Brain and Language 199 (2019) 104697

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Short communication

fMRI evidence that left posterior temporal cortex contributes to N400 effects of predictability independent of congruity

Ellen F. Lau[a,*], Anna Namyst[a,b]

[a] University of Maryland, Department of Linguistics, College Park, MD, United States
[b] NIMH MEG Core Facility, National Institutes of Health, Bethesda, MD, United States

ARTICLE INFO

Keywords:
N400
Prediction
Pre-activation
Congruity
MTG
ERP
fMRI

ABSTRACT

Previous electrophysiological work argues that predictability and semantic incongruity rapidly impact comprehension, as indicated by modulation of the N400 component between ~300 and 500 ms. An ongoing question is whether effects of predictability in fact reflect pre-activation in long-term memory as opposed to modulating the kind of integration processes triggered by incongruity. Using fMRI, we compared the impact of predictability and incongruity in adjective-noun phrases, in regions identified with lexical and phrasal localizer scans. We found that predictability impacted activity in left posterior middle temporal gyrus (pMTG), while incongruity impacted activity in left precentral gyrus. Together with parallel data from ERP, these data are consistent with the hypothesis that left pMTG activity is a key contributor to N400 effects of predictability and that the relevant mechanism is reduced activation of stored lexical representations. We tentatively suggest that the left precentral region may play a role in reanalysis when incongruity is encountered.

1. Introduction

Much psycholinguistic research suggests that during language comprehension readers and listeners routinely predict the content, and perhaps the form, of the upcoming input, and that predicted input is correspondingly processed more easily (Federmeier & Kutas, 1999; Altmann & Kamide, 1999; Wicha, Moreno, & Kutas, 2004; and many others). However, from the early days of this work, a central challenge for studying prediction in language has been dissociating effects of prediction from other processes. In particular, predictability manipulations were often confounded with congruity, such that in one condition the critical word was both highly predictable and semantically congruous, and in the other condition the critical word was both unpredictable and semantically incongruous (e.g. *He spread the warm bread with butter/socks*; Kutas & Hillyard, 1980). If neural activity is greater in the latter condition than the former, the difference could be attributed either to facilitated lexical or conceptual access, which would reduce activity in the predictable condition, or to increased effort towards computing the sentence- or discourse-level meaning, which would increase activity in the incongruous condition.

This ambiguity has resulted in ongoing debate about the functional interpretation of the well-known N400 effect in ERP, and in uncertainty about the extent to which the N400 effect can be used to study predictive processes in language comprehension. The N400 response refers to a broad negative deflection in the ERP response to words and other meaningful stimuli which peaks around 400 ms post-stimulus onset. Kutas and Hillyard (1984) and hundreds of studies since have observed that the amplitude of the N400 is smaller for words that are predicted by the context. This could be taken to indicate that the N400 indexes activation of lexical and conceptual networks, which can be activated less broadly when a predictive context allows the comprehender to narrow in on the correct candidate (Federmeier & Kutas, 1999). However, much of this data could also be taken to indicate that the N400 only indexes the integration of incoming input into the sentence- or discourse-level meaning, with unpredictable sentences describing scenarios that are less common/plausible and thus require more effort to construct the meaning (Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

Recent ERP work from our group and others provides new evidence that N400 amplitude is modulated by predictive facilitation. Deconfounding predictability and semantic congruity in sentence contexts has been challenging because even large corpora are too sparse to provide precise estimates of the likelihood of particular sentence continuations, and human sentence continuation data provide precise estimates for highly predictable endings but less precise estimates for unpredictable endings. Lau, Namyst, Fogel, and Delgado (2016) introduced a two-word adjective-noun paradigm, which allows better-supported estimates of predictability from corpus data. We compared

* Corresponding author.
E-mail addresses: ellenlau@umd.edu (E.F. Lau), anna.namyst@nih.gov (A. Namyst).

