Journal of Memory and Language 93 (2017) 203–216

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Goals and strategies influence lexical prediction during sentence comprehension

Trevor Brothers *, Tamara Y. Swaab, Matthew J. Traxler

University of California, Davis, USA

ARTICLE INFO

Article history:
Received 21 October 2015
revision received 8 October 2016

Keywords:
Prediction
Anticipation
Automaticity
N400
Post-N400 positivity
Self-paced reading

ABSTRACT

Predictive processing is a critical component of language comprehension, but exactly how and why comprehenders generate lexical predictions remains to be determined. Here, we present two experiments suggesting that lexical prediction is influenced by top-down comprehension strategies, and that lexical predictions are not always generated automatically as a function of the preceding context. In Experiment 1 ($N = 24$), participants read predictable and unpredictable sentence-final words while EEG was recorded from the scalp. When comparing two different sets of task instructions, the neural effects of cloze probability were enhanced when predictive processing was emphasized. In Experiment 2 ($N = 252$), participants read predictable and unpredictable sentence continuations in a self-paced reading task, and the overall validity of predictive cues was manipulated across groups using a separate set of filler sentences. There was a linear relationship between the benefits of a constraining sentence context and the global validity of predictive cues. Critically, no reading time benefits were observed as *prediction validity* approached zero. These results provide important constraints for theories of anticipatory language processing, while calling into question prior assumptions about the automaticity of lexical prediction.

© 2016 Published by Elsevier Inc.

Introduction

In the past decade there has been a shift in our understanding of how readers construct meaning during online comprehension. While it has long been appreciated that context can affect the comprehension of individual words, recent studies suggest that language comprehenders can also use contextual constraints to actively predict upcoming words in a discourse (Altmann & Mirković, 2009; Brothers, Swaab, & Traxler, 2015; Huettig, 2015; Kutas, DeLong, & Smith, 2011). Whether it is a listener making anticipatory eye-movements to a predicted object in a scene (Altmann & Kamide, 1999), a reader skipping over a predictable word in a passage of text (Rayner, Slattery, Drieghe, & Liversedge, 2011), or an eager interlocutor finishing a friend's sentence (Howes, Healey, Purver, & Eshghi, 2012), anticipation appears to be a fundamental mechanism in both language comprehension and human cognition more generally (Clark, 2013).

One important method for assessing the neural time-course of predictive processing is the event-related potential (ERP) technique. In previous ERP studies, the N400 component has been shown to be sensitive to the difficulty of processing meaningful stimuli, including words, pictures, and linguistic gestures. In addition to a number of low-level, lexical variables (Kutas & Federmeier, 2011; Van Petten, 2014), one of the primary factors that influences N400 amplitude during reading is the predictability of a word in context. Predictability in these studies is often operationalized as *cloze probability*, or the likelihood that a

* Corresponding author at: Department of Psychology, UC Davis, 1 Shields Avenue, Davis, CA 95616, USA.

E-mail address: tabrothers@ucdavis.edu (T. Brothers).

http://dx.doi.org/10.1016/j.jml.2016.10.002
0749-596X/© 2016 Published by Elsevier Inc.

participant will provide a particular word in an offline, sentence-completion task (Taylor, 1953). Previous studies have shown a strong linear relationship between cloze and N400 amplitude (DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984). It has been theorized that this N400 predictability effect is driven by pre-activation of semantic and lexical features, which ultimately results in facilitated neural processing of the predicted word (Kutas & Federmeier, 2000; Swaab, Ledoux, Camblin, & Boudewyn, 2012).

Another ERP component which is known to be sensitive to contextual probability is the post-N400 positivity (Delong, Urbach, Groppe, & Kutas, 2011; Van Petten & Luka, 2012). Unexpected sentence continuations, in addition to showing larger N400 amplitudes, also produce an enhanced late positivity over frontal electrode sites. It has been hypothesized that this component reflects the costs of encountering a disconfirmed lexical prediction (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Thornhill & Van Petten, 2012), or the costs of revising the preceding discourse context in light of new, unanticipated information (Brothers et al., 2015).

Despite recent progress, there are still many unknown variables that may influence how and when lexical predictions are generated. One critical, unanswered question is whether specific lexical predictions are activated *automatically* as a function of the preceding context, or whether predictions are generated *strategically* to facilitate upcoming text processing (for a discussion, see Huettig, 2015). According to an *automatic-activation* account of prediction, as readers process a semantically rich sentence context (*the web was spun by*...) they will rapidly activate a set of associated concepts (*spider*). Given a sufficiently constraining context, activation will accumulate for a specific lexical item, resulting in pre-activation even to the level of phonological or orthographic features. This would all occur automatically and unconsciously, simply as a function of the pre-existing links that a reader has formed in lexical-semantic memory.

Alternatively, generating specific lexical predictions at the form level may require additional processing mechanisms beyond simple spreading activation. It may be the case that true *prediction* requires the selection of some word candidates at the expense of others. Like other forms of lexical selection (cohort competition, lexical ambiguity resolution), this process would likely unfold gradually over time, requiring a resource-demanding inhibition of competing alternatives. While some forms of semantic priming are relatively resource-free, *anticipatory* priming effects may require additional attention and cognitive resources. Some support for this hypothesis comes from a dual-task study by Heyman, Van Rensbergen, Storms, Hutchison, and De Deyne (2015). In this experiment, participants showed selective impairments in anticipatory semantic priming when they were placed under high levels of concurrent working-memory load. Moreover, it appears that readers with better cognitive control abilities show larger anticipatory priming effects during both lexical decision and naming tasks (Hutchison, 2007; Hutchison, Heap, Neely, & Thomas, 2014).

If the generation of lexical predictions is *metabolically costly*, as suggested by Kuperberg and Jaeger (2016), then an efficient comprehender may suppress anticipatory mechanisms when they are no longer beneficial or relevant to the task at hand. Similarly, in an environment where predictive cues are particularly important (e.g. understanding a friend at a noisy party), comprehenders may increase the amount of resources devoted to top-down, anticipatory processing (Huettig, 2015; Lupyan & Clark, 2015).

Currently, the main source of evidence for strategic modulations of this type comes from single-word semantic priming tasks (den Heyer, Briand, & Dannenbring, 1983; Holcomb, 1988; Lau, Holcomb, & Kuperberg, 2013). For example, semantic priming effects in a lexical decision task are enhanced when the proportion of related prime-target pairs (doctor-NURSE) in the experiment is high, and reduced when the relatedness proportion is low (see Neely, 1991 for a discussion). While these studies are informative, it is still unclear if strategic mechanisms of this type can influence lexical prediction during language comprehension more generally.

In the present experiments we investigated whether top-down goals and strategies can influence lexical anticipation mechanisms during sentence and discourse processing. In Experiment 1 we used two neural indices of lexical prediction (the N400 and post-N400 positivity) to determine whether anticipatory effects at the discourse-level can be influenced by top-down comprehension strategies. In Experiment 2, we directly manipulated the environmental validity of predictive cues, to determine whether readers could strategically modulate their degree of anticipatory processing during a self-paced reading task.

One of the primary tests for establishing the automaticity of a cognitive mechanism is to determine whether it is *controllable* or *goal dependent* (Moors & De Houwer, 2006). To this end, we first recorded ERPs under two reading conditions: a standard *Comprehension* task in which participants answered true-false comprehension questions, and a separate *Prediction* task in which participants were instructed to actively anticipate the final word of each sentence. Using this task manipulation, we tested whether the standard ERP effects of cloze probability are generated automatically as a function of the preceding context, or if these effects could be altered as a function of readers' comprehension goals. If lexical predictions are subject to top-down control, we would expect an enhancement of both the benefits (N400) as well as the costs (PNP) of constraining contexts when a predictive reading strategy is emphasized.

Experiment 1

Materials and method

Participants

Twenty-four undergraduates (16 females) from the University of California, Davis participated in Experiment 1. The mean age of this group was 20.3 years (range 18–26, std = 2.3), and all were native English speakers with no history of neurological or reading deficits. These participants had normal or corrected-to-normal vision and were right-handed as determined by the Oldfield handedness inventory (1971). None had previously participated in any studies using these experimental stimuli.

