Linking Verbs to Syntax: Investigating Error-Based Learning using Pupillometry

Malathi Thothathiri (malathi@gwu.edu)
Department of Speech Language and Hearing Sciences, The George Washington University
Washington, DC 20052 USA

Vsevolod Kapatsinski (vkapatsi@uoregon.edu)
Department of Linguistics, 1290 University of Oregon
Eugene, OR 97403 USA

Jordan Becker (jordanbecker@gwmail.gwu.edu)
Department of Speech Language and Hearing Sciences, The George Washington University
Washington, DC 20052 USA

Abstract

Verbs show different statistical preferences for syntactic structures. An influential theory for how verb-syntax links are learned suggests that learning is based on, and proportional to, prediction error. However, the evidence is mixed and there is need for evidence from a diverse set of paradigms. We exposed 90 college-aged adults to an artificial language containing novel verbs and sentence structures and tested their production of utterances in the new language. We reversed verb-syntax links from one training block to another and found that participants were able to learn the reversal, as seen in a production test. Pupillometry detected surprise when participants heard sentences in the second training block that had the opposite verb-syntax links than in the first training block. Despite detecting both surprise and successful learning, we did not find an association between the two. Thus, we did not find evidence that learning was based on surprise. We discuss alternative learning mechanisms that can help language users adapt their language based on the input.

Keywords: prediction error, argument structure, reversal, artificial language, statistical learning, surprise

Introduction

Verb bias refers to the relative likelihood of a verb appearing in different possible sentence structures. For example, the verb *give* is biased towards appearing more often in double-object dative sentences (*John gave Lily the book*) than prepositional-object dative sentences (*John gave the book to Lily)* while *bring* is biased in the reverse direction. Children and adults show evidence for implicitly learning such biases from the language input, which then guide subsequent sentence comprehension and production (Lin & Fisher, 2017; Peter et al., 2015; Ryskin et al., 2017; Snedeker & Trueswell, 2004; Thothathiri, Evans & Poudel, 2017; Wonnacott, Newport & Tanenhaus, 2008).

Usage-based theories of language emphasize the dynamic nature of lifelong language acquisition, wherein language users continually adjust to ongoing input. For example, after being exposed to training in an artificial language containing novel verbs with different verb biases, adults show a tendency to subsequently use those verbs in the structures with which they were statistically associated during training (Thothathiri & Rattinger, 2016; Wonnacott et al, 2008). Intriguingly, this tendency to dynamically learn or adapt verb biases also applies to the participants' native language. Short lab-based training sessions containing sentences in the participants' native language can detectably alter subsequent sentence comprehension and production in the direction of the verb biases present during the experiment (Ryskin et al., 2017; Thothathiri et al., 2017).

What is the mechanism underlying this dynamic learning? Error-based learning is one influential proposal (Chang et al., 2000; Jaeger & Snider, 2008). Under this account, listeners predict upcoming words or structures as the sentence unfolds. Cases where what is heard does not match the prediction give rise to error, which in turn leads to learning. Thus, surprise arising from a mismatch between the current input and the learner's prediction based on prior input is a key driver of learning in this framework (Ellis, 2006; Ramscar et al., 2010).

Error-based learning was initially proposed as an account of cue competition effects (Rescorla & Wagner, 1972), including the blocking effect (Kamin, 1969) and the inverse base rate effect (Rescorla, 1968). These effects have also been reported in language learning (Ellis & Sagarra, 2010; Nixon, 2020; Ramscar et al., 2010, 2013). However, cue competition effects have also been difficult to replicate (Maes et al., 2016) and it is possible to account for cue competition without positing that surprising observations result in faster learning (Stout & Miller, 2007). This has led researchers to look for evidence of surprise being linked to faster learning independently of cue competition (e.g., Olejarczuk, Kapatsinski & Baayen, 2018).