https://doi.org/10.1016/j.bandl.2019.104697
Received 25 February 2018; Received in revised form 17 September 2019; Accepted 18 September 2019

the separate effects of predictability (*runny nose* vs. *dainty nose*) and semantic incongruity (*yellow bag* vs. *innocent bag*). When predictability was controlled, we found that congruity elicited only small, barely reliable N400 differences, while predictability drove a very large N400 reduction. These results suggest that predictability can modulate the N400 response independent of integration difficulty, by hypothesis through facilitated lexical and/or conceptual activation.[1]

Here we used the adjective-noun paradigm from Lau et al. (2016) to provide converging evidence from fMRI for the hypothesis that N400 effects of contextual predictability indeed reflect pre-activation of stored lexical and/or conceptual representations. We first used a localizer run to identify regions that were engaged in lexical and sentential processing across participants. Then, just as in our ERP study, participants were presented with two-word adjective-noun phrases and attended to the materials in anticipation of a post-experiment memory recall test. If N400 effects of predictability reflect mechanisms that support basic access of stored lexical representations, we would expect to see effects of predictability in posterior temporal areas that respond to manipulations of lexical content. Alternatively, if N400 effects of predictability reflect mechanisms that integrate stored lexical or conceptual representations, we would expect to see effects of predictability in frontal or anterior temporal regions that respond to manipulations of sentential structure and meaning.

Which brain regions generate N400 context effects has also been a matter of some debate. As reviewed by Lau, Phillips, and Poeppel (2008; see also Van Petten & Luka, 2006), fMRI effects in mid-posterior temporal cortex appear to best track semantic priming manipulations that generate N400 effects, and MEG work using sentence contexts is also consistent with this (although see e.g. Maess, Herrmann, Hahne, Nakamura, & Friederici, 2006; Hagoort, 2008 for dissenting perspectives). However, the results of fMRI studies using congruity/predictability manipulations in sentences have been much more variable. This likely reflects the fact that the temporally 'sluggish' fMRI response sums the response to most or all words in the sentence, resulting in increased variability that can mask the differential response to the single word that is manipulated. An additional virtue of the adjective-noun paradigm used here is that only the single context word and the target word contribute to the response.

2. Results

Overall accuracy on the post-run memory recall task was 69% for the predictability manipulation and 63% for the congruity manipulation. These scores were somewhat low, and reflect the fact that the task of recalling which adjectives and nouns had been presented together after viewing a list of ~100 phrases is a fairly challenging one.

2.1. Lexical/phrasal localizer

The results of the lexical/phrasal localizer were largely consistent with previous work. In the lexical contrast (scrambled sentences vs. consonant strings) we observed significant clusters of increased activity for scrambled sentences relative to consonant strings in left posterior MTG, left posterior STG, and left precentral/middle frontal gyrus. In the phrasal contrast (sentences vs. scrambled sentences) we observed a significant cluster of increased activity for sentences in left anterior STS (see table in Supplementary Materials for exact coordinates and extents). We evaluated effects of lexical predictability and semantic incongruity manipulations in these four ROIs.

We note that previous work using this kind of localizer has identified additional regions in similar lexical and phrasal contrasts, and indeed at a less conservative threshold we also observed phrasal effects in right hippocampus, right posterior STG, left IFG, and right anterior STS, and lexical effects in left anterior STG and right posterior STG, among other regions.

2.2. Predictability effects

The predictability manipulation had a reliable impact only in the left posterior MTG region of interest identified in the lexical localizer, such that highly predictable items demonstrated less activity than less predictable items (Fig. 1). In the omnibus ANOVA we observed a significant interaction between condition and region ($F(1, 23) = 2.7$, $p < .05$), and in follow-up ANOVAs in each of the four regions, pMTG was the only region to demonstrate a significant effect of condition ($F(1, 23) = 3.1$, $p = .05$). Follow-up pairwise comparisons among the three conditions in pMTG demonstrated a significant difference between high predictability and low predictability, low constraint conditions ($F(1, 23) = 5.1$, $p < .05$) and between high predictability and low predictability, high constraint conditions ($F(1, 23) = 4.4$, $p < .05$). In our previous ERP study using the same materials and paradigm (Lau et al., 2016) we observed the same pattern in N400 amplitudes to the critical noun, as illustrated in Fig. 3.