Materials

The experimental stimuli were 180 critical words, which appeared as the final word of a two-sentence discourse passage (see the Appendix A for examples). For each critical word two different passages were constructed. One passage was written to moderately constrain the final critical word (*Thomas didn't like the temperature of his drink. He thought it was much too* <u>*hot*</u>), while the other passage made this critical word highly *unpredictable* but still semantically coherent (e.g. *Thomas didn't like the look of the water. He thought it was much too* <u>*hot*</u>). During an offline sentence-completion task for these materials, participants produced the final critical word for 50.7% of the medium-cloze passages, and for 0.7% of the low-cloze passages, $t(179) = 107$, $p < .001$. While these passages differed considerably in the predictability of the final critical word, they did not differ in overall constraint (51.1% vs 51.8%, $t < 1$) which was defined as the cloze probability of the most likely completion for each passage. This suggests that participants should be equally fluent in generating lexical predictions across both types of passages. For additional information regarding these stimuli see Brothers et al. (2015).

In the current experiment, each participant read a total of 180 critical passages. These participants saw equal numbers of stimuli in the Comprehension and Prediction tasks, with 60 medium-cloze and 30 low-cloze passages appearing in each condition. All of these stimuli were counterbalanced across six lists in a Latin Square design to ensure (1) that each participant saw each critical word only once, (2) that each critical word appeared in equal proportions as a high and low-cloze target, and (3) that each item appeared equally often during the Comprehension and Prediction blocks. Each participant was randomly assigned to one of these six lists at the beginning of the experiment.

Procedure

During EEG recording, participants were seated in an electrically-shielded, sound-attenuated booth. At the beginning of each trial, the first sentence of a passage appeared, in full, on an LCD monitor. Participants read this first sentence carefully for comprehension at their own pace, and it remained on the screen until participants pressed a button indicating they were ready to proceed. After this button press, a fixation cross appeared centrally for 1000 ms. The second sentence of the passage was then presented one word at a time using rapid serial visual presentation (RSVP). Each word appeared for 300 ms with a stimulus onset asynchrony of 600 ms.

In the first half of the experiment, participants were instructed to read each passage carefully for comprehension. Approximately one-quarter of the stimuli were followed by a true-false comprehension question. In the second half of the experiment, we introduced the Prediction task. Participants were instructed to read both sentences carefully and to try to predict the final word of each passage before it appeared. Approximately 1700 ms following the offset of the final word, participants were instructed to respond whether their prediction was correct or incorrect. Participants were encouraged to be as honest as possible, and once their response was recorded, the experiment proceeded automatically to the next trial. Participants saw six blocks of sentences in total, and short breaks were provided when needed.

EEG recording

The electroencephalogram (EEG) was recorded from 29 tin electrodes mounted in an elastic cap (Electro-Cap International; Eaton, OH). Horizontal and vertical electrooculograms were also recorded to monitor eye movements and blinks. All electrode impedances were kept below 5 kΩ. The EEG signal was amplified using a Synamps Model 8050 Amplifier (Compumedics Neuroscan) with a bandpass of 0.05–100 Hz. The signal was digitally recorded at a sampling rate of 250 Hz. All channels were initially referenced to an electrode placed over the right mastoid and later re-referenced to the average of the right and left mastoids.

After EEG recording, independent component analysis (ICA) was used to isolate and remove EEG artifacts due to blinks. Trials with remaining EEG artifact due to amplifier drift, muscle artifacts, or eye movements were rejected prior to analysis (2.1% of all trials). Finally, event-related potentials (ERPs) were averaged across items within each condition (Medium-cloze vs Low-cloze) and task block (Comprehension vs Prediction), including a 200 ms baseline and 1000 ms of activity following the onset of the final critical word. Except where indicated, statistical analyses were performed on mean amplitude ERP amplitudes across conditions within a set of pre-defined latency windows. Latency analyses were performed using *peak latency* – the time when an effect reaches its maximum – as well as a *20% peak latency* – the time when an effect reaches 20% of its maximum amplitude (Luck, 2014). For all analyses with more than 1 degree of freedom, any reported *p*-values were first adjusted using the Greenhouse-Geisser correction for nonsphericity.

Results

Behavioral results

Participants spent more time reading the first sentence of discourse passages in the Comprehension block (mean = 5.6 s, *S.D.* = 1.6 s) than in the subsequent Prediction block (mean = 4.8 s, *S.D.* = 1.5 s), $t(23) = 3.01$, $p = .006$. In the Comprehension block, participants correctly answered 92% (*S.D.* = 7%) of the true-false comprehension questions. In the Prediction block, participants reported a correct prediction following 49.9% (*S.D.* = 13.3%) of the medium-cloze passages and 4.2% (*S.D.* = 6.2%) of the low-cloze passages. Consistent with previous findings, participants' prediction accuracies closely matched the offline cloze probabilities of these materials (50.7% and 0.9% respectively). These results suggest that participants were attending carefully to the discourse passages in both tasks.

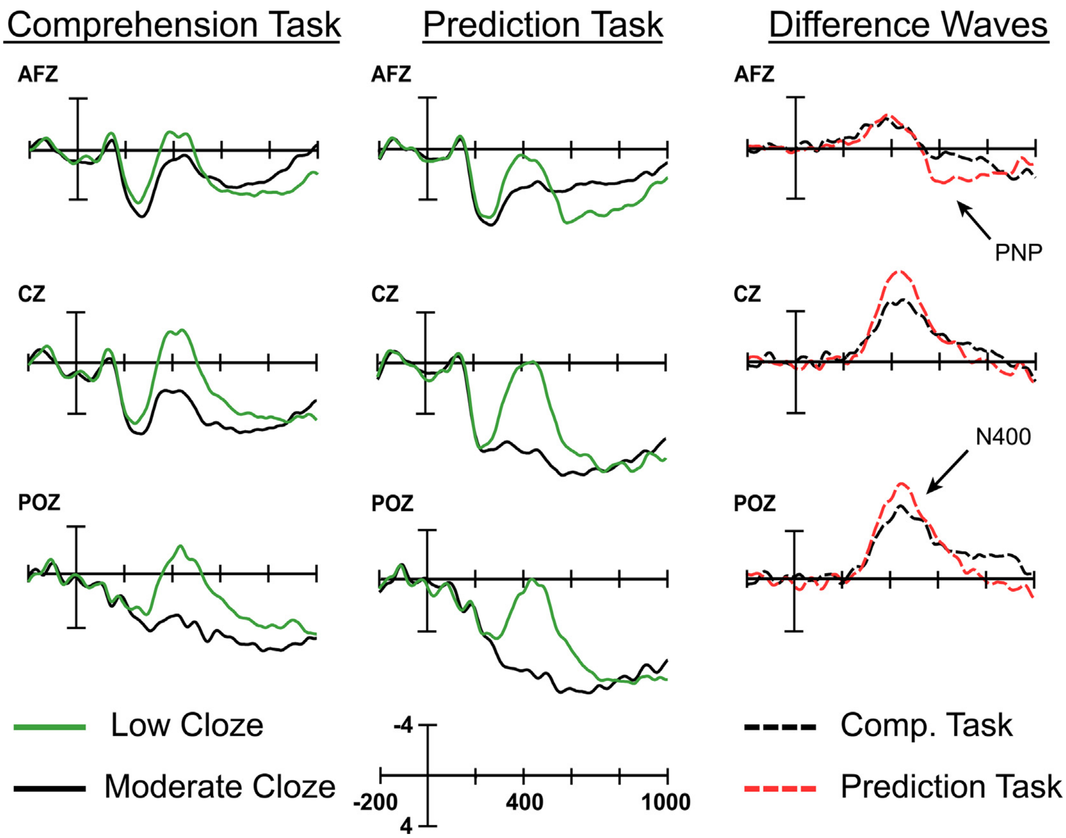

Fig. 1. Grand-average event-related potentials, time-locked to the final word of each passage. Waveforms are averaged separately for Moderate cloze (50%) and Low Cloze (0%) completions within the Comprehension and the Prediction Tasks. Predictability difference waves (*Low cloze* minus *Moderate cloze*) are also plotted for the two tasks. Note the enhancement of the N400 and the PNP effect when participants are asked to predict.

