Special Issue Article

QJEP

Quarterly Journal of Experimental Psychology
2018, Vol. 71(1) 37–45

DOI: 10.1080/17470218.2016.1240814
qjep@sagepub.com
SAGE

# Dynamic adjustment of lexical processing in the lexical decision task: Cross-trial sequence effects

**David A Balota[1], Andrew J Aschenbrenner[1] and Melvin J Yap[2]**


## Abstract

There has been growing interest in dynamic changes in the lexical processing system across trials, which have typically been assessed via linear mixed effect modelling. In the current study, we explore the influence of previous trial lexicality and previous trial perceptual degradation on the effect of lexicality and degradation on the current trial. The results of analyses of three datasets (two previously published studies and a new study) provide evidence for a robust four-way interaction among previous trial lexicality and degradation and current trial lexicality and degradation effects. Discussion emphasizes how priming of relevant dimensions (clear vs. degraded or word vs. nonword) within the experimental context modulates the influence of degradation on the current trial as a function of lexicality. These results suggest that in lexical decision there are robust lingering effects of the previous stimulus and response that carry over to the current stimulus and response, and participants cannot tune task-related systems to only the present trial. Importantly, although these complex relationships are theoretically important regarding lexical and decision level processes, these complexities also reinforce Keith Rayner's emphasis on on-line eye-tracking measures during reading as the most straightforward window into word-level processes engaged during reading.



### Keywords

Lexical decision; Sequential effects

Received: 14 July 2016; accepted: 18 September 2016


Keith Rayner was an expert at identifying important problems and tackling them in the most straightforward manner. As a postdoc, one of the current authors (D.B.) remembers him often saying that if you want to study how people read words, then you should study reading and stop spending so much time studying speeded pronunciation and lexical decision performance. Although Keith appreciated the insights that one can glean from speeded word pronunciation and isolated lexical decision performance, especially in relation to computational models of visual word recognition, these tasks bring on-line many operations that may have limited relevance to reading (e.g., decision processes in lexical decision). Likewise, one must acknowledge that reading brings on-line many operations that may have limited relevance to isolated visual word recognition (e.g., comprehension, syntactic analyses, eye-movement control).

Although there are many differences between reading and isolated lexical decision (the target of the current paper), one difference that has, to our knowledge, not been fully appreciated in the literature involves the temporal dynamics of the tasks. Specifically, reading is a continuous measure of language processing, whereas visual word recognition tasks (e.g., lexical decision, pronunciation, and semantic verification) are more punctate, trial-by-trial procedures that experimental psychologists have developed to isolate components of important aspects of cognition, including pattern recognition, lexical/semantic processing, attention, and memory. Of course, in reading, one does not simply process the current word; instead, the reader is processing linguistic information at many levels (lexical, syntactic, semantic) in parallel, based on the previous words

[1]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO, USA
[2]Department of Psychology, National University of Singapore, Singapore

**Corresponding author:**
Andrew J. Aschenbrenner, Washington University in St. Louis, St. Louis, MO 63130, USA.
Email: a.aschenbrenner@wustl.edu

and even some characteristics of upcoming words (e.g., Rayner, Pollatsek, Ashby, & Clifton, 2012). In contrast, researchers who use trial-by-trial procedures often make inferences based on the assumption that each trial is a unique event. Indeed, we often collapse across trials within a given condition to obtain an estimate of performance in that condition.

Although the trial-by-trial approach has dominated much of cognitive psychology and has led to important theoretical insights, researchers have recently capitalized on analytic techniques (e.g., linear mixed-effects analyses) that allow one to explore how the processing system adjusts across trials to influence performance. This work follows some early classic work with the Stroop task. Specifically, the Stroop effect on the current trial is smaller when the congruency of the previous trial is incongruent than when it is congruent (e.g., Gratton, Coles, & Donchin, 1992). This cross-trial sequence effect has received considerable attention because it suggests dynamic adjustments of control across adjacent trials and can provide important insights into the flexibility of the cognitive system. There are currently multiple theoretical accounts of such cross-trial effects, including (a) adaptations to the statistical environment (Kinoshita, Forster, & Mozer, 2008; Kinoshita, Mozer, & Forster, 2011), (b) increases in attentional control systems based on difficulty of preceding trials (Gratton et al., 1992), and (c) priming of relevant pathways (Aschenbrenner & Balota, 2015).