2.3. Congruity effects

The semantic congruity manipulation had a reliable impact only in the left precentral/middle frontal region of interest identified in the lexical localizer, such that incongruous items demonstrated more activity than congruous items (Fig. 2). In the omnibus ANOVA we observed a significant interaction between condition and region ($F(1, 23) = 5.3$, $p < .05$), and in follow-up ANOVAs in each of the four regions, the precentral gyrus ROI was the only one to demonstrate a significant effect of condition ($F(1, 23) = 5.1$, $p < .05$).

2.4. Exploratory whole-brain analysis

No voxels showed significant effects of predictability or congruity in the exploratory whole-brain analysis.

3. Discussion

The current study aimed to bring fMRI evidence to bear on the hypothesis that N400 effects of contextual predictability reflect lexical or conceptual pre-activation, as opposed to being solely driven by differences in post-lexical integration difficulty. In the short two-word materials used here, a manipulation of predictability robustly modulates N400 amplitude in ERP and a manipulation of semantic congruity does not (Lau et al., 2016). The corresponding fMRI results show that activity in a region of left posterior MTG picked out by a 'lexical' localizer tracks the N400 predictability pattern, and at the same time does not differentiate semantically congruous and incongruous items. Activity in the 'phrasal' region of left anterior STS was modulated by neither predictability nor congruity. However, we found that activity in a 'lexical' region of left precentral/middle frontal gyrus was increased in response to incongruity.

3.1. Predictability effects in posterior middle temporal cortex

Like our prior ERP study (Lau et al., 2016), these data argue against a view in which N400 predictability effects can simply be reduced to semantic integration difficulty. If semantic integration does occur when processing such short phrases, then a large difference in integration difficulty would be observed for incongruous phrases such as *innocent bag*. However, Lau et al. (2016) observed very little difference in the ERP response to incongruous phrases when compared to unpredictable congruous phrases, and in the current study we do not observe reliable

[1] These results leave open the possibility that integration difficulty separately impacts the ERP during this time-window, e.g. DeLong, Quante, & Kutas, 2014; Brothers, Swaab, & Traxler, 2015.

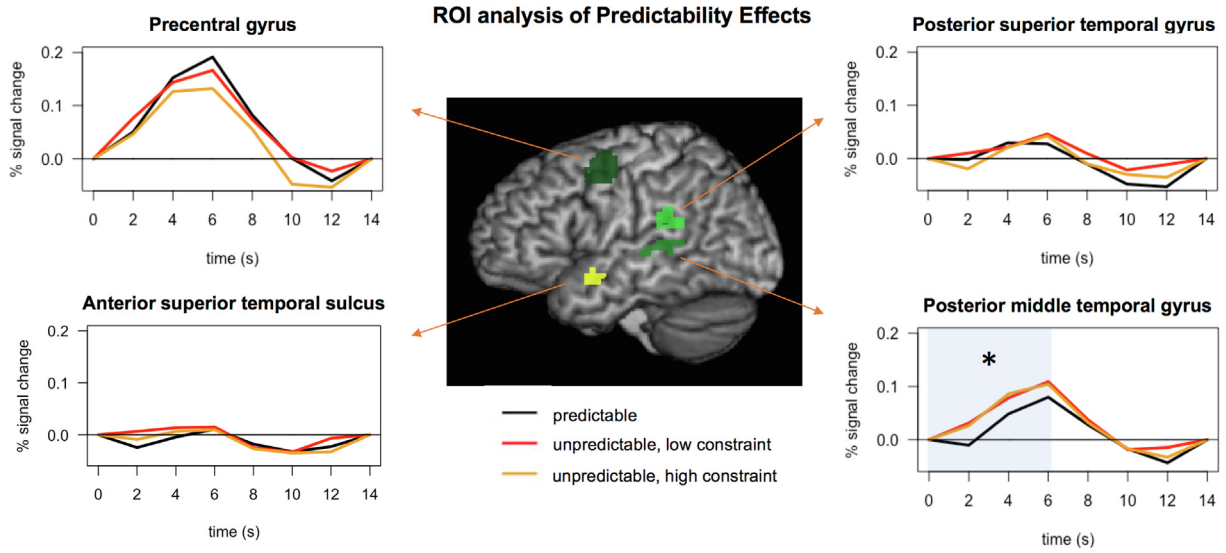

Fig. 1. Estimated hemodynamic timecourses for the predictability manipulation in regions of interest identified by the lexical and phrasal localizers.