ERP results

Fig. 1 shows grand-average event-related potentials, averaged separately as a function of task and cloze probability. Following initial visual components (P1, N1) the ERPs in both tasks showed a prominent N400 peak for low-cloze sentence final words and reduced N400 amplitudes for medium-cloze targets. In both tasks, these N400 cloze differences began at approximately 200 ms, with the largest differences occurring at approximately 400 ms over central-posterior electrodes. The size of these cloze effects also appeared to differ across tasks, with greater reductions in N400 amplitude for medium-cloze words in the Prediction task when compared to the Comprehension task. Beyond the N400 time window, there also appeared to be differences in the amplitude of the post-N400 positivity. Across both tasks, the PNP was enhanced for unpredictable target words, but again these differences were more pronounced during the Prediction task. Finally, we observed an overall Task effect on sentence-final words with ERPs in the Prediction task showing a general positive shift across the scalp.[1]

To assess these differences statistically, we performed a series of 2 × 2 (Task × Cloze) repeated-measures ANOVAs comparing mean ERP amplitudes across conditions. For each time-window of interest, we performed two separate ANOVAs to assess the topographic distribution of any ERP effects: one including Midline electrode sites (AFz, Fz, Cz, Pz, POz) with a five-level factor of Anteriority, and one over Lateral electrode sites which included a three-level factor of Anteriority: Frontal (FP1/2, F7/8, F3/4), Central (FC5/6, FC1/2, C3/4, CP1/2, CP5/6) and Posterior (T5/6, P3/4, O1/2), and a two-level factor of Hemisphere (Left, Right). Our analyses will focus first on ERP differences between the Comprehension and Prediction tasks, and then turn later to ERP differences between correct and incorrect predictions within the prediction block.

N400 time window. In the N400 time-window (300–500 ms), we observed a main effect of Task (Midline: $F(1,23) = 37.27$, $p < .001$, Lateral: $F(1,23) = 30.43$, $p < .001$), a main effect of Cloze probability (Midline: $F(1,23) = 184.23$, $p < .001$, Lateral: $F(1,23) = 146.29$, $p < .001$) and a significant Task by Cloze interaction (Midline: $F(1,23) = 11.71$, $p = .002$, Lateral: $F(1,23) = 8.26$, $p = .009$). Although the Cloze effect was significant for each of the tasks when analyzed separately (all $Fs > 100$), the amplitude of this effect was significantly larger over central-parietal electrode sites (Cz, Pz, CP1/2, P3/4) in the Prediction task (5.8 µV) than the Comprehension task (4.2 µV). Interest-

[1] For additional whole-head ERP plots and topographic distributions, see Supplementary Materials.

ingly, these differences in amplitude were not accompanied by overall differences in latency, with no significant differences in the timing of the N400 differences wave across tasks: *peak latency* (Comprehension: 457 ms vs Prediction: 444 ms, $p = .17$), *20% peak latency* (Comprehension: 328 ms vs Prediction: 333 ms, $p = .60$).

Post-N400 time window. In the PNP time-window (600–900 ms) we observed a similar pattern of results. Repeated–measures ANOVAs revealed a main effect of Task (Midline: $F(1,23) = 33.68$, $p < .001$, Lateral: $F(1,23) = 28.44$, $p < .001$), an interaction between Cloze probability and Anteriority (Midline: $F(4,92) = 28.15$, $p < .001$, Lateral: $F(1,23) = 10.92$, $p = .002$), and a Task by Cloze interaction (Midline: $F(1,23) = 4.62$, $p = .04$, Lateral: $F(1,23) = 3.64$, $p = .07$). In both tasks, unexpected sentence-final words produced a larger positivity over frontal electrode sites (AFz, Fz, FP1/2, F3/4, all $Fs > 6$), but again this difference was more pronounced in the Prediction task (1.65 μV) than the Comprehension task (0.96 μV).

Effects of prediction accuracy. In addition to assessing the differential effects of cloze probability across tasks, we also analyzed ERPs within the Prediction block to isolate the neural effects of accurate and inaccurate lexical predictions. For this analysis, we separately averaged ERPs to medium-cloze words based on each participant's self-reported prediction accuracy. In this way, we could compare brain responses to predicted medium-cloze, unpredicted medium-cloze, and low-cloze trials. Replicating previous results (Brothers et al., 2015), we observed reduced N400 amplitudes for successfully predicted sentence-final words beginning approximately 200 ms post stimulus-onset (see Fig. 2). Mean ERP amplitudes for *predicted* medium-cloze and *unpredicted* medium-cloze words differed in both an early 200–300 ms time window (Midline: $F(1,23) = 10.48$, $p = .004$, Lateral: $F(1,23) = 5.93$, $p = .02$), as well as the typical N400 time-window between 300 and 500 ms (Midline: $F(1,23) = 23.0$, $p < .001$, Lateral: $F(1,23) = 23.4$, $p < .001$), particularly over right-posterior electrode sites (Cloze × Hemisphere: $F(2,46) = 4.7$, $p = .04$, Cloze × Anteriority: $F(2,46) = 29.7$, $p < .001$).

Even after controlling for differences in prediction accuracy, N400 amplitudes also differed as a function of prior contextual support, with unpredicted *low*-cloze words showing larger N400 amplitudes than unpredicted *medium-cloze* words. While these differences were not present in the early 200–300 ms window (all $Fs < 1$), they were quite robust in the later time-window between 300 and 500 ms (Midline: $F(1,23) = 75.4$, $p < .001$, Lateral: $F(1,23) = 59.4$, $p < .001$), again particularly over right-posterior electrodes (Cloze × Hemisphere: $F(2,46) = 7.8$, $p = .01$, Cloze × Anteriority: $F(2,46) = 12.7$, $p < .001$). Latency analyses performed over a cluster of central-posterior electrodes revealed that the ERP effects of contextual support were delayed by approximately 100 ms relative to the effects of prediction accuracy: *peak latency* (Prediction effect: 378 ms vs Context effect: 478 ms, $t(23) = 8.41$, $p < .001$), *20% peak latency* (Prediction effect: 298 ms vs Context effect: 387 ms, $t(23) = 8.78$, $p < .001$).

The amplitude of the frontal PNP (600–900 ms) was also modulated as a function of prediction accuracy. Similar to the N400, the amplitude of the PNP showed a graded pattern over frontal electrode sites, with the smallest positivity for predicted final words (1.5 μV), a moderate positivity for unpredicted final words (2.4 μV), and the largest positivity for low-cloze final words (3.8 μV). While the amplitude difference between unpredicted and low-cloze words was highly reliable (Condition × Anteriority, Midline: $F(4,96) = 10.5$, $p < .001$), the difference between predicted and unpredicted words was only significant over a restricted set of frontal electrodes (FP1/2, AFz, Fz: $F(1,23) = 4.33$, $p = .048$). In sum, these results successfully replicate previous findings; correct lexical predictions reduced the amplitude of both the N400 and the frontal PNP. In addition, prediction accuracy had a rapid influence during word recognition, preceding the effects of contextual support by approximately 100 ms. This is consistent with the idea that lexical prediction also results in form-based pre-activation, which can provide facilitation during very early stages of word recognition (Lau et al., 2013; Luka & Van Petten, 2014).

Discussion

In Experiment 1, participants read two-sentence discourse passages under two different sets of task instructions. During the Comprehension task, participants were asked to read each passage carefully and occasionally answer true-false comprehension questions. In the Prediction task, participants were asked to actively anticipate the final word of each passage and, after a delay, report whether their prediction was correct. By separately averaging event-related potential activity to moderately-predictable and unpredictable passage-final words in these two tasks, we assessed whether comprehension strategies could modulate the neural effects of sentence constraint during lexical processing.