In the current study, we focus on cross-trial sequence effects in the lexical decision task. This work was originally motivated by an important series of observations by Masson and Kliegl (2013). These researchers tackled the influence of trial history on the relationship among three theoretically important variables in lexical decision performance: semantic priming, word frequency, and stimulus degradation. These variables are important for models of lexical processing because semantic priming interacts with both word frequency (larger priming effects for low-frequency words than high-frequency words) and stimulus degradation (larger priming effects for degraded words than clear words), but word frequency and stimulus degradation produce additive effects. Because additive effects can sometimes indicate that variables are influencing separable stages in processing (based on additive factors logic, Sternberg, 1969; but see McClelland, 1979), the additivity of stimulus degradation and word frequency has been particularly difficult to accommodate within current computational models that do not allow for separable stages. Masson and Kliegl argued that these apparent additive effects may actually be due to cross-trial sequence effects in lexical decision performance. For example, in their first experiment, they found underadditive effects of stimulus quality and word frequency when the previous trial was presented in *clear* format, whereas they found overadditive effects of stimulus quality and word frequency when the previous trial was presented in *degraded* format. When one collapses across the previous trial, as in the standard lexical decision experiment, one finds additive effects of the two variables—but, of course, this is simply because one is collapsing across the two levels of trial history that ultimately cancel each other out. Similar patterns were observed in a second experiment. If, indeed, this observation is correct, then a major pattern of results that has been used to constrain models of lexical decision performance (see for example, Borowsky & Besner, 1993; Yap, Balota, Tse, & Besner, 2008) would be called into question, and the implications of such cross-trial effects may extend to other variables and tasks.

Subsequent research suggested that the cross-trial modulation of the additive effects of word frequency and stimulus degradation may reflect the presence of a semantic priming manipulation within the same experiment. Specifically, Scaltritti, Balota, and Peressotti (2013; also see Borowsky & Besner, 1993) have shown that the presence of related primes can induce a list-wide relatedness checking process in lexical decision task that can modulate the presence of additive effects of word frequency and stimulus degradation. In further support of this argument, Balota, Aschenbrenner, and Yap (2013) reported analyses of three lexical decision studies in the available literature and showed that the additive effects of word frequency and stimulus degradation were not modulated by the characteristics of the previous trial when there were no semantically related primes embedded within the list. However, more recent research by Masson, Rabe, and Kliegl (2017) suggests that the presence of semantically related primes may not be the critical variable in the earlier study. Specifically, Masson et al. have reported two experiments (including the semantic priming manipulation) that failed to replicate their original effect. Hence, they suggested that this aspect of their data is likely to have been a Type I error.

Despite the ephemeral effects of cross-trial influences on the frequency × degradation additivity described above, Masson and Kliegl (2013) reported additional cross-trial sequence effects that appear quite consistent and are important for any model of lexical decision performance, and potentially for other lexical processing tasks. Specifically, for word stimuli, they reported a robust interaction among current stimulus quality, previous stimulus quality, and previous lexicality (also see Masson et al., 2017). This interaction indicates that responses on the current trial were faster when the previous trial had the same level of degradation, but *only* occurred when the previous trial was a word and not when the previous trial was a nonword. Although this pattern was not the focus of the Masson and Kliegl study, it is potentially theoretically important because it brings into question the basic premise of being able to measure the influence of a variable on Trial N, without taking into account previous trial history. These results also strongly suggest that there are carryover effects of decision processes that may be unique to the lexical decision task. Indeed, in their reanalysis of previous studies, Balota et al. (2013) also reported evidence for this three-way interaction.

Before discussing the implications of this three-way interaction, we further explore the strength of this interaction in three different datasets. Two of the data-sets were taken from published experiments conducted at Washington University (see Yap et al., 2008), and the third from a new experiment conducted at the National University of Singapore. Here, we describe the first two studies briefly, since full details are presented in the original report and subsequent analyses are also presented in Balota et al. (2013). In the first experiment of Yap et al. (2008), 28 participants were administered a lexical decision task that included a within-participant manipulation of degradation, with 50 high-frequency and 50 low-frequency words at each level of degradation, along with 100 clear and 100 degraded pronounceable non-words (e.g., FLIRP). Yap et al. (2008, Experiment 2, $N = 56$) used the same word stimuli but included 200 pseudohomophones as the non-words (e.g., BRANE).

In addition to examining the three-way interaction, we also explore the influence of current trial lexicality in the current analyses. The influence of current trial lexicality has not been examined in previous studies. Remarkably, as discussed below, in all three datasets there is a robust four-way interaction that occurs in both transformed data and non-transformed data. The generality of these findings in both transformed and non-transformed data analyses is particularly important because Balota et al. (2013) have shown that the common practice of transforming data via inverse or log transformations (commonly used to normalize the residuals in linear mixed effects analyses) can modulate the presence of additive or interactive effects. Because reaction time distributions are positively skewed and the shape of the distribution may be informative (see Ratcliff, 1979), transformations that change the shape of the distributions may lose important information. Indeed, because of this concern, Lo and Andrews (2015) have recently developed a generalized linear mixed-effect model to provide a solution to this problem without the need for data transformations. Importantly, for the present results the influence of data transformation is not critical, because we obtained the same results on transformed and non-transformed data, further supporting the robustness of this pattern.