In syntax, much evidence in favor of error-based learning comes from surprisal effects within a priming paradigm. Surprisal is defined as $-\log(p)$, where p is the probability of the structure given the verb. Surprisal quantifies the unexpectedness of the structure. Structural priming refers to a tendency to repeat a recently experienced sentence structure. For example, hearing a double-object dative sentence makes participants more likely to produce a double-object than a prepositional-object dative sentence and vice versa (Bock, 1986; Chang et al., 2000). With respect to error-based learning, hearing a verb that is biased towards one structure (e.g., prepositional-object dative) produced in an

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

alternative structure (e.g., double-object dative) can lead to surprise and stronger priming/learning than hearing a verb in the less surprising structure (Jaeger & Snider, 2008; Fazekas et al., 2020).

The empirical evidence for prime surprisal effects is mixed, however. A recent review found that the prime surprisal effect was found in less than half of the studies that tested it (Fazekas, Sala & Pine, 2024a). Using a much larger sample than previous experiments, a recent study found strong Bayesian evidence against a prime surprisal effect (Fazekas, Blything & Ambridge, 2024b) and suggested a need for additional paradigms to test the role of surprise in linking verbs to syntax.

We employed a reversal learning paradigm and pupillometry to evaluate (1) whether adults show evidence for dynamically adapting verb-syntax associations in their behavior; (2) whether they show surprise, operationalized as pupil dilation, when previously learned verb biases are reversed; and (3) whether individual differences in surprise correlate with learning rate, which would provide evidence for error-based learning. Pupil dilation has previously been shown to be sensitive to the predictability or unpredictability of auditory stimuli (see Zekveld, Koelewijn & Kramer, 2018, for a review). Therefore, we hypothesized that it might offer a way to measure surprise during sentence processing that can complement other paradigms (e.g., prime surprisal).

Methods

Participants

A total of ninety college-aged adults were tested (N=22-23 in each of four lists). Some participants did not produce usable sound recordings, had no utterances eligible for analyses and/or did not have usable pupillometry data. The number of participants in each analysis is indicated in Results below.

Stimuli

Participants were exposed to a new language in which eight verbs (*fenk, flern, gofe, gund, parn, pelk, semz, stoom*) appeared in verb-agent-patient (VAP) or verb-patient-agent (VPA) order. At the beginning of the experiment, participants were told that they would watch videos of puppets performing different actions and hear sentences in a new language. The new language used English nouns for the animals but novel verbs for the actions. There were two blocks, with training followed by test in each block.

In the first training block (hereafter, Block1), participants saw a video (e.g., a puppet tiger hugging a puppet zebra) and then heard a sentence (e.g., VAP: *stoom tiger zebra* or VPA: *stoom zebra tiger ka* (Note: “ka” is marker that indicates the more unusual order. This kind of marking is common in natural languages and has also been used in multiple previous studies containing artificial languages. Thothathiri & Rattinger, 2016; Wonnacott et al., 2008). Half of the verbs were only heard in VAP structures (VAP bias) and half only in VPA structures (VPA bias). After training, participants underwent the first production test (hereafter, Test1) where they saw new videos and were asked to describe the videos using sentences from the new language. The test videos contained different animals from the training, ensuring that participants had to generate new sentences themselves and could not simply repeat previously heard utterances.

After Test1 was completed, participants were exposed to a second training block (hereafter, Block2). The procedure was the same as in Block1 except that the verb biases were reversed. If a verb was heard only in VAP during Block 1, it was now heard only in VPA and vice versa. Following Block2, participants underwent Test2 (similar to Test1), where they saw new videos and described them.

In Block 1, four verbs appeared with higher frequency (18 times) and four with lower frequency (6 times) for a total of 96 sentences. These were split evenly across VAP and VPA sentence structures. Six different animals (*donkey, monkey, lion, tiger, zebra, giraffe*) appeared an equal number of times as agent or patient with each verb. Order was pseudorandomized such that verbs did not repeat in consecutive trials and the sentence structure could repeat a maximum of 3 consecutive times. No video repeated within or across blocks.