Consistent with a strategic view of anticipatory processing, the neural effects of cloze probability *did* differ across tasks. We observed larger contextual modulations of the N400 when active prediction was emphasized, suggesting that readers experienced greater facilitation after encountering a predictable lexical item. We also observed enhancements in a frontal PNP during the Prediction task, suggesting that readers may have also encountered greater costs when their lexical predictions were disconfirmed. If the content of the preceding discourse were the only relevant factor in generating upcoming lexical predictions we should have observed no differences across tasks, because all passages were fully counter-balanced across the Prediction and Comprehension blocks. The fact that anticipatory strategies can immediately influence the online comprehension of predictable and unpredictable words in context suggests that the generation of lexical predictions is at least partially subject to top-down control.

In this experiment we also observed a main effect of Task, with more positive ERP amplitudes within the Prediction task for both moderate-cloze and low-cloze sentence-final words. This ERP difference – which was maximal over central-parietal electrode sites – likely represents an

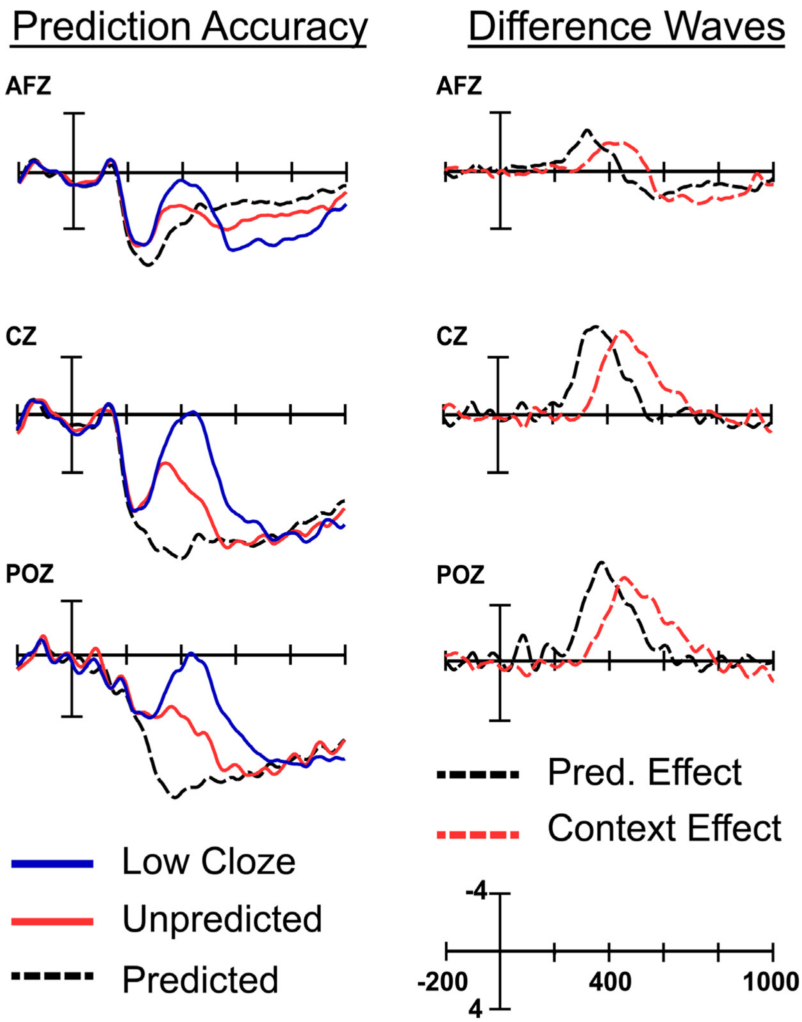

Fig. 2. Grand-average event-related potentials in the Prediction Task, averaged separately as a function of participants' self-reported prediction accuracy. Difference waves were calculated to isolate the effects of Prediction accuracy (*Medium-cloze Unpredicted* minus *Medium-cloze Predicted*) and Contextual support (*Low-cloze Unpredicted* minus *Medium-cloze Unpredicted).* Note the early onset and peak for the Prediction effect relative to the Context effect.

enhanced P300 (Donchin, 1981). This ERP difference was most likely triggered by the heightened task-relevance of sentence-final words in preparing for the upcoming prediction judgement. Critically though, a monophasic P300 difference cannot explain the biphasic enhancement of both the N400 and PNP effects across the two tasks. For additional discussion, see Brothers et al. (2015).

Some readers may wonder why the standard N400 cloze probability effect (*low-cloze* minus *medium-cloze*) did not differ in latency across tasks (Fig. 1) but that, within the Prediction block, large latency differences were observed between the effects of prediction accuracy and contextual support (Fig. 2). It stands to reason that, although we were unable to sort trials in the Comprehension task as a function of prediction accuracy, readers were likely experiencing a mix of correct and incorrect lexical predictions for these passages as well. One possible explanation for these latency effects is that correct lexical predictions produced facilitation in the 200–300 ms window across *both* tasks (Prediction and Comprehension), but the unique effects of prediction accuracy can only be uncovered by sorting trials based on participants' behavioral responses.[2] Therefore, it appears that the addition of a Prediction task does *not* influence the overall latency of the N400. Instead, activating and selecting a specific lexical prediction appears to uniquely drive ERP differences in the 200–300 ms time-window. Across both tasks, we observed similar neural consequences of contextual constraint with a similar time-course, but the addition of a Prediction task appeared to enhance the overall benefits (and costs) of predictive processing.

One potential limitation of the present results is that participants always performed the Comprehension task

[2] A 2 × 2 (Task × Cloze) ANOVA in the 200–300 ms time window supports this observation. In this time window, we observed a main effect of Cloze Probability (Midline: $F(1,23) = 8.21$, $p = .009$, Lateral: $F(1,23) = 4.09$, $p = .055$) with no significant differences across tasks.

in Block 1, prior to the Prediction task. While this arrangement was critical to ensure that participants were not implicitly performing the Prediction task in the second half of the experiment, it also raises the possibility that the ERP differences observed across tasks could be attributed to order effects. To address this concern, we ran a control experiment with the same set of materials ($N = 18$) in which participants received Comprehension instructions in both the first and second half of the experiment. While, in the main experiment, we found an enhancement in the size of the cloze probability effect when comparing the Comprehension and Prediction blocks (N400: 4.2 μV vs 5.8 μV – PNP: 0.96 μV vs 1.65 μV), there were no significant differences in the control experiment when comparing data from the first and second half (N400: 3.42 μV vs 3.05 μV – PNP: 1.1 μV vs 0.8 μV). Detailed results from this control experiment can be found in the Supplementary Materials.

As a whole, these data suggest that predictive processing has a unique influence in the early stages of lexical processing and that lexical prediction mechanisms *can* be influenced by top-down comprehension goals. Nonetheless, there are some potential methodological critiques that could be leveled against these findings. In this experiment, the second sentence of each passage was presented one word at the time, at a relatively slow, fixed rate. Considering that processing speed is a relatively important factor for anticipatory processing (Stanovich & West, 1979; Traxler & Foss, 2000; Wlotko & Federmeier, 2015) this presentation rate may have allowed more time for strategic mechanisms to influence comprehension. In addition, our manipulation to enhance predictive processing was relatively coarse, relying on explicit task instructions. In a follow-up study we addressed these limitations by using a different dependent measure (self-paced reading) as well as a more implicit strategic manipulation. In this study we wished to determine whether readers could naturally modulate their degree of anticipatory processing as a function of the statistical regularities of the reading environment.

Experiment 2

In Experiment 2, participants read sentences for comprehension at their own pace. Some of these sentences contained either a predictable or unpredictable critical word, which should produce differences in overall reading time (Ehrlich & Rayner, 1981; Smith & Levy, 2013). While these experimental sentences were held constant, a set of non-critical, filler sentences was manipulated across participants to alter the overall *validity* of predictive cues in the experiment as a whole. If lexical predictions are generated automatically as a function of the preceding sentence context, then all groups should show an identical effect of predictability on reading times within the set of critical sentences. In contrast, if (1) participants are sensitive to the overall validity of predictive cues in their environment, and (2) if they are able to strategically modulate their degree of anticipatory processing (Lupyan & Clark, 2015; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008), then the size of the predictability effect should scale proportionally with the overall likelihood of encountering a correct prediction.