## Re-analysis of Yap et al. (2008, Experiments 1 and 2)

Results from the previous Yap et al. experiments were analysed with linear mixed effects models (LME) using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). Stimulus quality of the current trial, lexicality of the current trial, stimulus quality of the previous trial, and lexicality of the previous trial were entered as factors (coded as -0.5/+0.5 contrasts) and random intercepts of subjects and items were also included.[1] Significance was determined using t tests of regression coefficients with degrees of freedom estimated using the Satterthwaite approximation with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015). Follow-up tests of the highest order interaction were conducted by running the same analyses within levels of current trial lexicality (i.e., separately for word and nonword targets).

The full datasets from Yap et al. (2008, Experiments 1 and 2) are displayed in the top row of Figure 1. Of course, both experiments produced highly reliable effects of degradation (78 ms, $p < .001$; 115 ms, $p < .001$), and lexicality (107 ms, $p < .001$; 114 ms, $p < .001$), for Yap et al.'s Experiments 1 and 2, respectively. Most importantly for the present discussion, for each experiment there was a highly reliable four-way interaction among current trial lexicality and degradation and previous trial lexicality and degradation [Experiment 1, $\beta = -160.01$, $t(9881) = -6.10$, $p < .0001$; Experiment 2, $\beta = -88.50$, $t(19482) = -3.59$, $p = .0003$]. Follow-up analyses of the three-way interaction among previous trial degradation and previous trial lexicality and current trial degradation yielded reliable effects for word targets in both experiments [Experiment 1: $\beta = 91.76$, $t(5021) = 5.39$, $p < .0001$; Experiment 2, $\beta = 60.95$, $t(10047) = 3.79$, $p = .0001$]. Turning to the nonwords, the same three-way interaction was reliable for Experiment 1, $\beta = -70.44$, $t(4851) = -3.60$, $p = .0003$,but did not reach significance for Experiment 2, $\beta = -28.69$, $t(9450) = -1.56$, $p = .12$.

Of course, describing the nature of a four-way interaction is complex, to say the least. In an attempt to simplify this, Figure 2 plots the degradation effect for words and nonwords on the current trial, as a function of previous trial lexicality and previous trial degradation. First, consider the word trials, which produce a very systematic pattern. As noted, the three-way interaction among previous trial degradation, previous trial lexicality, and current trial degradation is highly reliable in both datasets and replicates the general pattern observed by Masson and Kliegl (2013) for word stimuli. Specifically, the effect of degradation on the current trial is larger when the previous trial was a clear word or a degraded nonword than when the previous trial was a degraded word or clear nonword. Turning to the nonword data, one can see just the opposite pattern, although the effects are smaller. Specifically, when the previous trial is a clear word or a degraded nonword, the degradation effect on the current trial for nonwords is actually the smallest. Indeed, the previous trial conditions that produce the largest degradation effects for words on the current trial produced the smallest degradation effects for non-words on the current trial. Clearly, decision and degradation are modulating performance in lexical

