Episodic Memory Cues in Acquisition of Novel Visual-Phonological Associations: a Webcam-Based Eye-Tracking Study

Simone L. Calabrich (sml19pvv@bangor.ac.uk)
School of Psychology, Bangor University
Bangor, Gwynedd LL57 2AS, UK

Gary M. Oppenheim (g.m.oppenheim@bangor.ac.uk)
School of Psychology, Bangor University
Bangor, Gwynedd LL57 2AS, UK

Manon W. Jones (manon.jones@bangor.ac.uk)
School of Psychology, Bangor University
Bangor, Gwynedd LL57 2AS, UK

Abstract

When learning to bind visual symbols to sounds, to what extent do beginning readers track seemingly irrelevant information such as a symbol's position within a visual display? In this study, we used adult typical readers' own webcams to track their eye movements during a paired associate learning task that arbitrarily bound unfamiliar characters with monosyllabic pseudowords. Overall, participants' error rate in recognition (Phase 1) decreased as a function of exposure, but was not modulated by the episodic memory-based effect of 'looking-at-nothing'. Moreover, participants' lowest error rate in both recognition and recall (Phases 1 and 2) was associated with item consistency across multiple exposures, in terms of spatial and contextual properties (i.e., stimulus' screen location and co-occurrences with specific distractor items during encoding). Taken together, our findings suggest that normally developing readers extract statistical regularities in the input during visual-phonological associative learning, leading to rapid acquisition of these pre-orthographic representations.

Keywords: Episodic memory; looking-at-nothing; paired associate learning; cross-modal binding; reading.

Introduction

Associative learning is a key skill underlying reading development. In initial stages of literacy acquisition, written or printed symbols (i.e., graphemes), which at first appear meaningless, gradually begin to evoke specific linguistic representations (i.e., phonemes). Repeatedly accessing such phonological associations in response to visual stimuli (i.e., letters) progressively automatizes that process (Ehri, 2005; Ehri & Saltmarsh, 1995; Jones et al., 2018) resulting in the apparent effortlessness of skilled reading. Performance in visual-verbal versions of the paired associate learning task – an episodic memory paradigm which assesses the ability to accurately bind two distinct items together in memory (Scorpio et al., 2018) and retrieve them later as a single entity (Brockmole & Franconeri, 2009) - has been shown to discriminate typical readers from those with dyslexia (e.g., Jones et al., 2018; Toffalini et al., 2018; Wang, Wass, & Castles, 2017). Paired associate learning performance accounts for unique variance in reading ability, and impairments to the underlying skills appear to result in clinically significant reading difficulties (Litt & Nation, 2014; Wang et al. 2017), supporting the assumption that the task taps abilities that are crucial for skilled reading acquisition.

Reading acquisition thus appears to build on episodic memory. In episodic memory, contextual properties, such as temporal and spatial information, are encoded alongside salient task features (Tulving, 1972). These properties, which share patterns of neural activity, can be used as cues to aid memory retrieval (El-Kalliny et al., 2019). To illustrate, if Event A is encoded in temporal proximity to Event B, exploiting the temporal relationship between the two events may facilitate their subsequent retrieval from the episodic memory system when needed (Tulving, 1972; El-Kalliny et al., 2019). Episodic memory-based investigations focusing on learning of arbitrary visual-phonological associations demonstrated that typical readers, but not individuals with dyslexia, are sensitive to consistent spatial cues presented *across multiple trials* (Albano, Garcia, & Cornoldi, 2016; Jones et al., 2018; Toffalini et al., 2018). Typical readers' sensitivity to spatial cues extends to their oculomotor behavior: when given a visual cue, they fixate blank screen locations previously occupied by a target item, resulting in greater probability of accurate phonological recall (Jones et al., 2018).