In Block 2, all verbs appeared 6 times for a total of 48 sentences. As in Block 1, order was pseudorandomized. Each animal appeared 7-9 times as agent or patient with each verb. The 48 sentences were split into 6 mini blocks of 8 sentences each. Each verb appeared once in each mini block. Each animal appeared 1-2 times within each mini block. This mini structure within the bigger block structure allowed us to analyze different phases of Block 2 separately (see Results) while maintaining the counterbalancing.

Test1 and Test2 contained 32 trials each. The action corresponding to each verb was shown 4 times. Six animals that were different from training (*bear, pig, frog, cow, cat, dog*) appeared 5-6 times as agent or patient and never more than once in a given thematic role with any verb. Order was counterbalanced as with the training blocks. No video repeated within or across the test blocks.

The assignment of different verbs to different bias and frequency conditions was counterbalanced across four lists such that each verb was in the 4 possible combinations of high/low frequency and VAP/VPA bias once. Participants were randomly assigned to a list.

Procedure

The experiment began with a 9-point eyetracker calibration followed by a thirty-second white screen and a thirty-second black screen with a fixation cross in the middle. This was used to estimate each participant’s pupil dilatability. The participants were then familiarized with the verbs. Each verb was presented once, both orthographically and auditorily. The onset of the auditory stimulus began 500 ms after the onset of the orthographic stimulus. The orthographic presentation continued after the sound ended, for a total duration of 4 seconds per verb. Participants were then familiarized with the pictures of the possible event participants, with the corresponding English nouns superimposed over them orthographically. Each picture was presented for 2 seconds. This was followed by the first training block. On each trial, participants saw a video of one puppet animal performing the action designated by the verb on another. Once the video finished playing, a grey screen with a central fixation cross appeared (which participants had been instructed to look at), and eye recording began. After 500ms, the sentence describing the event played. Five seconds later, the experiment transitioned to the next trial. Test trials worked similarly, except that no sound played after the video. Instead, the verb appeared in the center of the screen and participants were asked to produce a sentence describing the event by speaking into the microphone. Participants were given 6 seconds to respond.

Behavioral Analyses

For the behavioral analyses, we coded the structure of participants' utterances in Test1 and Test2. We only analyzed trials containing the correct verb, correct nouns (or close synonyms e.g., *glove* for *mitten*), and a complete sentence that began with the verb and was followed by two nouns (i.e., VAP or VPA). We accepted sentences with or without the "ka" marker because we were primarily interested in word order. All other trials containing incorrect lexical items or only a subset of the sentence elements or other word orders were discarded. Mixed effects models, as implemented in the lme4 package (version 1.1-35.5) in R (Bates et al., 2015) with *p* values computed using the lmerTest package (version 3.1-3; Kuznetsova et al., 2017) were used to examine if the use of VAP or VPA order varied as a function of verb bias, frequency, and test number. An interaction between verb bias and test number would indicate that participants successfully adapted to the reversal of the verb biases from Block1 to Block2. Categorical variables were sum coded throughout.

To evaluate individual differences in learning, we examined the correlation between the bias effects in Test1 and Test2. The ranef function in lme4 was used to extract by-participant slopes for Verb Bias as a measure of how well each participant learned the verb-structure co-occurrences.

Pupillometry Analyses

The left pupil size of each participant was recorded during the two training blocks (Block1 and Block2) using the Eyelink 1000 Eyetracker with a chin and forehead rest. For each trial, the baseline pupil size was computed as the average during the 300ms following the onset of the training sentence (i.e., the onset of the verb). During this period, we do not expect any effects related to the structure of the sentence (VAP vs VPA). A baseline is necessary because pupil sizes can change over the course of an experiment due to factors of non-interest (e.g., fatigue). For each trial, we were interested in the change in pupil size as the sentence unfolded, relative to the baseline period. Time from the onset of the sentence to three seconds later was divided into 30 bins of 100ms each. The dependent measure for the analysis was the average left pupil size in each bin minus the baseline pupil size for the trial. Time points containing blinks, extended by 50ms on both sides, were excluded from these calculations due to the steep drop off or rise in pupil size from the eye closing or opening respectively. Trials that contained blinks during the entire 300ms after the onset of the sentence were excluded because we could not compute a baseline pupil size for these trials. Altogether, 18.6% of trials in Block1 and 25.5% of trials in Block2 were eliminated after preprocessing.