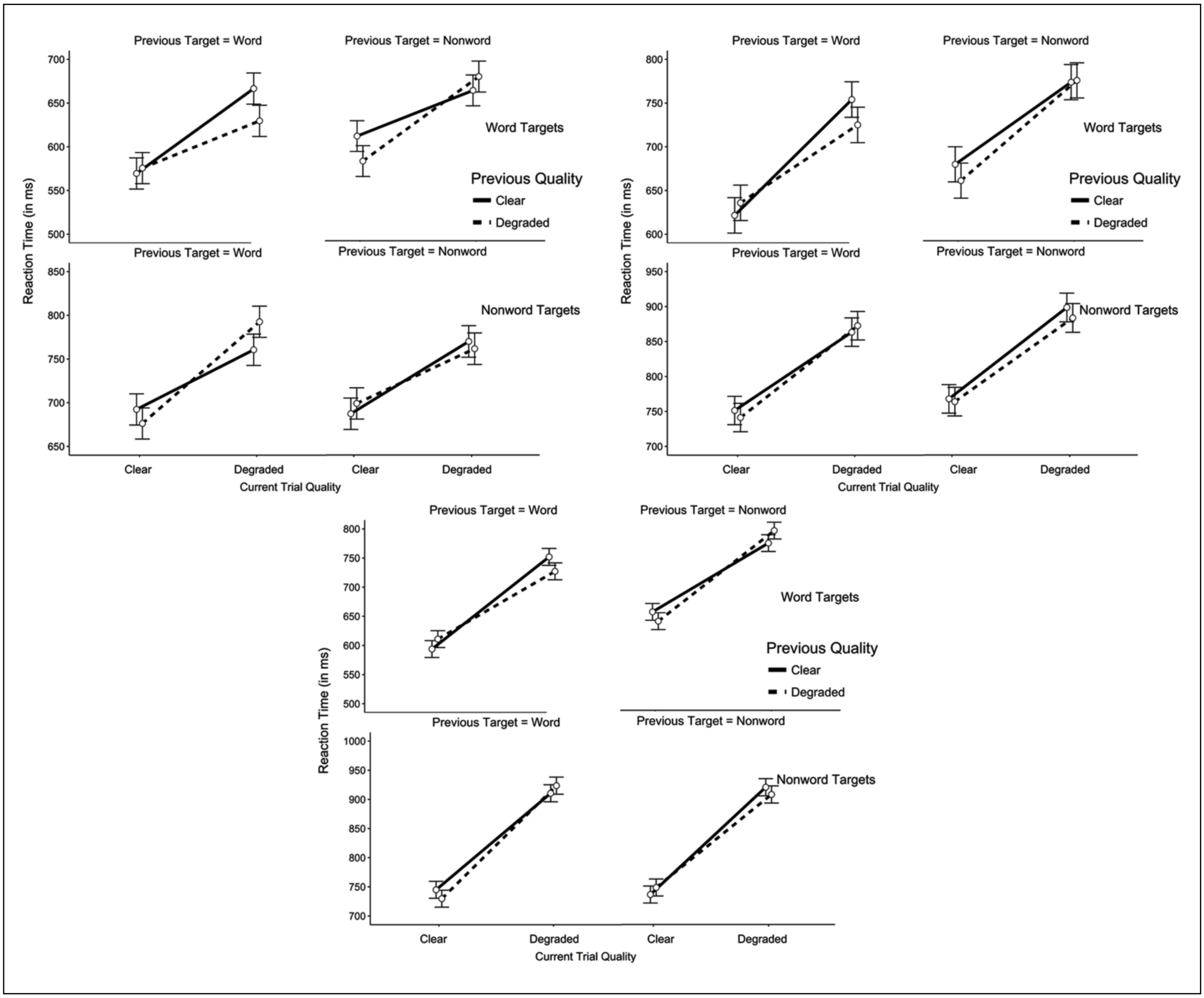


**Figure 1.** Reaction time as a function of current trial degradation and lexicality and of previous trial degradation and lexicality. Error bars are the SE of the mean for that condition. Yap et al. (2008) Experiment 1 is the top left panel, Yap et al. (2008), Experiment 2 is the top right panel and the Current Experiment is the bottom panel.

decision across trials. We now turn to the results from a new study that further explores this interaction in a different subject population. Given the complexity of a four-way interaction, it is important to ensure the strength and direction of this pattern.

## Experiment

### Method

*Participants.* A total of 160 undergraduates from the National University of Singapore participated for SGD15. The participants' first language was English, and they had normal or corrected-to-normal vision. The lexical decision task was embedded in a battery of other measures that included memory scanning, operation span, and antisaccade tasks. Although all participants have English as their primary language, most were also fluent in a second language, including Mandarin, Malay, and/or Tamil.

*Design.* Stimulus quality (clear or degraded) and word frequency (high or low) were manipulated within participants, the primary dependent primary dependent variable being RT. It is worth noting that a similar analysis on accuracy rates did not compromise the interpretation of these data.

*Stimuli.* A total of 120 words (60 high-frequency and 60 low-frequency) were selected (see Table 1 for descriptive statistics). High- and low-frequency words were matched on number of letters, number of syllables, number of morphemes, neighbourhood size, Levenshtein distance (Yarkoni, Balota, & Yap, 2008), log-transformed subtitle-based contextual diversity (Brysbaert & New, 2009), and concreteness (Brysbaert, Warriner, & Kuperman, 2014). In addition, 120 nonwords were generated using the multilingual pseudoword generator, Wuggy (Keuleers & Brysbaert, 2010). These non-words were matched to their yoked controls on number of syllables and number of letters, as well as subsyllabic structure and transition frequencies.