Returning to a spatial location in which salient information was originally presented is an unconscious oculomotor behavior that is triggered by the reactivation of internal memory representations (Ferreira et al., 2008; Richardson & Spivey, 2000). This behavior is believed to play a functional role in memory retrieval (Richardson & Spivey, 2000; Scholz, Klichowicz, & Krems, 2018), modulating retrieval of both visual and auditory information (Scholz, Mehlhorn, & Krems, 2016). The phenomenon seems to occur even when encoding of spatial information is task-irrelevant (Richardson & Spivey, 2000) and thus encoded incidentally. The episodic memory-based effect of 'looking at nothing' when trying to

remember something gradually diminishes as learning unfolds and representations strengthen over time (Scholz et al., 2016; Wantz et al., 2016).

To date, however, the effect of presentation consistency in the episodic trace on visual-phonological binding accuracy in typical readers is relatively underexplored. Here, we begin to elucidate the cognitive underpinnings of efficient orthographic-phonological representations in typical readers.

The Current Study

We examine whether typical readers efficiently use a combination of spatial and contextual cues to aid learning of novel cross-modal bindings, taking a full and accurate snapshot of the episodes to facilitate the visual-phonological binding. To test this, we designed a paired associate learning task in which we manipulated consistency of stimuli's spatial locations and their co-occurrences across multiple exposures. Our goal is to probe whether these episodic cues, when combined, modulate recognition of novel visual-phonological associations in typical readers. We also examined whether 'looking-at-nothing' behavior would emerge in the current study at the trial level, and if so, whether directing one's gaze towards relevant empty screen locations would aid recognition of the novel associations.

We tracked participants' eye movements remotely with their webcams during a paired associate learning task in which Kanji characters – which were unfamiliar to these native British English speakers – were arbitrarily but consistently bound to monosyllabic pseudowords adhering to phonotactic constraints in English. On each trial, as in Jones et al. (2018), participants were prompted to encode three characters, one at a time, along with their corresponding pseudowords. An auditory cue with the target pseudoword followed the encoding phase. After a blank screen, during which we tracked participants' eye movements, participants were then tested on their ability to recognize the corresponding character associated with the auditory cue. Our manipulation of consistency of stimuli's locations and intra-trial co-occurrences ('context', henceforth) resulted in four different trial types. Consistent location involved Kanji characters appearing in the same screen location across trials, whereas consistent context involved characters appearing with the same distractor items across trials. A separate cued-recall task was administered to assess lasting retention of the visual-phonological associations.

Based on previous empirical findings that typical readers gradually automatize retrieval of visual-phonological associations over time (Jones et al., 2018), performance in later blocks should be superior as a function of repetition, which, in turn, will be an indication of incremental learning.

If typical readers are able to efficiently use *multiple* episodic cues present during encoding in order to aid recognition of the novel visual-phonological associations, then they should err less when both location and context are kept consistent across trials, as compared to when they are not. Furthermore, if encoding under the consistent location/ consistent context condition is indeed more robust than in the other conditions as a consequence of the regular episodic cues, then we will also observe longer-lasting retention of the bindings encoded under this condition (as assessed by a separate cued-recall task following the main recognition task) showing that typical readers not only efficiently detect regularities in the stimuli but also use them to their advantage.

Considering that visually revisiting empty screen locations previously occupied by targets has been shown to aid memory retrieval, we expected looking-at-nothing behavior to also emerge in our study.

Finally, one unique methodological aspect of this study is its use of a webcam-based method for remote eye-tracking. Previous research on the role of looking-at-nothing behavior in paired associate learning has been conducted in-lab with specialized hardware. Here, we set out to investigate the phenomenon remotely using WebGazer.js, an open-source webcam-based eye-tracking JavaScript library (Papoutsaki et al., 2016) which has been shown to reliably detect fixations and replicate findings of in-lab cognitive science studies with reasonably comparable accuracy (Semmelmann & Weigelt, 2018). Without transmitting videos or pictures, WebGazer.js uses participants' webcams to infer on-screen gaze locations with an average error of approximately 100 pixels. Thus, this study provides a test of the method's suitability as a flexible, low-cost alternative for 'looking-at-nothing' research.