Mixed effects models were used to examine if pupil sizes were larger when participants heard sentences in Block2 that reversed the verb biases experienced during Block1. Such an increase would be consistent with participants experiencing surprise – the language processing system expects the structure that was associated with the verb during previous exposure, but in Block2, the order of the arguments heard is different from what is expected from prior exposure. This effect might be expected to be especially strong when an a priori unexpected order (PA) is experienced.

In addition, we expected to find an effect of the frequency with which a verb-structure pair appeared in Block 1 on surprise in Block 2. Since each verb always co-occurred with a particular structure, the learners should have predicted the structure more confidently based on a frequent verb, which occurred 18 times in Block 1, than based on a rare verb, which occurred only 6 times. Therefore, participants should have had more prediction error when the structure was switched in Block 2 for a frequent verb compared to a rare one, resulting in larger pupil dilations in response to frequent verbs.

Correlations between Learning and Surprise

To examine whether the extent of surprise was related to the speed of reversal learning, we extracted by-participant random slopes from the pupillometry and behavioral analyses and tested their relationship in a linear regression model, using the lm function in R. To account for any physiological differences between participants in pupil dilatability, we computed the difference between the maximal pupil size observed during the final 25% of a blank black screen (when pupils are expected to expand) and the minimal pupil size observed during the final 25% of a blank white screen (when pupils are expected to constrict) and entered it as a covariate.

Results

Behavioral Results

Data from 77 participants was eligible for the mixed model analysis. The model contained the fixed effects of verb bias (VAP/VPA), frequency (High/Low), and test number (1/2) and their interactions, as well as random intercepts and slopes by participant, and random intercept by verb. The results are shown in Table 1 and Figure 1. There was a significant interaction between verb bias and test number ($b(se)$ = –0.99(.14), p<.001), consistent with participants switching their word order preferences according to the verb bias experienced during Block1 versus Block2 training, with Block2 reversing the biases from Block1. However, there was no significant effect of frequency, suggesting that frequent and infrequent verb biases were learned equally well in both blocks. Given this result, the frequency manipulation may not have been strong enough to affect learning, or participants may have assumed that all verbs have categorical structural biases because this was true of every verb.

Data from 65 participants was eligible for the correlational analysis (some participants produced usable utterances in one test but not the other). There was a significant positive correlation between the bias effect in Test1 and that in Test2 (r=.28, t(63)=2.36, p<.05. See Figure 2). That is, participants who showed better accuracy on the verb biases after Block1 training also showed better accuracy on the opposite verb biases after Block2 training, despite the fact that the biases were reversed between blocks.

Table 1: Behavioral (production) results

Fixed Effect	Estimate (SE)	Z	p
Intercept	**-1.9(.4)**	**-5.32**	**<.001**
Bias	**.6(.2)**	**3.57**	**<.001**
Frequency	.1(.1)	1.32	.19
TestNumber	.2(.2)	1.04	.30
Bias*Freq	-.1(.1)	-1.48	.14
Bias*Test	**-1(.1)**	**-7.02**	**<.001**
Freq*Test	-.1(.1)	-1.47	.14
Bias*Freq*Test	0(.1)	-.12	.90