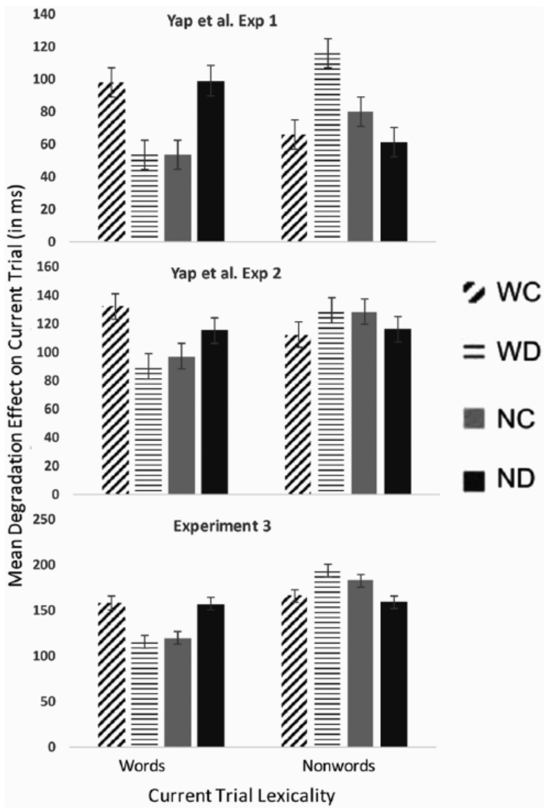


**Figure 2.** Degradation effects for words and nonwords as a function of previous trial degradation and lexicality. WC = previous word, previous clear trials, WD = previous word, previous degraded trials, NC = previous nonword, previous clear trials, ND = previous nonword, previous degraded trials. Error bars are the SE of the difference between Means for the Degraded minus the Clear current trial conditions.

*Procedure.* PC-compatible computers running E-prime software (Schneider, Eschman, & Zuccolotto, 2001) were used for stimulus presentation and data collection. Participants were individually tested in sound-attenuated cubicles; they were positioned approximately 60 cm from the computer screen. Participants were instructed to decide whether the letter string presented formed a word or nonword by making the appropriate button press (slash key for words and "Z" key for nonwords). Participants were encouraged to respond quickly but not at the expense of accuracy. There were 16 practice trials, followed by four experimental blocks of 60 trials each, with breaks between blocks. The order in which stimuli were presented was randomized anew for each participant. Stimuli were presented in uppercase 18-point Courier New, and each trial comprised the following order of events: (a) a fixation point (+) at the centre of the monitor for 400 ms, (b) a blank screen for 200 ms, and (c) the target. The target remained on the screen until the participant responded. For incorrect responses, the word "Incorrect" was displayed slightly below the fixation point for 1600 ms. Using the same degradation procedure as in Yap et al. (2008), half the targets were degraded by rapidly alternating letter strings with a randomly generated mask of the same length. For example, the mask @$#&% was presented for 14 ms, followed by a five-letter target word for 28 ms; the two alternated rapidly until a response was detected. Mask patterns were consistent within a trial and were generated from random permutations of the following symbols: &@?!$*%#?. Across participants, targets were counterbalanced across degraded and clear conditions.

## Results

As shown in the bottom row of Figure 1, the results from the new study replicate the general pattern obtained in the

**Table 1.** Descriptive characteristics of stimuli used in experiment.

| Variable | High frequency (*N* = 60) | | Low frequency (*N* = 60) | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| Log SUBTL-CD | 3.32 | 0.36 | 2.32 | 0.46 |
| Log SUBTL-WF | 3.66 | 0.48 | 2.51 | 0.53 |
| Letters (N) | 4.80 | 0.95 | 4.82 | 0.79 |
| Syllables (N) | 1.42 | 0.50 | 1.30 | 0.46 |
| Morphemes (N) | 1.08 | 0.28 | 1.07 | 0.25 |
| Orthographic neighbourhood size | 4.32 | 4.23 | 4.32 | 4.14 |
| Phonological neighbourhood size | 12.67 | 10.04 | 11.07 | 9.92 |
| Orthographic Levenshtein distance 20 | 1.72 | 0.34 | 1.74 | 0.32 |
| Phonological Levenshtein distance 20 | 1.46 | 0.44 | 1.57 | 0.46 |
| Concreteness | 4.01 | 0.88 | 3.99 | 0.86 |

previous studies. Importantly, there was again a highly reliable four-way interaction among current trial lexicality and degradation and previous trial lexicality and degradation, $\beta = 132.2$, $t\ (33250) = -6.71$, $p < .0001$. The follow-up comparisons indicate that the three-way interactions are reliable both for word targets, $\beta = 81.00$, $t\ (17177) = 6.44$, $p < .0001$, and for nonword targets, $\beta = -49.32$, $t(16168) = -3.33$, $p = .0009$. More importantly, as shown in the bottom panel of Figure 2, when considering the degradation effect, one again finds the same pattern observed in Experiments 1 and 2. Clearly, this complex four-way interaction is highly robust across participant populations and stimuli.

## General discussion

The results from these analyses are quite clear. There are robust cross-trial carryover effects of lexicality and degradation onto the current trial. These results are consistent with Masson and Kliegl's (2013; also see Masson et al., 2017) observation of a three-way interaction among current trial degradation and previous trial degradation and lexicality and, importantly, indicate that this pattern does not extend to nonwords, which actually produce the opposite pattern, hence leading to the four-way interaction. This four-way interaction is robust in all three of the datasets examined. Moreover, because the four-way interaction occurs for both transformed and non-transformed analyses, we can be confident that transformation of the data is *not* modulating the pattern.