Method

Participants

Fourteen university students (age: M = 22.6, SD = 4.21, 13 females) participated remotely in this experiment. One additional participant was excluded due to an error rate more than three standard deviations above the group mean. All were native speakers of British English, recruited through Bangor University, and none reported any history of psychiatric and/or neurological diseases, visual acuity, hearing, or any other risk factors. Crucially, all participants self-reported normal or skilled reading ability in the Adult Reading Questionnaire (Snowling, Dawes, Nash, & Hulme, 2012). All participants were naïve to the purpose of the experiment, and had never seen nor heard any of the stimuli before. The experiment was approved by the Bangor University Ethics Committee, and participants provided informed consent and received payment for participation.

Stimuli, Design and Procedure

Phase 1: Recognition Task The task was programmed and hosted on Gorilla Experiment Builder (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2018). Participants were not allowed to do the task on mobile phones or tablets. Participants' physical distance from the screen was calculated with the Virtual Chinrest task (Li et al., 2020), which indicated an average sitting distance of 50.88 cm from their monitors (SD = 8.59). Participants were instructed to sit still, and to avoid head movements and/or to look away from the screen during the task. Pictorial instructions were included in an attempt to collect higher data quality. A 5-

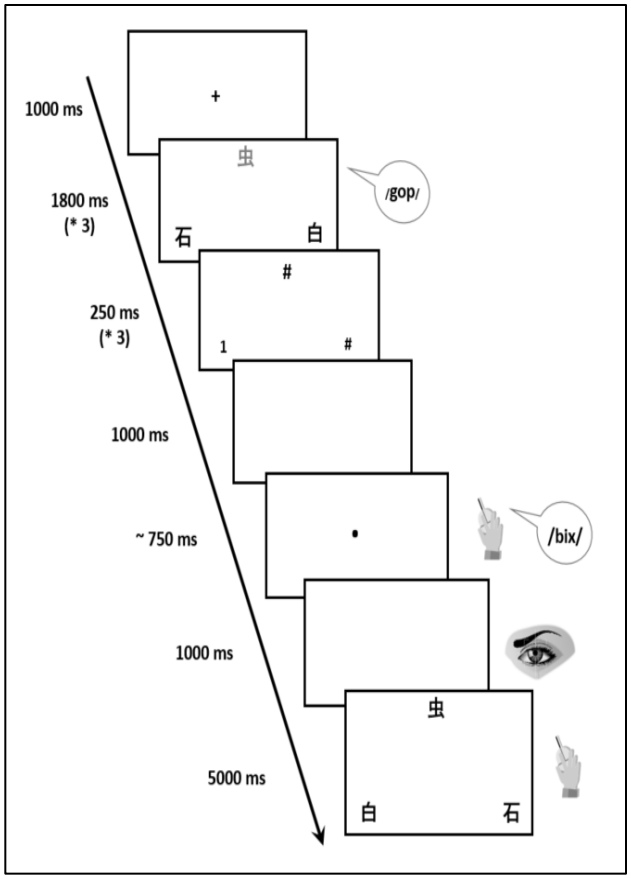

Figure 1: Timeline of a single trial in the recognition task. The eye denotes the screen in which we expected to detect 'looks-at-nothing' behavior.

point calibration was performed at the beginning of the main task and every 18 trials (i.e., mid-block and at the onset of a new block). Calibration was re-attempted whenever the calibration prediction for at least one of the five calibration points approximated an incorrect one.

Thirty-six Kanji characters were arbitrarily matched to 36 monosyllabic pseudowords (e.g., 'kig'), generated with Wuggy (Keuleers & Brysbaert, 2010) according to English-like phonotactic constraints. The auditory stimuli were recorded by a female native speaker of British English. Character-sound pairings were kept constant across the experiment such that each character was always bound to the same pseudoword.