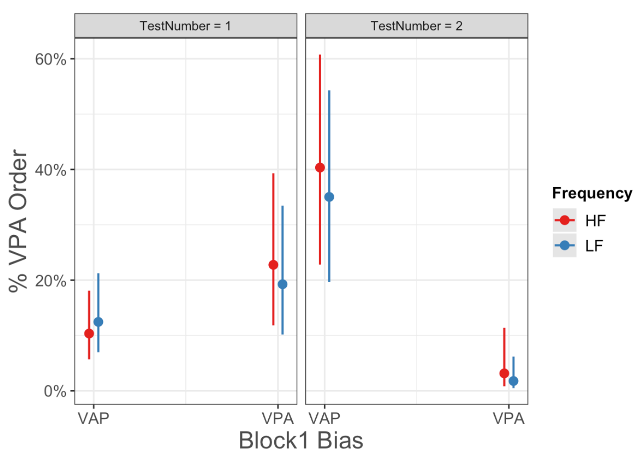

Figure 1: Word order produced by participants. In Test1, participants produced more VPA order with verbs that had a VPA bias during Block1 training. Block2 training switched verb biases (VAP verbs were now VPA-biased and vice versa). Participants successfully flipped their utterances in Test2, producing more VPA order with verbs that were VAP-biased in Block1 and VPA-biased in Block2.

Pupillometry Results

Data from 77 participants was eligible for the pupillometry analysis. The model contained the fixed effects of block (1/2), bias (VAP/VPA) and frequency (High/Low) and their interactions, as well as random intercepts and slopes by participant, and random intercept by verb. The results are shown in Table 2 and Figure 3. There was a significant effect of Block (b(se)=23.6(10.3), t=2.3, p<.05). Pupil sizes were larger at the beginning of Block2 (mini-blocks 1 and 2. See Methods) relative to Block1. This is consistent with an effect of surprise on pupil dilation because verb biases were reversed during Block2 from what they were during Block1. As in the behavioral data, there was no effect of verb frequency and no interactions between frequency and the other predictors.

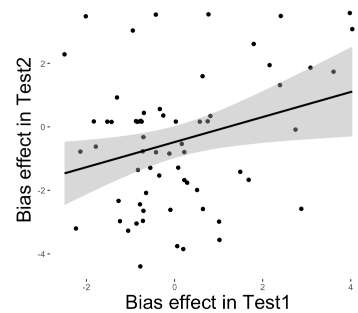

Figure 2: Correlation between the bias learning effects in the two tests. Each data point is one participant.

Table 2: Pupillometry results: Block 1 vs. start of Block 2

Fixed Effect	b(se)	df	t	p
Intercept	**113.5(16.6)**	**187.9**	**6.83**	**<.05**
Block	**23.6(10.3)**	**128.4**	**2.30**	**<.05**
Bias	1.1(6.2)	165.9	.18	.85
Frequency	4.6 (6.6)	175.8	.69	.49
Block*Bias	-3.2(5.0)	88.9	-.65	.52
Block*Freq	4.8(5.4)	98.8	.90	.37
Bias*Freq	-.8(3.9)	108.3	-.21	.84
Block*Bias*Freq	-3.0(3.3)	64.3	-.91	.36

Figure 3: Pupil sizes were larger at the beginning of training Block2 relative to Block1, consistent with surprise. HF=High Frequency verbs, LF=Low Frequency verbs

Participants heard training sentences during Block1, produced their own utterances during Test1, and then heard training sentences again during Block2. We considered the possibility that the pupil size effect in Figure 3 resulted from the change in task that occurred at the beginning of Block2 rather than surprise at the reversal of the verb biases.

Therefore, we conducted a secondary analysis comparing Block1 to the end of Block2 (mini-blocks 5 and 6. See Methods), which does not involve a change in task. The model contained fixed effects of Block (1/2) and Bias (VAP/VPA). Because there were no effects of frequency in any previous analyses, we did not include frequency as a predictor. The results are shown in Table 3 and Figure 4. There was a significant interaction between Block and Bias. For verbs that switched from VAP in Block1 to VPA in Block2, the pupil sizes were significantly larger at the end of Block2 ($b(se)$=20.8(5.6), t(68.8)=3.71, p<.001) than during Block1. For verbs that switched from VPA in Block1 to VAP in Block2, there was no effect of Block ($b(se)$=2.1(8.0), t(70)=.26, p=.80).

To summarize, we found pupil effects consistent with surprise at the beginning of Block2, when the verb biases were reversed. We also found pupil effects consistent with surprise at the end of Block2, where there was no confound of a change in task.