Method

Participants

Two hundred and fifty-two UC Davis undergraduates (158 females) participated in Experiment 2 for course credit. The mean age of this group was 20.8 years (range 18–33, std = 2.5), and all were native English speakers.

Materials

For Experiment 2 we selected a set of sixty critical words (length = 4.9 characters, frequency = 52 per million). Each word was included in two different sentence frames, which made the critical word either highly predictable "*The web had been spun by the large <u>spider</u> on the porch*." (cloze = 96.9%, std = 3.6%), or unpredictable "*Alex said he wanted to watch the large <u>spider</u> on the porch*." (cloze = 0.4%, std = 1.1%). The position of the critical word was held constant across the two sentence frames, and the sentences were matched in overall word length. In addition, an average of two words *prior* to the critical word (range = 1–5) and all of the words *following* the critical word were identical across the two sentences (see Appendix B for examples).

To manipulate the overall strategic validity of predictive information, we also developed a set of 180 highly constraining filler sentences. These sentences could be completed with either an expected final word "*The volleyball shot barely made it over the <u>net</u>*." (cloze = 95.2%, std = 4.7%) or with an unexpected final word "*The volleyball shot barely made it over the <u>car</u>*." (cloze = 0.5%, std = 1.5%). These two alternate completions did not significantly differ in word length (high-cloze mean = 4.9, low-cloze mean = 5.1, $p > .1$) or word frequency (high-cloze mean = 75 per million, low-cloze mean = 121 per million, $p > .2$). As in Experiment 1, cloze values were obtained for all critical and filler sentences using an offline sentence-completion task. This task was performed using a new group of UC Davis undergraduates ($N = 60$ per sentence frame).

The sixty critical sentence pairs were counterbalanced to ensure that each participant saw each critical word only once – with 30 critical words appearing in a predictive context and 30 words appearing in a non-predictive context. The filler items were also manipulated across lists to create three separate groups: a High Validity group with 100% predictable fillers, a Medium Validity group with 50% predictable and 50% unpredictable fillers, and a Low Validity group with 100% unpredictable fillers. For the Medium Validity group, two additional lists were constructed to ensure that each critical item was preceded equally often by a predictable or unpredictable filler sentence. This counterbalancing scheme resulted in a total of eight experimental lists. Participants were randomly assigned to one of these lists at the beginning of their experimental session, with 63 participants in the High Validity group, 126 participants in the Medium Validity group, and 63 participants in the Low Validity group. After combining the 60 critical and 180 filler sentences, each Validity condition contained the following proportions of predictable sentence continuations – High Validity: 87.5%, Medium Validity: 50.0%, Low Validity: 12.5%.

Procedure

The experiment was run in an HTML web browser using Ibex Farm (http://spellout.net/ibexfarm). Participants were instructed to read each sentence carefully for comprehension using a self-paced moving window paradigm (Just, Carpenter, & Woolley, 1982). On each trial, a sentence appeared on the screen with all non-space characters replaced by a dash. Participants pressed the spacebar to view one word of the sentence at a time, and reading time durations were recorded as the time between button presses. A true-false comprehension question was presented in full on the screen following 25% of the sentences, and participants received immediate feed-back on their responses. Across the three Validity groups (High, Medium, Low), the set of sixty critical sentences was kept constant, and only the final words of the filler sentences were manipulated. Each subject saw ten filler sentences at the beginning of the experiment before encountering any critical items. The critical and filler sentences were then presented in a fixed random order that was constant across all eight lists.

Results

Reading time data was analyzed from the set of 60 critical sentences (see Table 1). Before performing any statistical analyses, reading times for the critical word and the subsequent spillover word ($n + 1$) were combined into a single scoring region (see Smith & Levy, 2013). All reading times in this region more than 3 standard deviations from the subject/condition mean were replaced with this cutoff value, affecting less than 1% of the data. Reading times were then analyzed using mixed-effect models with random slopes and intercepts for subjects and items. Following Barr, Levy, Scheepers, and Tily (2013), we first specified a maximal random effects model, and then simplified this model when appropriate using a backwards "best path" algorithm. Predictability was coded as a categorical variable and prediction Validity was treated as a continuous, between-subjects predictor. Both of these factors were mean-centered prior to analysis. All reported p-values were obtained via model comparison using log-likelihood ratio tests.

Reading time data

Accuracy on the comprehension questions was uniformly high (mean = 97%, std = 3%) with no significant differences across groups, $F < 1$. Our analysis of reading times in the critical region revealed a main effect of Predictability, with faster reading times when words appeared in a predictable context ($b = 16$ ms, $t = 3.8$, $p < .0001$). We saw no overall differences in reading time across Validity groups ($b = 32$ ms, $t = 0.5$, $p = .62$), but we did observe a significant Validity by Predictability interaction ($b = -28.5$, $t = 2.6$, $p = .009$). Separate analyses within each group suggested that Predictability effects were most pronounced in the High Validity group (b = 26 ms, $t = 2.72$, $p = .007$), and that they decreased in size for the Medium Validity group ($b = 18$ ms, $t = 3.74$, $p = .0003$) and the Low Validity group ($b = 4$ ms, $t = 0.97$, $p = .33$) in a roughly linear fashion (see Fig. 3).

To understand this result, it was important to determine whether these group differences were driven by a reduction in the *benefits* of a predictive context (i.e. reduced pre-activation), or a reduction in *costs* for encountering an unexpected continuation (i.e. reduced reanalysis). To answer this question, we first partialled out variability in reading speed across participants by calculating length-adjusted residual reading times (Ferreira & Clifton, 1986).[3] Two mixed-effects models were then estimated (separately for the predictable and unpredictable conditions) using length-adjusted, residual reading times as the dependent variable. As can be seen in Fig. 4, reading times in the unpredictable condition remained essentially flat across the three validity groups ($b = -5$ ms, $t = -0.37$, $p = .72$), while reading times for predictable material became faster as the validity of predictive cues increased ($b = -33$ ms, $t = -2.3$, $p = .022$). This suggests that differences in the Predictability effect across groups were driven by a reduction in contextual facilitation as predictive cues became less valid.

Another important question was whether these group differences were caused by a global shift in reading strategies, or whether they were driven by temporary adaptations to a disconfirmed prediction on the immediately preceding trial. To determine whether *local* or *global* adaptation effects could better account for these effects, we analyzed data from the Medium Validity group ($N = 126$) as a function of Predictability and the Validity of the previous filler sentence (Valid vs Invalid). This model again showed a significant main effect of Predictability ($b = 18$ ms, $t = 3.7$, $p < .001$), with faster reading times for predictable words, but no effects of the previous trial type ($b = -1$ ms, $t < 1$) and no significant interaction ($b = 4$ ms, $t < 1$). The size of the Predictability effect was essentially unchanged when the previous trial was Valid (16.3 ms) or Invalid (19.9 ms). This suggests that participants' reading behavior was influenced by the global statistical regularities of the environment, but not by a recently encountered prediction success or failure.

Finally, although the three Validity groups did not differ significantly in either comprehension accuracy or overall reading speed (all pairwise ts < 1.2), we wanted to verify that the validity manipulation was having a *specific* influence on predictive mechanisms, as opposed to a more general effect on motivation or attention. Along with

[3] For this analysis, regression slopes and intercepts were calculated for each subject, predicting reading times within the set of filler sentences as a function of word length. These slopes and intercepts were then used to calculate a predicted reading time estimate for each subject in each critical and spillover region. Length-adjusted residuals were calculated as the difference between the predicted and observed reading times on each trial. This process was necessary before performing our across-groups, supplementary analysis in order to partial-out participant-to-participant variability in reading speed. For additional information on this procedure see Ferreira and Clifton (1986).