Each trial began with a 1000-ms fixation cross. Then, three Kanji characters appeared in black on white background in triangle formation (See Fig. 1). Each character occupied 20x20 units within a 4:3 window in Gorilla Experiment Builder's screen space. One at a time, each character would pseudo-randomly highlight in red while its corresponding pseudoword played in the background (participants were encouraged to use earphones or headphones to listen to the stimuli). A 1000-ms blank screen followed, and then a visual backward masking phase, during which hash symbols and numbers replaced the target stimuli on the screen to minimize iconic memory. Then the 'testing phase' began. A black dot appeared in the centre of the screen; participants were instructed to click on it to hear one of the three pseudowords: the 'target' for the testing phase. This clicking instruction also provided a crucial attention check: participants were automatically excluded from the experiment if they failed to click on the dot within 10 seconds in three consecutive trials. A 1000-ms blank screen followed, during which participants' eye movements were recorded via WebGazer (Papoutsaki et al., 2016) with a sampling rate of 60 Hz. The three Kanji characters re-appeared; to encourage participants to encode character-sound associations, characters' spatial positions changed between the encoding and testing phases in two thirds of the trials[1]. Participants were prompted to click on the character that corresponded to the target audio. The characters remained on the screen for up to 5000 ms or until a mouse-click was detected, after which a 250-ms blank screen terminated the trial.

We orthogonally manipulated two aspects of the encoding phase: 1. *Location consistency*: whether a target character consistently appeared in the same spatial location throughout the experiment, and 2. *Context consistency*: whether a target character consistently appeared with the same two other characters throughout the experiment. Thus, of the 36 Kanji characters, 18 always appeared in the same screen position, whereas 18 characters varied in position. Similarly, half of the stimuli consistently co-occurred with the same two other characters, whilst the remaining 18 did not have any fixed co-occurrences.

To ensure attention to the phonological component of the bindings, we interspersed cued-recall trials within each block at regular intervals (i.e., every six trials). In each trial, a Kanji character was shown in the middle of the screen (see Fig. 2), after which participants were prompted to articulate the corresponding pseudoword. The target for each interspersed recall trial (N= 36) was a character randomly selected from one of the six preceding recognition trials.

The 252 trials (216 recognition trials plus 36 interspersed cued-recall trials) were presented over 6 blocks, between which participants were encouraged to take short breaks. Trials' assignment to blocks was pseudo-randomized to ensure that all conditions were equally frequent within a block. Presentation of blocks and of trials within each block was randomized across participants to avoid order effects.

Five practice trials (i.e., four recognition trials and one recall trial) representative of those used in the actual experiment were presented in order to familiarize the participants with the procedure. None of the practice items were used during the experiment. Feedback was provided to participants during the practice block, but not in the experimental trials.

[1] Due to the automatic and unconscious nature of the 'looking-at-nothing' behavior (Ferreira et al., 2008); Richardson & Spivey, 2000), we did not expect this manipulation to prevent participants from re-fixating relevant screen locations.

Phase 2: Cued Recall A separate cued-recall task comprising the same visual-auditory stimuli from the previous task was administered on Gorilla Experiment Builder (Anwyl-Irvine et al., 2018) immediately after Phase 1. The task consisted of a single block with 36 trials. Each trial, methodologically identical to the above mentioned interspersed cued-recall trials, started with a 1000-ms fixation cross, followed by a Kanji character presented in black on a white background (See Fig. 2). The character was presented in the center of the screen for 1000-ms, and occupied 20x20 units of screen space within a 4:3 window. Three black dots, presented in the same triangle formation as Phase 1, indicated that a voice response was required. Participants were allowed 4 seconds to provide a verbal response. A 250-ms blank screen terminated the trial. Trial presentation was randomized across participants to avoid order effects. Eye-tracking metrics were not recorded in this task.

Total experiment duration averaged 105 minutes. An automatic time limit of 150 minutes ensured that participants would complete the experiment in one sitting.

Data Analysis

Eye tracking. Eye-tracking metrics recorded by Gorilla Experiment Builder (Anwyl-Irvine et al., 2018) include a face convergence value column, which comprises a score ranging from 0 to 1 for the face model fit. The face convergence value indicates how strongly the image detected resembles a face: 0 means no fit and 1 means perfect fit. Gorilla's recommendation is to trust face convergence values over 0.5. We excluded eyetracking estimates below that threshold in our analyses.