It should be noted that the cross-trial effects observed in the current analyses are relatively large. This is seen most clearly in the word data shown in Figure 2. Specifically, the effect of degradation on the current trial changes by 40–50 ms depending upon degradation and lexicality in the previous trial. Indeed, in the reanalysis of the Yap et al. (2008) Experiment 1 data, the degradation effect nearly doubles depending on the previous trial degradation and lexicality. Hence, these are relatively large modulations of performance on the current trial, based on the previous trial characteristics.

What are the implications of the present results? At the most general level, the present results converge on the notion that when considering performance in standard trial-by-trial experiments, one needs to be concerned about carryover effects from the previous trial, which modulates effects on the current trial. We are obviously not the first to note this. For example, as discussed, there is already a rich literature on cross-trial sequence effects in attentional selection tasks (e.g., Gratton et al., 1992).

Lexical decision performance is not typically considered as placing high demands on attentional selection, at least not in the same way as Stroop incongruent trials do, so the mechanisms accounting for the current results may not be the same. However, there has been work within visual word recognition that does show an influence of difficulty of previous trials on current trial performance. These data have been accommodated by the Adaptation to the Statistics of the Environment (ASE) model (e.g., Kinoshita et al., 2008). Specifically, if the response on the previous trial has been relatively slow, then the initiation of the response on the current trial will be later in time than when the previous trial has been relatively fast. Hence, participants adapt to the difficulty of the trials within the experimental context. Possibly, the present results simply reflect this cross-trial carryover effect of difficulty. Although this model is useful, it cannot accommodate the present results, because the influence of the difficulty of the previous trial changes depending on whether the current trial is a word or a nonword. Masson and Kliegl (2013) also note that the ASE model cannot account for the three-way interaction for only word stimuli. Hence, it appears that one will need to appeal to a different mechanism to account for the present cross-trial effects.

Although a complete account of these results will likely demand a formal model of lexical decision, such as the Bayesian Reader (Norris, 2006) or the Diffusion model (Ratcliff, Gomez, & McKoon, 2004), a descriptive account seems reasonable at this point. The framework presented here is based on the similarity of the previous and current trial along the dimensions of lexicality and degradation. The notion is that the dimensions of lexicality and degradation on the current trial may be primed by the previous trial. By priming, we mean that the system is tuned to the dimensions of the upcoming stimulus by the previous trial. These dimensions include the stimulus–response mapping of word or nonword and the perceptual difficulty of processing degraded or clear stimuli. So, for a nonword in the degraded presentation, the system is tuned for nonwords and for degraded stimuli. The notion here is that there is a lingering effect of this tuning on the following trial. In describing this priming account, we consider the degradation effects in Figure 2, since this variable is of greatest interest (at least more so than the lexicality effect) to researchers in visual word recognition.

First, consider the conditions that produce the largest degradation effect on the current trial. As shown in Figure 2, the degradation effect on the current word trial is largest when the previous trial is either a clear word or a degraded nonword. This pattern is not surprising. For example, if the previous trial is a clear word, then on the following trial participants are primed for a clear word but not for a degraded word, and hence there is a relatively large degradation effect for words. If the previous trial is a degraded nonword, the participant is primed for a nonword response for this degraded stimulus, which is inconsistent with the word response, and so there is a slowdown on degraded trials, thereby also increasing the degradation effect. In the same vein, how might this perspective account for the relatively small degradation effects for words following degraded words and clear nonwords? This pattern also follows a similar logic. Specifically, the degradation effect following degraded words should be relatively small because the participant has just received a degraded word stimulus and hence this should prime processing of another degraded word stimulus and potentially slow processing of a clear stimulus, thereby producing a relatively small degradation effect. Why should the degradation effect following clear nonwords be relatively small? Here, the clear word stimulus on the current trial might be slowed due to the primed clear nonword response, thereby slowing response latency in the clear condition, which would again decrease the degradation effect.

Of course, an interesting aspect of the present results is that the previous trial characteristics actually produce the opposite pattern when considering the current trial degradation effect for nonwords. This is not surprising, given the above framework. For example, there is a relatively large effect of degradation on the current nonword trials if the previous trial was a clear nonword. This makes sense because the previous clear nonword would prime the current clear nonword and disrupt the current degraded nonword, thereby decreasing the degradation effect. Likewise, if the previous trial was degraded, then one might expect the degradation effect on the current trial to be relatively small, because this will disrupt clear nonword processing on the current trial and prime degraded nonword processing on the current trial, thereby reducing the degradation effect, as observed.