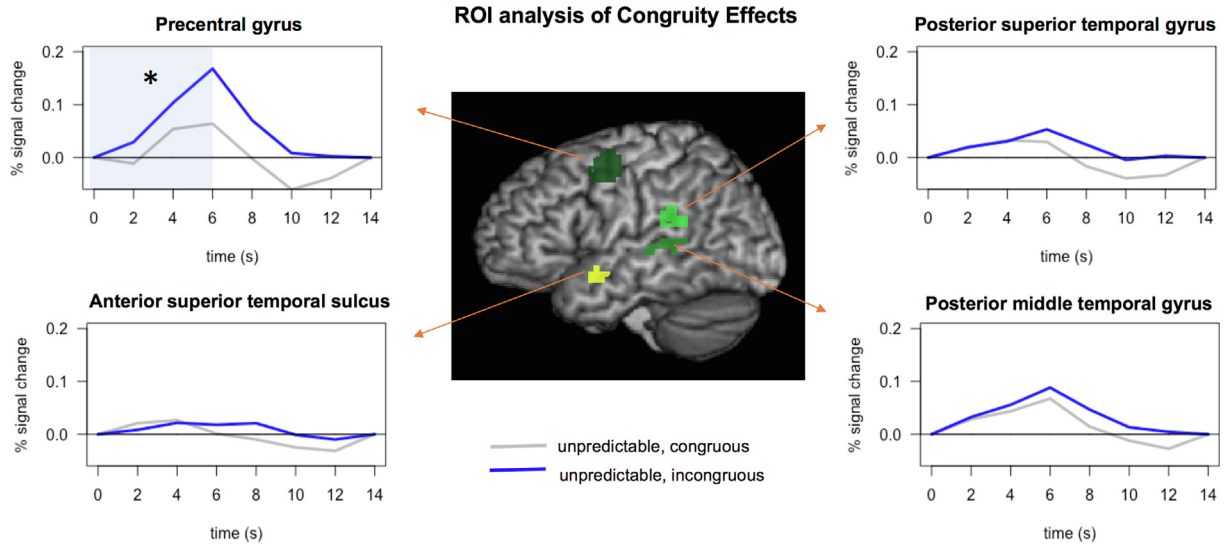

Fig. 2. Estimated hemodynamic timecourses for the congruity manipulation in regions of interest identified by the lexical and phrasal localizers.

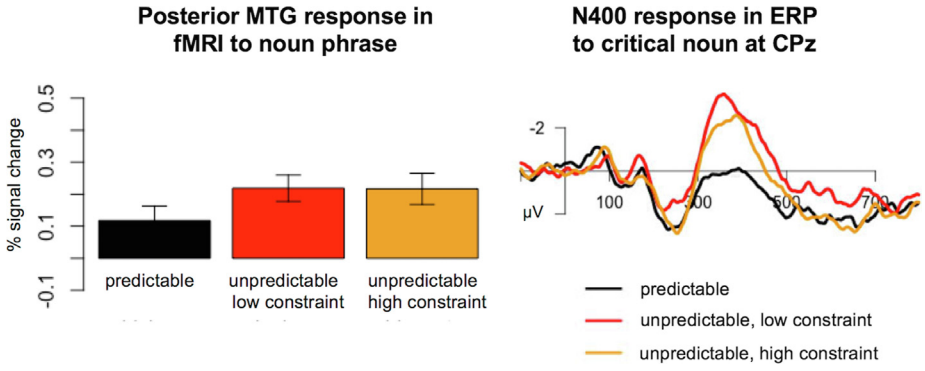

Fig. 3. Side-by-side comparison of the effect of the predictability manipulation on the estimated hemodynamic response (across 0–6 s) in the posterior MTG ROI in the current study and in the ERP study reported by Lau et al. (2016). Error bars represent standard errors.