Under ideal conditions, WebGazer.js (Papoutsaki et al., 2016) is able to generate up to 60 eyetracking estimates (i.e. predictions) per second with x and y coordinates of where on the screen the subject is predicted to be looking. However, the number of predictions largely varies depending on participants' hardware, lighting conditions, among other things. In addition to these predictions, Gorilla Experiment Builder (Anwyl-Irvine et al., 2018) translates the coordinate data into a 'normalized' space, in which -0.5 and 0.5 will always be the center of the screen regardless of its size. This normalization allows eye movements detected across different screen sizes to be compared. We used the normalized coordinates in our analyses.

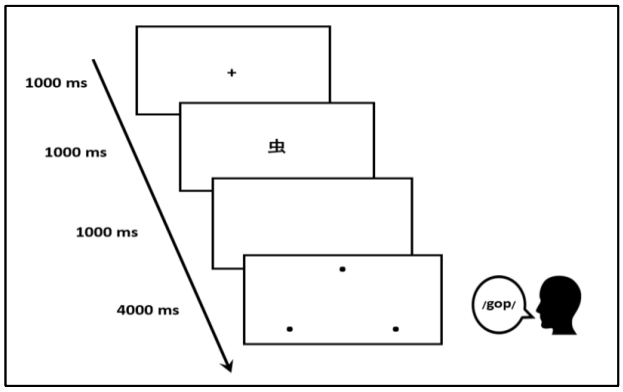

Figure 2: Timeline of a single trial in the cued-recall task.

Regression analyses. Analyses used confirmatory logistic mixed effects regression, via the glmer::binomial function in the lme4 v1.1-23 library (Bates, Mächler, Bolker, & Walker, 2014) in R v4.0.0 (R Core Team, 2020), including maximal random effects structures (Barr, Levy, Scheepers, & Tily, 2013) reverting to a 'parsimonious' approach in the case of convergence errors (Bates et al., 2015). For the recognition task in Phase 1, error rate was modelled as a function of *Location consistency* ("LocationC", i.e., whether a target character consistently appeared in the same spatial location throughout the experiment; consistent = -0.5, inconsistent = 0.5), *Context consistency* ("ContextC", i.e., whether a target character consistently appeared with the same two other characters throughout the experiment; consistent = -0.5, inconsistent = 0.5), and *Block*, a predictor tracking target repetition, log-transformed to account for the fact that repetition effects follow a logarithmic function. Following Jones et al. (2018), to probe whether participants' looks back at blank screen locations previously occupied by targets would facilitate recognition of those items, we also included two eyetracking-related binomial predictors: (1. a binomial predictor indicating whether we identified fixations on any

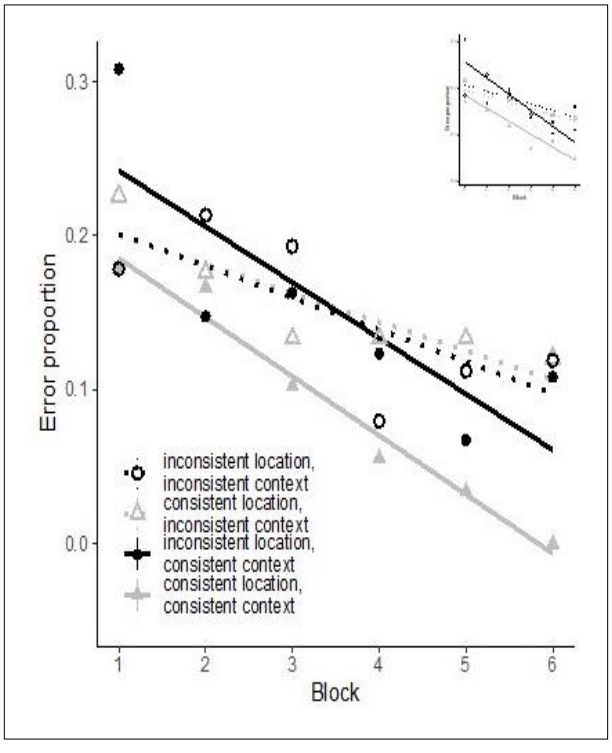

Figure 3: Error rate by condition in the Phase 1 recognition task. The main figure depicts the pattern in the restricted dataset; the inset shows the same pattern when including trials without valid eyetracking data.

region of interest during the blank screen immediately preceding the testing phase ("FixatedAnyROI", no = −0.5, yes = 0.5), and 2. a nested binomial predictor indicating whether the participant fixated the former location of the target more than the former locations of the distractors ("PrimaryFixation", target = -0.5, distractor = 0.5, no fixations = 0.0). All predictors were contrast-coded and centered. For Phase 2's cued-recall task, accuracy was modelled as a function of *Location consistency* and *Context consistency*, as described above.