It is noteworthy that in each of these admittedly *post hoc* accounts there is an influence of compatibility of both degradation of the previous and current trial and the lexicality of the previous and current trial in modulating the current trial degradation effect. Hence participants appear to be carrying across trials (priming) both dimensions of the experimental context. In this light, one may interpret these results within a model proposed by Turner, Van Zandt, and Brown (2011) which suggests that similarity of stimuli and responses across trials will strengthen the connection between a particular response and a particularly condition, via a type of Hebbian learning mechanism, which could be in play because in all experiments analysed, feedback was given when an error was made. An alternative approach may involve trial-by-trial adjustments to the criteria, based on the "monitor and adjust" principle of the Leaky Competing Accumulator Model of lexical decision developed by Dufau, Grainger, and Ziegler (2012). Here, cross-trial adjustments are made in the "word" and "nonword" response criteria based on the accuracy of the previous trial, again based on feedback. Both models rely heavily on the importance of feedback and error trials, and so the cross trial adjustments would be particularly strong after error trials. Although it is beyond the scope of this report to evaluate these models in accounting for present results, the consistency of the results across experiments affords a clear target for such models.

We have been emphasizing the cross-trial effects of degradation and lexicality in the present study. However, in each of the experiments presented there was a large effect of word frequency (Yap et al., 2008: Experiment 1,52 ms, $p < .001$; Experiment 2, 75 ms, $p < .001$; Current Experiment, 53 ms, $p < .001$). One might naturally ask whether there is evidence of cross-trial effects of word frequency when frequency is included in the models. We explored this in a series of LME analyses, and the only analysis that produced a hint of an influence on frequency was a Previous Trial Degradation × Current Trial Frequency interaction in the current experiment, indicating that the frequency effect was slightly larger (by 12 ms) following degraded stimuli than following clear stimuli ($p < .03$). Because this marginal effect occurred only in one dataset and was related to vocabulary scores ($p < .04$, in this bi/multilingual population), we believe this effect is very weak (or spurious) and so may not generalize to other populations.

It is potentially noteworthy that the current word frequency effect is only minimally influenced by previous trial history of lexicality and degradation. This suggests that there may be differential sensitivity of variables across trials. Because frequency is additive with stimulus degradation in all of these datasets, one might not expect degradation to interact with frequency across trials. In contrast, stimulus degradation and lexicality produce large interactions in all three datasets in the current trial data. Possibly, the extent to which one finds cross-trial effects will be partly modulated by the presence of interactions in the current trial of the manipulated variables. For example, because word frequency interacts with semantic priming, one might expect cross-trial interactions of these two variables. Clearly, further work needs to be conducted to explore how the lexical processing system adjusts across trials within an experiment.

Of course, we are clearly not the first to emphasize the importance of cross-trial sequence effects in the lexical decision task. For example, in addition to the Masson and Kliegl (2013) study discussed above, Perea and Carreiras (2003), Ratcliff and McKoon (1995), and Lima and Huntsman (1997) have all reported evidence of cross-trial sequence effects in lexical decision performance across various conditions. However, the strength of the present study is the consistent and robust four-way interaction among degradation and lexicality effects using linear mixed effects modelling. Hence, we believe that the present results clearly add to the arsenal of findings that any complete model of lexical decision will need to accommodate.

At this point, one may ask whether the present interactive effects would generalize to other lexical processing tasks. We know of only one study that addresses this possibility. O'Malley and Besner (2013) reported a similar study and analyses in which participants pronounced words and nonwords aloud instead of making lexical decisions. The results of their study yielded no evidence of a three-way interaction among previous trial lexicality, previous trial degradation, and current trial degradation for current word trials. They did not report the results from the current nonword trials. Of course, the influence of lexicality on lexical decision is much larger than on pronunciation, and so this may limit the ability to detect the interaction. Alternatively, it may be the case that the present interactive effects are reflective of requiring forced binary decisions from the participants. In this light, an examination of the semantic verification task would be useful to further examine the interactive effects when a binary decision is required in another major lexical processing task.

In closing, we began this paper noting how Keith Rayner cautioned researchers about making inferences about *word reading* from tasks such as lexical decision. If one is interested in studying how people read words, then it is best to do so while they are actually reading words embedded in text. The robust four-way interaction in which previous trial degradation and lexicality influence current trial degradation and lexicality is clearly an example of an idiosyncratic aspect of the lexical decision task. This finding may also have more general implications for the trial-by-trial procedures commonly employed by cognitive psychologists in other tasks. In this light, Keith was correct (as was typically the case) in noting that lexical decision is not simply word reading. Although the present results may have limited value in informing how participants read words in text, we also believe that Keith would have appreciated the utility of such work in understanding how people adjust the lexical processing system across time within this particular task and see the potential generality of this approach to explore such dynamic changes in other lexical processing tasks (e.g., semantic verification), cognitive tasks in other domains (e.g., speeded episodic recognition), and perhaps even in reading.