Table 1
Raw Reading Times in the self-paced reading task, for the pre-target word (*large*), the critical target word (*spider*), and the adjacent spillover word (*on*). Average reading times are presented separately for Predictable and Unpredictable sentence frames. Finally, average per-word reading times and standard errors for the filler sentences are also presented for each group (excluding sentence-final words).

Region	High validity (87.5%)			Medium validity (50%)			Low validity (12.5%)		
	Pred	Unpred	*diff*	Pred	Unpred	*diff*	Pred	Unpred	*diff*
Pre-target ($n-1$)	292	296	*3*	294	297	*3*	281	280	*−1*
Target word (n)	295	306	*11*	296	301	*5*	288	289	*0*
Spillover ($n+1$)	298	313	*15*	301	314	*13*	292	297	*4*
Average RT (std)	322 (16)			321 (11)			307 (13)		

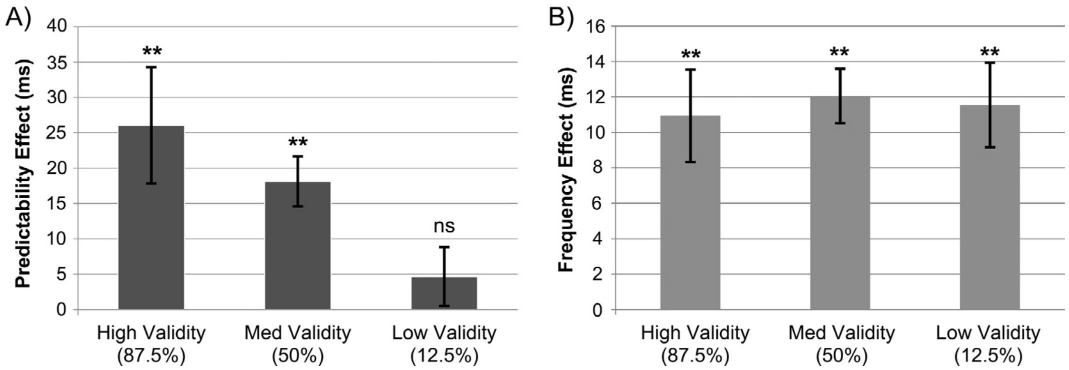

Fig. 3. (A) Differences in reading time between the Predictable and Unpredictable condition across the three Validity groups. The size of the predictability effect decreased monotonically as the validity of predictive cues decreased. (B) In contrast, predictive validity had no influence on the reading time differences between high-frequency and low-frequency words. Error bars represent ±1 standard error of the mean. **$p < .01$, ns = not significant.

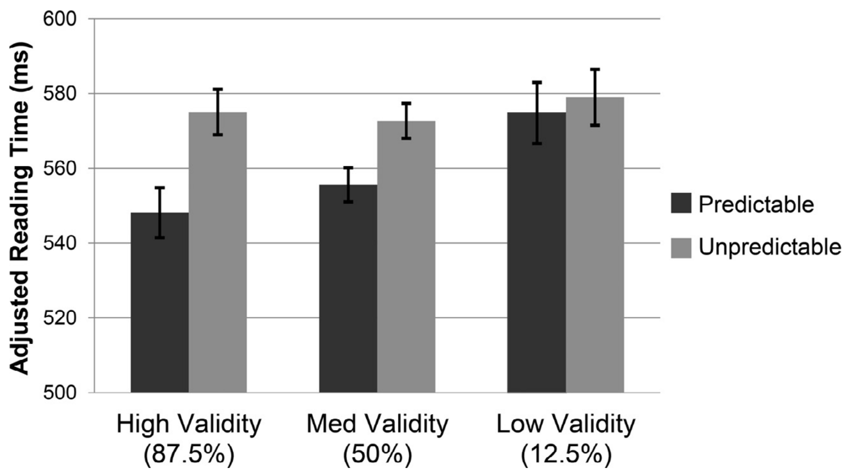

Fig. 4. Average length-adjusted reading times in the two-word critical region as a function of Predictability and the Validity of predictive cues (see text for an explanation). Reducing the Validity of predictive cues resulted in slower reading times for predictable words, but had no influence on unpredictable words. Note: a fixed value of 600 ms was added to each adjusted mean to improve the interpretability of the graph. Error bars represent ±1 standard error of the mean.

contextual predictability, word frequency also has also been shown to influence reading times (Inhoff & Rayner, 1986; Just & Carpenter, 1980). To assess whether these frequency effects would also be influenced by the Validity manipulation, we separated all non-sentence-final, content words into two frequency bins based on a median split (high-frequency: 1545 per million, low-frequency: 19 per million, 630 words per condition). We then calculated the difference in reading time between high frequency and low-frequency content words for each participant across the three Validity groups. As can be seen in Fig. 3b, the overall frequency effect was highly significant ($t(251) = 10.00$, $p < .001$) with no differences in the size of this effect across groups, $F < 1$, (High Validity: 11 ms, Med. Validity: 12 ms, Low Validity: 12 ms). The same pattern of results was observed using a mixed-effects analysis with log-frequency as a continuous predictor (frequency: $b = -8.7$ ms, $t = -3.86$, $p < .001$, frequency × Validity: $b = -0.2$, $t = -0.06$, $p = .66$). These analyses suggest that all three groups were similarly attentive and motivated while reading for comprehension and that the Validity manipulation had a selective influence on participant's global anticipatory strategies.[4]

Discussion

In Experiment 2, participants read sentences for comprehension at their own pace, while the probability of encountering a confirmed or disconfirmed lexical prediction was manipulated across participants. The reading time data showed that readers were indeed sensitive to these probabilistic cues. When the experimental context contained a large proportion of sentences with highly expected endings, participants showed robust predictability effects, with faster reading times on critical words that could be anticipated based on the preceding context. In contrast, in an environment where disconfirmed predictions were the norm – and predictive strategies were no longer valid – these reading time benefits disappeared. Moreover, this global context effect appeared to have a selective influence on the benefits of predictive constraint. As predictive validity increased, this led to greater facilitation for contextually predictable words without influencing the processing of unpredictable material.

As in Experiment 1, care was taken to equate the critical sentences across conditions. Once again, the lexical content of the predictable and unpredictable critical items was kept constant across groups and only the higher-level, strategic validity of prediction was altered. Lexical predictability effects did not appear to be generated automatically through spreading activation or lexical co-occurrence; instead the benefits provided by a supportive sentence context appeared to be modulated by top-down comprehension strategies.

The pattern of validity effects observed in the current study appears to diverge somewhat from "relatedness proportion" effects observed in word-pair semantic priming tasks. As discussed earlier, congruent semantic primes provide robust facilitation during both lexical decision and naming tasks, and this priming effect increases with the overall proportion of related prime-target pairs in the experiment (Keefe & Neely, 1990; Neely & Keefe, 1989). Critically though, semantic priming effects in these tasks typically remain significantly greater than zero even as the proportion of related pairs approaches 0%, (den Heyer et al., 1983; Tweedy, Lapinski, & Schvaneveldt, 1977). This *significant intercept* phenomenon has been cited to support multi-route models of semantic priming (Neely, 1977; Neely & Keefe, 1989). In these models, priming that is caused by conscious, attention-demanding mechanisms can be influenced by strategic effects such as proportion manipulations, while priming caused by spreading-activation mechanisms occurs automatically – largely outside of conscious control. According to the logic of these studies, the higher this intercept, the more dependent a particular priming effect is on automatic association mechanisms.

Recall that, in the current study, no significant intercept was observed. Participants in the Low Validity group showed a non-significant 4 ms effect of contextual predictability. Assuming valid prediction probabilities of approximately 87.5, 50%, and 12.5% for the High, Medium and Low Validity groups – this results in a negligible predictability intercept of 2 ms as the strategic validity of predictive cues approaches zero. This result suggests that most of the benefits of lexical predictability in this reading task could be attributed to strategic or controlled factors, with little influence from automatic mechanisms such as spreading activation.