Results

Phase 1 (Recognition Task)

We excluded 30 (out of 3024) trials without behavioral responses (i.e., mouse clicks), leaving 2994 trials. The eye tracking procedure generated 52,204 fixation estimates across these 2994 behaviorally valid trials. We excluded 1.39% of those estimates (*N*=726), due to face convergence values below 0.5, indicating low-confidence eyetracking estimates. Finally, to address questions about looking-at-nothing behavior, in this paper, we focus our analyses on just the 2093 behaviorally valid trials with at least one valid eyetracking estimate; as illustrated in Figure 3, this restricted dataset is behaviorally very similar to the larger dataset. The mean face convergence value for these remaining trials was 0.77 (*SD* = 0.12), suggesting a sufficient basis for estimating eye movements. Participants primarily fixated the former locations of the target in 17% (*N* =366) of these trials, former locations of distractors in 18% (*N*=386), the center of the screen in 41% (*N*=874), and elsewhere in 22% (*N*=467).

Error rate data are illustrated in Figure 3, and do not suggest floor or ceiling effects in the recognition task. As described in the Method section, we used logistic mixed effects regression to model error rates as a function of location consistency, context consistency, target repetition, and eye fixation patterns (Table 1). Participants benefitted from stimulus repetition, erring less in later blocks (OR: 0.36:1, $\beta_{log(Block)}$ = -1.02, *SE* = .22 $p < .001$), and this benefit was stronger for targets that repeatedly appeared with the same distractors than those appearing with different distractors (OR: 1.93:1, $\beta_{log(Block) \times Context}$ = 0.66, *SE* = .22, $p = .003$). Finally, as illustrated in Figure 3, participants particularly benefitted from the combination of a consistent context with a consistent location (OR: 0.42:1, $\beta_{Location \times Context}$ = -0.87, *SE* = .44, p = .046).[2]

On average, participants correctly articulated 19.3 out of 36 pseudowords in the interspersed cued-recall trials (*SD* = 6.95). Since these trials were only included to ensure participant engagement with the task, they were not further analysed.

Phase 2 (Cued-Recall Task)

Due to slow Internet connections, two participants' audio recordings from the cued-recall task failed to properly upload to Gorilla Experiment Builder's server, leaving a total of 12 participants for these analyses. On average, participants correctly articulated 20 out of 36 pseudowords in the cued-recall task (*SD* = 10.91). Participants' mean error proportions per trial type (i.e., whether location and/or context were consistent) can be found on Table 2.

We used logistic mixed effects regression to model error rates as a function of location consistency and context consistency (Table 3). As in the recognition task, these factors significantly interacted to affect cued recall performance (OR: 0.30:1, $\beta_{Location \times Context}$ = -1.21, *SE* = .61, $p = .049$): as in the Phase 1 recognition task, target location consistency only appeared to affect error rates when the target had been consistently presented with the same pair of distractors.[3]

Discussion

In this study, we examined the conditions under which typical readers optimally learn to associate visual-phonological information, simulating the process of acquiring orthographic-phonological representations. Specifically, we investigated the extent to which ostensibly task-irrelevant episodic details modulate visual-phonological binding performance in typical readers. To this end, we tested whether encoding new visual-phonological associations over multiple exposures was modulated by whether targets consistently appeared in the same screen locations or with the same pair of non-target distractors. To assess whether visual attention, in the form of 'looking-at-nothing' behavior, modulated these episodic effects, we also used participants' webcams to remotely track their eye movements.