Table 3: Pupillometry results: Block 1 vs. end of Block 2

Fixed Effect	$b(se)$	df	t	p
Intercept	**100.5(13.0)**	**147.2**	**7.71**	**<.05**
Block	10.9(6.8)	118.0	1.61	.11
Bias	-6.8(4.7)	95.6	-1.43	.16
Block*Bias	**-9.9(4.3)**	**66.6**	**-2.33**	**<.05**

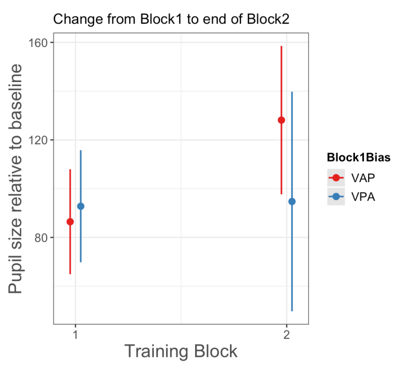

Figure 4: Pupil size was larger at the end of training Block2 for verbs that switched from VAP to VPA (red) but not for verbs that switched from VPA to VAP (blue).

Error-Based Learning Results

Sixty-two participants had analyzable results from both behavioral and pupillometric analyses for computing correlations. Random slopes for the interaction between verb bias and test number in the behavioral analyses (corresponding to Table 1/Figure 1) were used to index individual differences in reversal learning. From the pupillometry analysis, random slopes for the effect of training block at the beginning of Block2 (corresponding to Table 2/Figure 3) were used to index individual differences in surprise. The regression model contained the behavioral effect as the dependent variable, the measure of surprise as the independent variable of interest, and a measure of each participant's pupil dilatability as a control variable (see Methods). The association between the behavioral and pupillometric (surprise) effects was not significant ($b(se)$=.00004(.003), t=.01, p=.99). Similarly, the association between the behavioral effect and the surprise effect for VAP verbs at the end of Block 2 (corresponding to the VAP portion of Figure 4) was not significant ($b(se)$=–.003(.004), t =–0.79, p=.44). In contrast, the two pupillometry effects were significantly correlated even after accounting for individual differences in dilatability ($b(se)$=.33(.10), t=3.19, p<.005), suggesting that there was sufficient variation in the measures to be able to detect a correlation, if one existed. The results are summarized in Figure 5. Participants with larger pupil size effects at the beginning of Block2 also showed larger effects at the end of that block. However, neither of these effects was correlated with the behavioral effect.

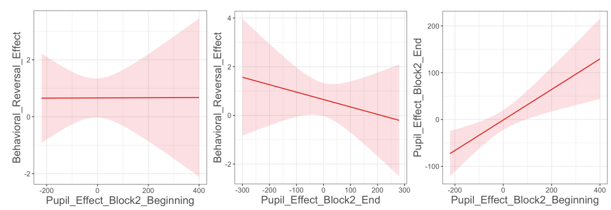

Figure 5: Correlations between individual differences in surprise and reversal learning. Left and Middle: No significant correlations between pupil effects and behavioral effect. Right: Significant correlation between pupil effects.

Discussion

This study investigated whether verb-syntax links are learned and updated by an error-based learning mechanism. Participants heard sentences in an artificial language containing novel verbs and two novel verb-initial word orders. The bias of individual verbs towards one of the two structures (VAP vs VPA) was reversed between training blocks. Production data showed that participants successfully learned this reversal.

Pupillometry showed evidence of surprise when participants encountered an unexpected structure. Pupil sizes increased in the second training block relative to the first one. Further, this increase persisted for verbs that switched to the less expected order of arguments in English (patient before agent; PA), as would be expected based on the subjects' cumulative experience in and out of the lab. Since the AP order is expected in English, the native language of the participants, it may not cause surprise even if the preceding verb had previously appeared in the PA order. In contrast, encountering the PA order after a verb that had previously appeared in AP seems to have led to sustained surprise that persisted across Block2.

Despite finding positive evidence for both surprise and verb bias learning, we did not find any support for an association between the two. There was no suggestion that participants who showed more surprise when verb biases were reversed were the ones who showed better reversal learning. One possibility is that the null regression finding is a false negative, arising from a lack of statistical power.