While perhaps surprising, this result is in line with much of the previous literature on automatic associative priming during sentence comprehension. Previous studies have shown that a locally consistent prime word can result in facilitation – on both reading time measures (Camblin, Gordon, & Swaab, 2007; Traxler, Foss, Seely, Kaup, & Morris, 2000;) and N400 amplitudes (Boudewyn, Gordon, Long, Polse, & Swaab, 2012) for an upcoming target (e.g. "He carefully arranged the (tables/papers) and **chairs**."). Critically though, these association effects are somewhat weak and short-lived, often disappearing across clause boundaries (Carroll & Slowiaczek, 1986) or when two or more words intervene between the prime and target (Federmeier, Van Petten, Schwartz, & Kutas, 2003; Van Petten, Weckerly, McIsaac, & Kutas, 1997). In the present experiment, predictive semantic information was often placed earlier in the sentence, with an average of two words prior to the critical word remaining constant across the high-cloze and low-cloze sentence frames. Because of this, any benefits of lexical predictability may have required readers to actively maintain an anticipated word or concept as the sentence unfolded. This maintenance component of lexical prediction may, in turn, be particularly influenced by strategic control. Future studies that directly manipulate the point in time when predictive information is first available (early in a discourse or immediately preceding the critical target) could help to address this question (see Fitzsimmons & Drieghe, 2013).

In general, the results of Experiment 2 are consistent with previous evidence for syntactic and phonological adaptation during language comprehension (Bradlow & Bent, 2008; Fine, Jaeger, Farmer, & Qian, 2013; Kaschak & Glenberg, 2004; Norris, McQueen, & Cutler, 2003). For example, when readers encounter an unfamiliar grammatical structure such as the *needs construction* (e.g. *This lantern needs lit*), they initially show large increases in reading time. After only a few exposures, processing diffi-

[4] Although the differences in overall reading speed across groups were no greater than would be expected by chance, we wished to verify that these small differences could not account for the observed interaction between Cloze and Validity. To this end, we re-ran our mixed-effects analyses after excluding 10 Low Validity participants, in order to more closely match all three groups on reading speed (new per-word reading times – High Validity: 322 ms, Med. Validity: 321 ms, Low Validity: 322 ms). The exclusion of these participants did not affect the pattern of predictability effects across groups (High Validity: 26 ms, Med. Validity: 18 ms, Low Validity: 5 ms, Cloze × Validity: $b = -28$ ms, $t = -2.36$, $p = .019$).

culty rapidly decreases, as readers gain cumulative experience with this sentence type (Kaschak & Glenberg, 2004). It has been suggested that syntactic adaptation effects of this kind are driven by an implicit learning mechanism with the primary goal of minimizing the degree of *prediction error* encountered during language processing (Chang, Dell, & Bock, 2006; Fine et al., 2013).

While implicit learning mechanisms likely played a role in generating the present "prediction validity" effect, it is unlikely that these reading time differences were caused by the implicit updating of readers' *linguistic representations*, either at the semantic or syntactic level. In this study, encountering a disconfirmed prediction based on one set of stored semantic relationships (web-spin-spider) resulted in reduced lexical anticipation in later sentences which shared very different semantic links (return-book-library). Therefore, we believe the present adaptation effects likely altered, not the *contents* of stored linguistic knowledge, but rather the strategic *application* of this knowledge within a particular comprehension environment.

Beyond the realm of language comprehension, similar strategic anticipation effects have also been observed during simple visual and auditory processing tasks. It is well-established – from both single-cell recording and functional magnetic resonance imaging (fMRI) data – that repeated presentation of the same stimulus leads to a reduction in neural responses over time (Henson & Rugg, 2003). In human fMRI studies it has been shown that this *repetition suppression* effect is much larger when stimulus repeats can be predicted in a top-down manner based on the statistical regularities of the current environment (Summerfield et al., 2008). Based on these findings, some researchers have argued that repetition priming is mainly driven by top-down sensory expectations, which can influence neural responses in very early perceptual areas such as primary visual cortex (Larsson & Smith, 2012). In combination with the current results, these findings suggest that *flexible* anticipatory strategies may be a more general feature of human perception and cognition.

General discussion

Across two experiments, we have shown evidence that readers' anticipatory language processing (the degree to which they anticipate upcoming lexical items from context) can be influenced by global, top-down factors. In Experiment 1, explicit task instructions were able to modulate the amplitude of two neural signatures of lexical prediction, the N400 and the frontal PNP, even while the semantic constraints of the preceding context were held constant. In Experiment 2, the benefits of word predictability during self-paced reading were eliminated in an environment where these predictive cues were no longer valid. These results provide important constraints for any theoretical model of predictive language processing; in particular they provide strong evidence against bottom-up, stimulus-driven accounts of anticipatory pre-activation.

Two essential questions for any model of anticipatory language processing are: (1) *when* do readers generate lexical predictions (at what time and under which circumstances), and (2) *what* sources of information are combined in order to pre-activate some words rather than others. One basic source of information that readers could use to generate predictions is lexical co-occurrence. Because some words or phrases are more likely to occur together in a sequence ("grant permission" vs "grant protection"), readers may be able to use these transitional probabilities to anticipate upcoming words in a sentence (Levy, 2008; Smith & Levy, 2013). While there was some initial evidence that bigram statistics or "forward transitional probabilities" might play a role in sentence comprehension (McDonald & Shillcock, 2003), later work suggested that these differences could be attributed to uncontrolled differences in cloze probability (Frisson, Rayner, & Pickering, 2005; Ong & Kliegl, 2008; Smith & Levy, 2011). In other words, rather than tracking bigram frequencies, it is more likely that readers calculate probabilities based on much larger portions of text (see Staub, 2015 for a recent discussion).[5] Intra-lexical spreading activation may also provide a source of information for generating lexical predictions. While associated words clearly receive facilitation within simple semantic priming tasks (doctor – nurse), these effects are less pronounced during comprehension and can be overruled by information from the global discourse or the message-level representation of a text (Hess, Foss, & Carroll, 1995; Morris, 1994; van Berkum, Hagoort, & Brown, 1999).

As we have seen from the present experiments, word co-occurrence and automatic lexical association alone cannot explain the full pattern of prediction effects. Even when the content of a local discourse context is held constant, the amount of facilitation provided by a constraining context can change as a result of top-down comprehension goals or the recent success or failure of predictive processing. For example, when prediction is beneficial or necessary to perform a particular task (for example in the face of degraded, bottom-up input) readers may form stronger or more specific predictions for individual words. In contrast, if a reader encounters a large number of prediction errors, they may begin to shift away from an anticipatory processing strategy and instead put a stronger emphasis on bottom-up stimulus evaluation (see Lupyan & Clark, 2015 for a similar suggestion).

One mechanism that could account for these effects is a strategic, forward-modelling approach to prediction (Pickering & Garrod, 2007, 2013). According to this account, the language production system can be recruited during comprehension to covertly imitate incoming lan-

5 While we believe that lexical co-occurrence, on its own, cannot provide a plausible psycholinguistic model for how lexical predictions are generated, co-occurrence measures still have their uses. For cases where acquiring cloze norms would be difficult or prohibitively expensive (for example when analyzing large corpora of natural text) conditional probability *can* provide a quick and dirty approximation of lexical predictability. Still, considering the modest correlation between these two measures (Ong & Kliegl, 2008; Smith & Levy, 2011; $r = .5$), these measures should be interpreted with caution, particularly when other confounding lexical factors, such as word frequency, have not been carefully controlled.

guage input. By mirroring the output of a speaker in real time, the comprehender can also internally generate anticipated lexical items before encountering them in the speech stream. One desirable property of this system is that production-based, forward-modelling is an *optional* element during comprehension. In the current experiments, anticipatory input from the production system could be dialed up or down as a function of a reader's strategic goals. In this way, different anticipatory strategies could be applied quickly and flexibly without altering the internal structure of the comprehension system itself.

Recently, Huettig (2015) has proposed a possible synthesis of these different approaches, suggesting that linguistic predictions are generated using *multiple* processing pathways. In his model, rapid, automatic (Type I) systems such as spreading activation and lexical co-occurrence, can be combined with slow, strategic (Type II) systems to provide lexical pre-activation. This model assumes that these two processing routes are not encapsulated, but can interact in real time to dynamically modulate activation for multiple word candidates.