Recognition accuracy for novel orthographic-phonological bindings improved with repetition (see Fig. 3), in line with previous evidence in the paired-associate learning literature (e.g., Jones et al., 2018), and suggesting an incremental development of stable visual-phonological associations with repetition.

Recognition, as well as later recall, was also modulated by the consistency of extraneous cues that were present during encoding. Participants more accurately recognized visual symbols from associated nonword cues for targets that were

[2] In a post-hoc analysis, we examined the effect of varying stimuli positions between encoding and testing phases. Although participants erred significantly more when stimuli positions were mismatched across the two phases (OR: 2.23:1, $\beta_{EncodingVersusTestingPositions}$ = -1.02, SE = .24 $p < .001$), the overall pattern of results indicated in the main analysis stayed largely the same ($\beta_{log(Block)}$ = -1.07; $\beta_{Context}$ = 0.58; $\beta_{log(Block) \times Context}$ = 0.73; $\beta_{Location \times Context}$ = -0.90; all *p*s < .05).

[3] Observed power for the significant results: Recognition task: 1-$\beta_{log(Block)}$ = .99; 1-$\beta_{Context}$ = .83; 1-$\beta_{log(Block) \times Context}$ = .84; 1-$\beta_{Location \times Context}$ = .58. Separate recall task: 1-$\beta_{Location \times Context}$ = .62. Due to the noisier nature of webcam-based eyetracking, we did not have a good basis for a pre-hoc power calculation for the current study. We intend to use the current findings to estimate sample and effect sizes that are suitable for the context of webcam-based eyetracking in future paired-associate learning/ looking-at-nothing experiments.

Table 1: Summary of a logistic mixed effects regression analysis of recognition error frequency.

	Coef (ß)	SE (ß)	*p*	*OR (exp(ß))*
(Intercept)	-2.49	0.37	**<.001**	0.08
log(Block)	-1.02	0.22	**<.001**	0.36
LocationC	0.23	0.33	0.489	1.26
ContextC	0.50	0.20	**0.011**	1.65
PrimaryFixation	0.39	0.29	0.175	1.48
FixatedAnyROI	-0.21	0.16	0.195	0.81
Block x LocationC	0.19	0.29	0.509	1.21
Block x ContextC	0.66	0.22	**0.003**	1.93
LocationC x ContextC	-0.87	0.44	**0.046**	0.42
Block x LocationC x ContextC	-0.32	0.45	0.477	0.73

Table 2: Summary of subject-weighted mean error proportions in the Phase 2 cued-recall task.

		Context	
		Consistent	Inconsistent
Location	Consistent	.454 (.274)	.491 (.340)
	Inconsistent	.500 (.320)	.493 (.216)

Table 3: Summary of a logistic mixed effects regression analysis of cued-recall error frequency.

	Coef (ß)	SE (ß)	*p*	*OR (exp(ß))*
(Intercept)	-0.32	0.46	0.481	0.73
LocationC	-0.12	0.31	0.690	0.88
ContextC	-0.17	0.29	0.562	0.84
LocationC x ContextC	-1.21	0.61	**0.049**	0.30

consistently presented in *both* the same screen location and with the same distractor symbol/nonword pairs.This finding suggests that, during the process of building an episodic representation of a novel visual-phonological binding, typical readers not only incorporate all the features available at the time of encoding, a typical occurrence in episodic memory formation (Tulving, 1972), but they also appeared to use the consistent features as an aid to help them retrieve these representations from memory. This pattern also emerged in the subsequent cued-recall task, which demonstrated superior accuracy for the bindings that participants had encoded in the consistent location and consistent context condition, suggesting that multiple co-occurring statistical frequencies in the input enable typical readers to quickly acquire accurate visual-phonological bindings, even after relatively few exposures.

In our experiment, participants were prompted to encode three bindings in each trial. In the consistent context condition, all three bindings repeatedly co-occurred over the course of the experiment. We might speculate that participants encoded all three bindings and stored them together, such that when the locations of these items were *in*consistent across trials, separating one item representation from the others for recall became problematic.