## Acknowledgements

We thank Simon Liversedge, Denis Drieghe, and two anonymous reviewers for helpful comments on a previous version of this manuscript.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This research was supported in part by an National Institute on Aging (NIA) Aging and Development Training Grant [grant number T32 AG000030] awarded to Dave Balota.

## ORCiD

Andrew J. Aschenbrenner http://orcid.org/0000-0002-4317-7282
Melvin J. Yap http://orcid.org/0000-0002-5372-4323

## Note

1. With the advent of LMEs as a primary analytical framework for psycholingustic experiments, there has been much debate on how best to determine the random effects structure, with some authors suggesting it is optimal to include all possible (maximal) random effects (Barr, Levy, Scheepers, & Tily, 2013). This assertion has not gone unchallenged, but, most importantly for the present purposes, significance tests of fixed effects to do not appear to be substantially changed by the inclusion of additional random effects (Bates, Kliegl, Vasishth, & Baayen, 2015). Indeed, when we re-tested the major analyses in the current data sets (i.e., the four-way interaction of current and previous degradation with current and previous lexicality, as well as the three-way current and previous degradation with previous lexicality within levels of current trial lexicality) with additional, uncorrelated random slopes of each factor, the same conclusions were reached as in our primary analysis.

## References

Aschenbrenner, A. J., & Balota, D. A. (2015). Interactive effects of working memory and trial history on Stroop interference in cognitively healthy aging. *Psychology and Aging*, *30*, 1–8. doi:10.1037/pag0000012

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *39*, 1563–1571. doi:10.1037/a0032186

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*, 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015). Parsimonious mixed models. Retrieved from arXiv:1506.04967 (stat.ME).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi:10.18637/jss.v067.i01

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *19*, 813–840.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. doi:10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. doi:10.3758/s13428-013-0403-5

Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say "No" to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *38*, 1117–1128. doi:10.1037/a0026948

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*, 480–506.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627–633. doi:10.3758/BRM.42.3.627

Kinoshita, S., Forster, K. I., & Mozer, M. C. (2008). Unconscious cognition isn't that smart: Modulation of masked repetition priming effect in the word naming task. *Cognition*, *107*, 623–649. doi:10.1016/j.cognition.2007.11.011

Kinoshita, S., Mozer, M. C., & Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology: General*, *140*, 622–636. doi:10.1037/a0024230

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmertest: Tests in linear mixed effects models. R Package version 2.0-29. Retrieved from http://CRAN.R-project.org/package=lmerTest

Lima, S. D., & Huntsman, L. A. (1997). Sequential dependencies in the lexical decision task. *Psychological Research*, *60*, 264–269.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. doi:10.3389/fpsyg.2015.01171

Masson, M. E. J., & Kliegl, R. (2013). Modulation of additive and interactive effects in lexical decision by trial history. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *39*, 898–914.

Masson, M. E., Rabe, M. M., & Kliegl, R. (2017). Modulation of additive and interactive effects by trial history revisited. *Memory & Cognition*. 45, 480–492. doi:10.1037/a0029180

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327–357. doi:10.1037/0033-295X.113.2.327

O'Malley, S., & Besner, D. (2013). Reading aloud: Does previous trial history modulate the joint effects of stimulus quality and word frequency? *Journal of Experimental Psychology: Learning, Memory & Cognition*, *39*, 1321–1325. doi:10.1037/a0031673

Perea, M., & Carreiras, M. (2003). Sequential effects in the lexical decision task: The role of the item frequency of the previous trial. *The Quarterly Journal of Experimental Psychology*, *56A*, 385–401. doi:10.1080/02724980244000387

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182. doi:10.1037/0033-295X.111.1.159

Ratcliff, R., & McKoon, G. (1995). Sequential effects in lexical decision: Tests of compound-cue retrieval theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1380–1388.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *The psychology of reading*. New York: Psychology Press.

Scaltritti, M., Balota, D. A., & Peressotti, F. (2013). Exploring the additive effects of stimulus quality and word frequency: The influence of local and list-wide prime relatedness. *The Quarterly Journal of Experimental Psychology*, *66*, 91–107. doi:10.1080/17470218.2012.698628

Schneider, W., Eschman, A., & Zuccolotto, A. (2001). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*, 421–457.

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review*, *118*, 583–613. doi:10.1037/a0025191

Yap, M. J., Balota, D. A., Tse, C. S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *34*, 495–513. doi:10.1037/0278-7393.34.3.495

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979. doi:10.3758/PBR.15.5.971