PLOS ONE, 18*(3), e0281788. https://doi.org/10.1371/journal.pone.0281788

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*(1). https://doi.org/10.1016/j.cognition.2008.02.003

Ordin, M., & Polyanskaya, L. (2021). The role of metacognition in recognition of the content of statistical learning. *Psychonomic Bulletin & Review.* https://doi.org/10.3758/s13423-020-01800-0

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology, 69*(12). https://doi.org/10.1080/17470218.2015.1112825

Polyanskaya, L. (2022). Cognitive mechanisms of statistical learning and segmentation of continuous sensory input. *Memory & Cognition, 50*(5). https://doi.org/10.3758/s13421-021-01264-0

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading, 23*(1). https://doi.org/10.1080/10888438.2018.1485680

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science, 37*(7). https://doi.org/10.1111/cogs.12050

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

Shoaib, A., Wang, T., Hay, J. F., & Lany, J. (2018). Do infants learn words from statistics? Evidence from English-learning infants hearing Italian. *Cognitive Science, 42*(8), 3083-3099. https://doi.org/10.1111/cogs.12673

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods, 49*(2), 1-15. 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*(3), 692-727. https://doi.org/10.1111/cogs.12556

Toro, J. M., Pons, F., Bion, R. A. H., & Sebastián-Gallés, N. (2011). The contribution of language-specific knowledge in the selection of statistically coherent word candidates. *Journal of Memory and Language, 64*(2). https://doi.org/10.1016/j.jml.2010.11.005

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), B25-B34.

https://doi.org/10.1016/j.cognition.2005.01.006

von Koss Torkildsen, J., Arciuli, J., & Wie, O. B. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and Individual Differences, 69*, 60-68. https://doi.org/10.1016/j.lindif.2018.11.003

Wang, T. L., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.00953

Tables

Table 1. Test items in three SL conditions

Word Type	Items
Target words	mei1nei1 [me 1ne 1], rao1dia1 [ra 1t a1], ruo1se1 [r 1se1], lai1fo1 [la 1fo1], te1nue1 [t'e1ny 1], re1rou1 [re1ruo1]
Partwords	se1mei1 [sə1me 1], dia1re1 [t a1re1], nei1te1 [ne 1t'e1], nue1ruo1 [ny 1ruo1], fo1rao1 [fo1ra 1], rou1lai1 [ruo1la 1]
Nonwords	rao1re1 [ra 1re1], ruo1lai1 [r 1la 1], mei1te1 [me 1t'e1], se1nei1 [sen1e 1], nue1rou1 [ny 1ruo1], fo1dia1 [fo1t a1]

Table 2. Unstandardized coefficients of fixed effects in LMM (estimate, SE, t-value, and p-value)

Fixed effect	Estimate	SE	t-value	p-value
Intercept				
Condition: SEL				
Word type: partword				
Word type: nonword				
Condition × partword				
Condition × nonword				
SEL condition: word	< 0.001			
SEL condition: word	< 0.001			

Table 3. Unstandardized coefficients of fixed effects in LMM (estimate, SE, t-value, and p-value)

Fixed effect	Estimate	SE	t-value	p-value
Intercept				
Condition: SEL				
Word type: partword				
Word type: nonword				
Condition × partword				
Condition × nonword				
SEL condition: word	< 0.001			
SEL condition: word	< 0.001			

Figures

Fig. 1 [Figure 1: see original paper] Schematic representation of three conditions of verbal SL task

Fig. 2 [Figure 2: see original paper] Familiarity ratings across word types in baseline and SEL conditions

Fig. 3 [Figure 3: see original paper] Familiarity ratings across word types in SEL and LEL conditions

Fig. 4 [Figure 4: see original paper] Correlations of familiarity ratings across word types in three conditions

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.