It is worth noting that our superadditive interaction of location consistency and context consistency for novel orthographic/phonological bindings resembles on its surface, at least, a very well-known superadditive effect in which relative location consistency interacts with context consistency to support perception and recall of overlearned orthographic-phonological bindings: 'the word superiority effect' (Baron & Thurston, 1973). This resemblance is intriguing because models of that effect often attribute it to robust connections between well-established representations (e.g. Rumelhart & McClelland, 1982). If a shared mechanism underpins both effects, our results would further demonstrate continuity between the earliest stages of binding acquisition and the distant goalpost of seemingly automatic skilled reading.

Although this study was partly motivated by previous reports that 'looking-at-nothing' modulates paired associate learning, we did not detect any such significant effects in this dataset. Contributing factors may simply be power and webcam-based eyetracking data quality: though the regression analysis identified trends in the expected directions, webcam-based eyetracking is still in its infancy, and thus, due to the inevitable increase in noise engendered by remote webcam-based eyetracking, the method used in our study may potentially not have detected fixations as consistently as specialized laboratory hardware.

References

Albano, D., Garcia, R. B., & Cornoldi, C. (2016). Deficits in working memory visual-phonological binding in children with dyslexia. *Psychology & Neuroscience*, *9*(4), 411.

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). Gorillas in our Midst: Gorilla.sc. *Behavior Research Methods*.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive psychology*, *4*(2), 207-228.

Brockmole, J. R., & Franconeri, S. L. (2009). Introduction to the special issue on binding. *Visual Cognition*, 17, 1-9.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of reading*, *9*(2), 167-188.

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing: An Interdisciplinary Journal*.

El-Kalliny, M. M., Wittig, J. H., Sheehan, T. C., Sreekumar, V., Inati, S. K., & Zaghloul, K. A. (2019). Changing temporal context in human temporal lobe promotes memory of distinct episodes. *Nature communications*, 10(1), 1-10.

Jones, M. W., Kuipers, J. R., Nugent, S., Miley, A., & Oppenheim, G. (2018). Episodic traces and statistical regularities: Paired associate learning in typical and dyslexic readers. *Cognition*, 177, 214-225.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, 42(3), 627-633.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). controlling for participants' Viewing Distance in Large-Scale, psychophysical online experiments Using a Virtual chinrest. *Scientific reports*, 10(1), 1-11.

Litt, R. A., & Nation, K. (2014). The nature and specificity of paired associate learning deficits in children with dyslexia. *Journal of Memory and Language*, *71*(1), 71-88.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. Psychological review, 89(1), 60-94.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. *Proceedings of the Twenty-Fifth IJCAI* 2016.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, 76(3), 269-295.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451-465.

Scholz, A., Klichowicz, A., & Krems, J. F. (2018). Covert shifts of attention can account for the functional role of "eye movements to nothing". *Memory & Cognition*, 46(2), 230-243.

Scholz, A., Mehlhorn, K., & Krems, J. F. (2016). Listen up, eye movements play a role in verbal memory retrieval. *Psychological research*, 80(1), 149-158.

Scorpio, K. A., Islam, R., Kim, S. M., Bind, R., Borod, J. C., & Bender, H. A. (2018). Paired-Associate Learning. In *Encyclopedia of Clinical Neuropsychology*. Springer.

Snowling, M., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia*, *18*(1), 1-15.

Toffalini, E., Tomasi, E., Albano, D., & Cornoldi, C. (2018). The effects of the constancy of location and order in working memory visual–phonological binding of children with dyslexia. *Child Neuropsychology*, 24(5), 671-685.

Tulving, E. (1972). 12. Episodic and Semantic Memory. *Organization of memory*/Eds E. Tulving, W. Donaldson, NY: Academic Press, 381-403.

Wang, H. C., Wass, M., & Castles, A. (2017). Paired-associate learning ability accounts for unique variance in orthographic learning. *Scientific Studies of Reading*, 21(1), 5-16

Wantz, A. L., Martarelli, C. S., & Mast, F. W. (2016). When looking back to nothing goes back to nothing. *Cognitive processing*, 17(1), 105-114.