However, we were able to detect other correlations—between the behavioral effects in Test1 and Test2, and between the pupillometry effects at the beginning and end of Block2. Furthermore, BIC comparisons between regression models with and without surprise yielded $p(H0) = 89\%$ (following the procedure in Wagenmakers, 2007), indicating substantial evidence for a lack of correlation. At the same time, a reliable but weak correlation can never be conclusively ruled out.

These results align with those of Thothathiri and Levshina (2023), who examined learning of new verb biases for English verbs, and tested error-based learning by examining whether there was better learning when an experimental verb bias deviated more from the verb's prior bias (Thothathiri & Levshina, 2023). Bayesian analysis provided evidence for the learning of experimentally manipulated biases but no support for error-based learning. However, one reason for the null result in that study could be the distinctive contexts in which the verb biases were initially learned vs. reversed (real-world experience vs. lab). If listeners learn and maintain *context-specific* statistical information, they might not engage in predictively processing sentences in the current context based on input experienced in a very different context. This explanation does not apply to the present study because the switch in bias occurred in the same context in which the original biases were learned. Thus, the present null result for error-based learning requires a different explanation.

One possibility is that while learning clearly occurred within this paradigm, it was not driven by prediction error. For example, it is possible that the relevant verb-order associations were acquired by Hebbian learning, which posits that learning rate is independent of prediction error (e.g., Hadley & Hayward, 1997; McClelland, 2006). It is also possible that the learning proceeds by instance-based storage of the encountered sentences as episodes that encompass the verb context and the order of the arguments (Johns et al., 2020). From this perspective, there may be no need to learn an *association* between a verb and an argument order. Instead, the verb and the order would be stored as a single configural vector, and production at test would proceed by averaging over the vectors containing the presented test verb.

Other evidence in the literature supports the perspective that, while prediction is a key component of language processing, it may not be used by everyone under every circumstance. For example, older adults are less likely to show prediction signatures in electrophysiological studies (Wlotko, Federmeier & Kutas, 2012). Less literate younger adults also show less evidence of prediction (Huettig & Pickering, 2019) Prediction may also not occur under high cognitive load or time pressure. In general, for error-based learning to occur, the learner must have enough time to predict the outcome before the perceptual evidence for its presence or absence becomes available (Kapatsinski, 2018). It may be that learners in the present experiment did not consistently use the sentence-initial verbs to predict the noun they would hear next, before hearing it. If so, introducing a delay between the verb and the nouns, reducing the number of different nouns, or using the same nouns in test as training might make prediction and error-based learning more likely.

It is also possible that the magnitude of pupil dilation in an individual may not closely track the magnitude of surprise the individual experienced. For example, Becker et al. (2024) found that only very large differences in surprise affected the magnitude of pupil dilation in a task where people used visual stimuli (faces) to predict spoken vowels. Similarly, in the present study, we did not find any effect of verb frequency on pupil dilation, even though participants may be expected to be more surprised when encountering a reversal of an association based on 18 trials vs. 6. However, a limitation of the present design is that all verbs categorically preferred one structure or another, and all switched in unison. Participants therefore could determine the bias of a verb from a single trial, potentially eliminating frequency effects. Future studies could use stronger within-subject manipulations (e.g., manipulations of some switching verbs having much higher pre-switch probability of being followed by AP) to examine whether the presence of surprise is associated with learning.

Before closing, the correlation between performance in Test 1 and Test2 warrants some discussion. Learners who did well in Test1 had more to unlearn in Block2, so how were they able to perform better in Test2, if learning was not error-based? One possibility is that some learners were better able to attend to the relevant cues. For example, Kruschke (1996) showed that drawing attention to relevant stimulus dimensions allows learners to rapidly reverse the associations of the cues residing on these dimensions. It may be that the better learners were attending to the dimensions that matter (verbs/actions) while others were attending to less relevant dimensions such as which animals were involved (see also Thothathiri & Levshina, 2023). Attending to the wrong cues would result in poor learning in both blocks, creating a positive correlation between the two. Future studies could examine if verbs whose biases were learned well in the first block were also learned well in the second block.