effects of congruity in the pMTG region in which we observed effects of predictability. These data are consistent with the view that N400 effects of predictability reflect the activation levels of stored lexical and conceptual representations in temporal cortex (Lau et al., 2008), where more predictable items elicit reduced neural responses to the bottom-up input.

Over the years, many have suggested that parts of left pMTG and/or pSTS contain either the units that encode wordforms, or more abstract 'lexical' or 'lemma' units that map directly to conceptual representations (e.g. Dronkers, Wilkins, Van Valin Jr., Redfern, & Jaeger, 2004;

Hickok & Poeppel, 2007; Wilson, Bautista, & McCarron, 2018). Conversely, generalization across conceptual representations has frequently been argued to be supported by a 'semantic hub' of units in anterior temporal cortex (e.g. Patterson, Nestor, & Rogers, 2007; Binder & Desai, 2011). The current fMRI results showing lexical predictability effects in left pMTG are consistent with this picture, as predictability could facilitate processing at the wordform and/or lexical level[2]. The current results are also consistent with several recent fMRI studies using a continuous approach which reported effects of bigram probability (Brennan, Stabler, Van Wagenen, Luh, & Hale, 2016) and lexical surprisal (Willems, Frank, Nijhof, Hagoort, & Van den Bosch, 2015) in left posterior temporal cortex, among other regions.

As reviewed in the Introduction, previous work has frequently localized N400 effects to temporal cortex, but the position of those effects has varied from anterior to posterior temporal cortex. In a review of most extant semantic priming fMRI studies at that time (Lau et al., 2008) we found that the predominant location of effects was in posterior MTG. However, in subsequent studies using semantic priming in a within-subjects EEG-MEG/fMRI design, we found that N400 effects localized instead to a more anterior position in STG/STS (Lau, Gramfort et al., 2013; Lau, Weber, Gramfort, Hämäläinen, & Kuperberg, 2014). We have suggested that the explanation lay in the emphasis of the experiments and tasks. The majority of the early semantic priming studies used lexical decision tasks, where emphasis is on identification of a specific lexical item (e.g. Gold et al., 2006). On the other hand, the experiments in Lau, Gramfort et al. (2013) and Lau et al. (2014) used a conceptual probe task (*Press the button when you see an animal word*), where emphasis is on identification of a conceptual category (see Lau et al., 2014, for further discussion). In the current study we observed reliable effects of predictability in our posterior temporal ROI, but found no reliable effects of predictability in our anterior temporal ROI.[3] Although speculative, we believe that the posterior locus of the current effect can be reconciled with the task emphasis explanation. Here participants performed a memory probe task at the end of the block requiring them to distinguish complete adjective-noun phrases they had seen from permuted catch trials. This might have resulted in greater attention to the particular words that were observed on each trial, and corresponding facilitation in a posterior temporal lexical network.

3.2. Why are hemodynamic correlates of N400 prediction effects relatively weak?

Although the ROI analysis revealed reliable effects of predictability in left posterior MTG/STS that corresponded to patterns of N400 amplitude, no effects of predictability were robust enough to survive the whole-brain analysis even at a relatively liberal threshold. This is fairly consistent with prior hemodynamic work: in our own previous fMRI semantic priming work we observed relatively small and circumscribed effects (Lau, Gramfort et al., 2013; Lau et al., 2014; Weber, Lau, Stillerman, & Kuperberg, 2016), and this is also true in many previous studies (e.g. Gold et al., 2006). In fMRI experiments that used sentence-level manipulations that typically lead to N400 effects, the results were even more variable, as reviewed in Lau et al. (2008). However, against the ERP literature this might seem surprising; compared to other ERP responses N400 predictability effects are relatively large in magnitude, long in duration (several hundred milliseconds) and can often be clearly seen in datasets from individual participants.