The present results highlight the importance of Type II mechanisms during sentence processing. In particular, the results of Experiment 2 suggest that strategic mechanisms are largely responsible for the reduction in reading times occurring for predictable words in context. We do recommend some caution in interpreting these results because self-paced reading and serial visual presentation are only rough approximations of normal reading comprehension. Clearly, it will be important to replicate these findings using additional paradigms and dependent measures. Moreover, it is possible that *neural* indices of lexical pre-activation, such as N400 amplitudes, may show a differential sensitivity to automatic and strategic priming effects – a question which is currently under investigation in our lab. In future work, we also hope to investigate whether pre-activation occurring at distinct representational levels (semantics, phonology, syntax) may be differentially sensitive to top-down strategies. For example, because phonological information is pre-activated more slowly (Ito, Corley, Pickering, Martin, & Nieuwland, 2016), these types of predictions may also be more sensitive to strategic control. Clearly, additional experimental evidence will be needed to address these questions.

Finally, the results of the present study also provide indirect evidence for potential costs incurred following an unsuccessful lexical prediction. If generating incorrect lexical predictions were cost-free, then there would be little motivation to modulate predictive strategies in response to the surrounding language environment. With a cost-free system, readers should continue to generate strong lexical predictions even if these anticipated words are rarely encountered (in these circumstances, *some* lexical prediction should always be better than none). The fact that readers reduce their level of anticipatory processing in situations of low predictive validity suggests that lexical prediction *does* entail some cost and that, at times, the cost-to-benefit ratio of anticipatory processing can become unfavorable. It is still unknown at what stage(s) these costs are most pronounced (generation, maintenance, disconfirmation/revision), but if the costs of lexical prediction play a role in guiding reading behavior, then an important goal for future research should be to better identify how and when these costs operate.

Acknowledgments

We thank Gina Kuperberg, Fernanda Ferreira, Megan Boudewyn, and Liv Hoversten for helpful discussions regarding this work. This research was partially funded by NSF (1024003) and NIH (R21 11601946).

Appendix A

A sample of medium-cloze (Med) and low-cloze (Low) passages used in Experiment 1. Critical sentence-final words are underlined here for emphasis. Participants saw only one passage from each pair.

Med:	Thomas didn't like the temperature of his drink. He thought it was much too hot.
Low:	Thomas didn't like the look of the water. He thought it was much too hot.
Med:	The author is writing another chapter about the fictional detective. To date, he thinks it will be his most popular novel.
Low:	Everyone congratulated the chef on all his hard work. To date, he thinks it will be his most popular novel.
Med:	The old school teacher wanted to draw the diagram up on the board. Unfortunately, he could not find his chalk.
Low:	The lawyer's pencil broke, and he couldn't finish writing his idea. Unfortunately, he could not find his chalk.
Med:	Jim thought he heard someone call his name. He looked back through the open door.
Low:	The cross country runner loved exploring new places. He ran ahead through the open door.
Med:	To test them, Steve smelled the clothes in the laundry basket. He realized that they were all dirty.
Low:	Steve wondered why the room was so dark, so he checked the blinds. He realized that they were all dirty.

Appendix B

A sample of high-cloze, low-cloze, and filler sentences used in Experiment 2. Critical words are underlined here for emphasis. Participants saw only one sentence from each pair. For filler sentences, the predictable and unpredictable completions are separated by a slash.

High: The web had been spun by the large spider on the porch.
Low: Alex said he wanted to watch the large spider on the porch.

High: At the vineyard, he bought a fancy bottle of wine for Debra.
Low: While wandering around, he found a new type of wine for Debra.

High: There was only one case that Sherlock Holmes was unable to solve in London.
Low: There was only one thing that my uncle was unable to solve in London.

High: The birthday boy blew out all the candles with Donna's help.
Low: The woman went to buy all the candles with Donna's help.

High: Because it was overdue, Billy returned to book to the library that was downtown.
Low: After finishing up at the movies, they went to the library that was downtown.

Filler: The volleyball shot barely made it over the net/car.
Filler: The patient was prepped for surgery at the local hospital/clinic.
Filler: The security team thought the suspicious briefcase posed a serious threat/concern.
Filler: I had to go inside because the jackhammer was so loud/close.
Filler: He only had seven dollars and twelve cents/dimes.

C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2016.10.002.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Altmann, G., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*, 583–609.

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.

Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., & Swaab, T. Y. (2012). Does discourse congruence influence spoken language comprehension before lexical association? Evidence from event-related potentials. *Language and Cognitive Processes, 27*, 698–733.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*, 707–729.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition, 136*, 135–149.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*, 103–128.

Carroll, P., & Slowiaczek, M. L. (1986). Constraints on semantic priming in reading: A fixation time analysis. *Memory & Cognition, 14*, 509–522.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181–204.

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology, 48*, 1203–1207.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.

den Heyer, K., Briand, K., & Dannenbring, G. L. (1983). Strategic factors in a lexical-decision task: Evidence for automatic and attention-driven processes. *Memory & Cognition, 11*(4), 374–381.

Donchin, E. (1981). Surprise!... surprise? *Psychophysiology, 18*, 493–513.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641–655.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging, 18*(4), 858–872.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75–84.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE, 8*, e77661.

Fitzsimmons, G., & Drieghe, D. (2013). How fast can predictability influence word skipping during reading? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1054–1063.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 862–877.

Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia, 41*, 263–270.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124*, 62–82.

Heyman, T., Van Rensbergen, B., Storms, G., Hutchison, K. A., & De Deyne, S. (2015). The influence of working memory load on semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 911–920.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*(1), 66–85.

Howes, C., Healey, P. G., Purver, M., & Eshghi, A. (2012). Finishing each other's... Responding to incomplete contributions in dialogue. In *Proceedings of the 34th annual conference of the cognitive science society* (pp. 479–485).

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118–135.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 645–662.

Hutchison, K. A., Heap, S. J., Neely, J. H., & Thomas, M. A. (2014). Attentional control and asymmetric associative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 844–856.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*, 431–439.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157–171.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.
Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General, 133*, 450–467.
Keefe, D. E., & Neely, J. H. (1990). Semantic priming in the pronunciation task: The role of prospective prime-generated expectancies. *Memory & Cognition, 18*, 289–298.
Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*, 32–59.
Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. *Predictions in the Brain: Using Our Past to Generate a Future*, 190–207.
Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.
Larsson, J., & Smith, A. T. (2012). FMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cerebral Cortex, 22*, 567–576.
Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*, 484–502.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177.
Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, MA: MIT Press.
Luka, B. J., & Van Petten, C. (2014). Prospective and retrospective semantic processing: Prediction, time, and relationship strength in event-related potentials. *Brain and Language, 135*, 115–129.
Lupyan, G., & Clark, A. (2015). Words and the world predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science, 24*, 279–284.
McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*, 648–652.
Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.
Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 92–103.
Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226–254.
Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective/retrospective processing theory. *The Psychology of Learning and Motivation, 24*, 207–248.
Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. *Basic Processes in Reading: Visual Word Recognition, 11*, 264–336.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.
Ong, J. K. Y., & Kliegl, R. (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research, 2*, 1–7.
Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110.
Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*, 329–347.
Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 514–528.
Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1637–1642).
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*(3), 302–319.
Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition, 7*, 77–85.
Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9*, 311–327.
Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11*, 1004–1006.
Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. (2012). Language-related ERP components. In S. J. Luck & E. Kappenman (Eds.), *Oxford handbook of event-related potential components*. .
Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly, 30*, 415–433.
Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*, 382–392.
Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1266–1282.
Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence processing: Intralexical spreading activation, schemas, and situation models. *Journal of Psycholinguistic Research, 29*, 581–595.
Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition, 5*, 84–89.
van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*, 657–671.
Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology, 94*, 407–419.
Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*, 176–190.
Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science, 8*, 238–242.
Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex, 68*, 20–32.