Analyses of pupil size averaged across time bins revealed similar results as above. However, future studies could employ cross-validation or other approaches for dealing with autocorrelated data (Mathot & Vilotijevic, 2023).

In conclusion, the present study adds to the literature showing impressive statistical learning of syntactic patterns (verb-structure dependencies) with no evidence for the use of an error-based learning mechanism. The accumulating evidence includes the mixed findings for prime-surprisal in structural priming studies and the lack of an interaction between prior and current verb biases in experimental studies that manipulate verb bias. We have suggested here that even when there is detectable surprise at an unexpected outcome (as indexed by pupil dilation), the rate of learning may not be related to the degree of surprise. We believe that this research area would benefit from explicating the contexts in which prediction-based learning is expected and not expected, the variety of learning mechanisms that can enable language users to flexibly adapt language based on ongoing input, and the precise relationship between surprise and pupil dilation.

References

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Becker, J., Viertler, M., Korn, C. W., & Blank, H. (2024). The pupil dilation response as an indicator of visual cue uncertainty and auditory outcome surprise. *European Journal of Neuroscience*, *59*(10), 2686-2701.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*, 355–387.

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, *29*, 217–229.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164-194.

Ellis, N. C., & Sagarra, N. (2010). The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition*, *32*(4), 553-580.

Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science*, *7*(11), 180877.

Fazekas, J., Sala, G., & Pine, J. (2024a). Prime surprisal as a tool for assessing error-based learning theories: A systematic review. *Languages*, *9*(4), 147.

Fazekas, J., Blything, L., & Ambridge, B. (2024b). No surprises: Strong evidence against a prime surprisal effect for the dative construction, with a large sample and an LLM-based surprisal predictor.

Hadley, R. F., & Hayward, M. B. (1997). Strong semantic systematicity from Hebbian connectionist learning. *Minds and Machines*, *7*, 1-37.

Huettig, F., & Pickering, M. J. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences*, *23*(6), 464-475.

Jaeger, T. F., & Snider, N. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 827-812. Austin, TX: Cognitive Science Society.

Johns, B. T., Jamieson, R. K., Crump, M. J., Jones, M. N., & Mewhort, D. J. K. (2020). Production without rules: Using an instance memory model to exploit structure in natural language. *Journal of Memory and Language*, *115*, 104165.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Cambell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279 –296) New York, NY: Appleton-Century-Crofts.

Kapatsinski, V. (2018). *Changing minds changing tools: From learning theory to language acquisition to language change*. Cambridge, MA: MIT Press.

Kuznetsova A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, *8*(2), 225-248.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 830-845.

Lin, Y., & Fisher, C. (2017). Error-based learning: A mechanism for linking verbs to syntax. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Maes, E., Boddez, Y., Alfei, J. M., Krypotos, A. M., D'Hooge, R., De Houwer, J., & Beckers, T. (2016). The elusive nature of the blocking effect: 15 failures to replicate. *Journal of Experimental Psychology: General*, *145*(9), e49-e71.

Mathot, S., & Vilotijevic, A. (2023). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods, 55*(6), 3055-3077.

McClelland, J. L. (2006). How far can you go with Hebbian learning, and when does it lead you astray. *Processes of change in brain and cognitive development: Attention and performance XXI*, 33-69. San Diego: Academic Press.

Nixon, J. S. (2020). Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition*, *197*, 104081.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1-15.

R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*(6), 1017-1023.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909-957.

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 781.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238-299.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, *114*(3), 759-783.

Thothathiri, M., Evans, D. G., & Poudel, S. (2017). Verb bias and verb-specific competition effects on sentence production. *PloS One*, *12*(7), e0180580.

Thothathiri, M., & Rattinger, M. G. (2016). Acquiring and producing sentences: Whether learners use verb-specific or verb-general information depends on cue validity. *Frontiers in Psychology*, *7*, 404.
Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779-804.
Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975-988.
Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165-209.
Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, *22*, 2331216518777174.