Why is such a strong and reliable effect in the ERP measure so much less robust in fMRI? The question has not been resolved, but there are many possibilities. One is that it results from different properties of the averaging in ERP and fMRI. Some ERP differences can arise from increased synchronization of oscillations rather than a net difference in neural activity; if this were the mechanism driving N400 effects (Roehm, Schlesewsky, Bornkessel, Frisch, & Haider, 2004), it would be unlikely to lead to a net change in hemodynamic activity (see Van Petten & Luka, 2006 for discussion, and Brázdil et al., 2005 on the P3 component). Because the N400 response is smeared across the scalp in EEG, slightly different distributions from individual to individual still sum together to give a net effect, but if the response is actually generated in somewhat different regions of temporal cortex from person to person, then spatially more precise fMRI would not end up summing these effects across individual participants (see Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010 for broader discussion of this issue). Another possibility is that even responses that last a few hundred milliseconds in ERP are too short to generate reliable differences in measurable hemodynamic activity, and that standard fMRI effects of language manipulations tend to reflect differential activity across many seconds. We believe that these are important questions for future research, as it cannot be assumed in advance that hemodynamic measures will be sensitive to the processes driving a particular ERP response; rather, for each response this assumption must be motivated with converging evidence from methods like MEG and intracranial recordings. Newer fMRI acquisition approaches with improved temporal precision such as multiband EPI sequences (Feinberg et al., 2010; Todd et al., 2016) may also be helpful in resolving hemodynamic effects that are temporally localized to particular parts of the HRF, as suggested to be the case here.

3.3. Semantic incongruity effects in dorsolateral prefrontal cortex

Although we did not predict that lexicality or semantic incongruity would modulate activity in left precentral/middle frontal gyrus, it is also not wholly surprising. Dorsolateral prefrontal cortex has been implicated in a range of high-level executive functions, from working memory to monitoring to decision-making. In the scrambled sentence vs. consonant string localizer contrast, this activity may reflect an effort to relate the unstructured words. Similarly, incongruous phrases may have triggered the engagement of executive routines, whether to examine if an analysis mistake was made, to search for an alternative word meaning, or to generate a metaphorical interpretation. Increased activity in this region for incongruous sentences has been observed in previous fMRI studies, although not consistently (e.g. Kuperberg et al., 2003). An interesting question for future investigation is whether this differential dorsolateral prefrontal activity is the source of the minor differences in N400 amplitude that we observed for the same congruity manipulation in ERP.

While isolated two-word phrases provide critical methodological advantages for determining the hemodynamic correlates to brief ERP effects, results from the comprehension of isolated two-word phrases can in no way be taken to provide a full picture of the processes involved in comprehension of sentences and discourses. Many integration processes may not be engaged by two-word phrases. Both classic fMRI work that investigated semantic incongruity in full sentences (e.g. Friederici, Rueschemeyer, Hahne, & Fiebach, 2003; Kuperberg et al., 2003; Hagoort, Hald, Bastiaansen, & Petersson, 2004) and recent fMRI work investigating lexical surprisal (an information-theoretic measure of expectedness) in more naturalistic discourses (Willems et al., 2015; Henderson, Choi, Lowder, & Ferreira, 2016; Brennan et al., 2016) has reported effects in a much broader set of regions. Similarly, our functional localizer, which involved passive reading with no explicit task, is unlikely to provide a comprehensive picture of the brain network that supports more naturalistic sentence comprehension. In particular, the

[2] A reviewer points out that the same logic could hold if the pMTG region were involved in processing individual phonemes rather than whole lexical items, assuming that phoneme processing were facilitated by prediction, more emphasized by lexical decision tasks, and more engaged by word-like stimuli than consonant strings. Although we favor a lexical role for this region, we agree that this is a logical possibility.

[3] Although it is worth noting that the anterior MTG/STS region picked out by the localizer is slightly inferior to the anterior STG region that showed priming effects in our prior work.

absence of robust IFG modulation in our functional localizer prevented us from evaluating semantic congruity effects in isolated phrases. Interestingly, an exploratory analysis using a BA45/47 cluster derived from the localizer at a higher threshold (p = .003) resulted in a marginally significant congruity effect in the later part of the hemodynamic response (8–12 s). Given the greater functional-anatomical variability associated with IFG (Amunts et al., 1999), it is likely that individual functional localizer data will be needed to better evaluate this response, which could be easily accomplished in future studies by slightly lengthening the localizer portion of the study to allow sufficient power to accurately map language regions for individual participants.

A final caveat of note is that our congruity manipulation always followed our predictability manipulation. Although a recent analysis of two-word semantic priming in ERP shows no reduction in N400 effect size across as many as 300 trials (Delaney-Busch, Morgan, Lau, & Kuperberg, 2019), and our Lau et al. 2016 ERP study showed similar results when predictability and congruity were run separately vs. being intermixed, future work should confirm that the same patterns hold when materials are randomized. More broadly, while rapid serial visual presentation must be used for reading studies in ERP to avoid pernicious eye-movement artifacts, it is important that future work aim to replicate these observations using a more natural presentation method such as the auditory modality.

3.4. Conclusion

We have presented the results of an fMRI study that replicated a recent ERP manipulation of predictability and congruity. This multimodal approach allowed us to provide novel support for the hypothesis that N400 effects of contextual predictability do in fact reflect the impact of predictive context on activation of lexical or conceptual representations, and not only later supra-lexical processes such as semantic integration. Although the current contribution is modest, we hope this type of multimodal ERP-fMRI mapping work will ultimately lay the foundation for a more integrated cognitive neuroscience of language.

4. Methods

4.1. Participants

Participants were 24 right-handed (Oldfield, 1971) native English speakers (15 female; mean age 22.6, range 19–29) who participated in the study for monetary compensation and gave informed consent in accordance with the Institutional Review Board of the University of Maryland. One additional participant was excluded from analysis for excessive movement.

4.2. Lexical/phrasal localizer

The lexical/phrasal localizer was composed of three conditions: sentences, word lists, and consonant string lists. All stimuli were nine words long. Forty sentences with varied but relatively simple structures were selected (20 from Rogalsky & Hickok, 2009) and divided across two presentation lists. Scrambled sentences for each list were created by scrambling words across all the sentence items from the other list. The 20 nonword stimuli in each list consisted of a sequence of random consonant strings, where the length of each consonant string was matched to the length of the words in the sentences. Each participant saw only one list. All stimuli are available in Supplementary Materials.

Each trial consisted of 9 words (300 ms on, 100 ms off), preceded by a fixation cross (200 ms on, 200 ms off), for a total time of 4 s per trial. 20 trials were randomly intermixed from each of 3 conditions: sentences, scrambled sentences, and consonant strings. In order to optimize deconvolution of the fMRI signal, the 60 4-s trials were intermixed with 80 s of rest (fixation) using the optseq algorithm (http://surfer.nmr.mgh.harvard.edu/optseq), and 10 s of rest was added to the end of the run for a total run-time of 330 s.

4.3. Predictability and congruity manipulations

Predictability and congruity manipulations were designed as subexperiments conducted in two separate runs, following the design of the initial two ERP experiments in Lau et al. (2016) (the current study was designed at the same time). The congruity run always followed the predictability run. Experiment 3 in Lau et al. (2016) demonstrated that the ERP effects of predictability and congruity were relatively unchanged when the materials were fully intermixed.

The predictability manipulation consisted of three conditions: predictable (*runny nose*), unpredictable, low-constraint *(dainty nose*), and unpredictable, high-constraint *(runny yogurt)*. Lau et al. (2016) collected ERPs from all three conditions in their Experiment 1, but focused their report on the contrast between the first two conditions only and its relation to effects of congruity. We included the additional high-constraint condition here in order to evaluate any BOLD responses associated with inhibiting the originally predicted noun (such inhibition is known not impact the N400 response; e.g. Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Lau et al. (2016) describes materials creation in greater detail. We derived predictability counts from the COCA corpus (Davies, 2009). For predictable items, p(noun|adjective) > 0.5, mean 0.65; for unpredictable items, p(noun|adjective) was < 0.02. For low-constraint adjectives there existed no noun for which p(noun|adjective) was > 0.15. 120 3-item sets were distributed across two lists in a Latin Square design, such that each participant saw exactly one version of each item, 40 items per condition.

The congruity manipulation consisted of two conditions: *congruous, unpredictable* (*yellow bag*) and *incongruous, unpredictable* (*innocent bag*). All items had a p(noun|adjective) < 0.005, and all adjectives were relatively unconstraining. Eighty 2-item sets were distributed across two lists in a Latin Square design. Each participant saw 40 items from each condition and exactly one version of each item, such that each participant saw exactly one version of each item, 40 items per condition.

In each trial, the adjective was presented for 500 ms, followed by a 100 ms blank screen, and then the noun was presented for 900 ms, followed by at least 100 ms of fixation, for a total time of 2 s per trial. In the predictability block, the 120 2 s trials were intermixed with 240 s of rest using the optseq algorithm to optimize deconvolution of the fMRI signal, and 10 s of rest was added to the end of the run to fully capture the responses to the last trials, for a total run time of 490 s. In the congruity block, the 80 2 s trials in were intermixed with 160 s of rest, and 10 s of rest was added to the end of the run, for a total run time of 330 s.

Participants' task was to complete a memory recall test after each run. This test consisted of 20 bigrams, of which 10 had appeared in the preceding block and 10 were mismatched adjective-noun pairs from the stimulus set. Participants made a binary button-press decision to indicate whether the pair had appeared in the previous run or not.

4.4. Recording

fMRI data were acquired at the Maryland Neuroimaging Center on a Siemens TRIO 3 T scanner (Siemens Medical Systems) using a 32-channel head coil. We first collected a T1-weighted high-resolution structural image in the sagittal plane (.9mm isotropic MPRAGE, 192 slices; TR 1.9 s; TE 2.32 ms; flip angle 9°). The subsequent functional scans were acquired with a T2*-weighted gradient echo sequence (TR, 2 s; TE, 24 ms; flip angle, 70°; interleaved acquisition). Each volume consisted of 36 axial slices (AC-PC aligned, 3 mm isotropic voxel with .3mm gap).

4.5. Analysis

AFNI software (Cox, 1996) was used to analyze the functional MRI data. After slice-time correction, functional images were motion corrected to the second time point of the predictability manipulation run using a 6-parameter rigid-body transformation (Cox & Jesmanowicz, 1999), aligned to the anatomical image, and warped to Talairach space. Functional images were spatially smoothed using a Gaussian 6 mm FWHM kernel. For the predictability and congruity manipulations, first-level General Linear Model (GLM) analyses were carried out on the functional images with a finite impulse response (FIR) model that gave estimates of the hemodynamic response at each TR (every 2 s) between 0 and 14 s using AFNI's 3dDeconvolve function with the 'tentzero' parameter, which sets activity at the first and last TRs to zero. Activity estimates were computed for each condition at each voxel in each participant, and the estimated values at 2, 4, and 6 s post-stimulus onset were summed to create an estimate for each condition and each voxel to carry forward to the second-level analysis (Ashby, 2011). Group-level analyses were conducted on these summed activity estimates using AFNI's 3dANOVA2 function.

We focused on the first three of the six time points estimated for the BOLD response because our goal was to measure the fMRI response that corresponds to the relatively rapid and short-lived N400 response in EEG. In Lau, Gramfort et al. (2013), Lau, Holcomb et al. (2013) and Lau et al. (2014), we conducted parallel N400 manipulations in EEG, MEG, and fMRI, which strongly indicated that increased BOLD response in left anterior STG corresponded to N400 effects observed in EEG. Unpublished exploratory analyses conducted during that work indicated that these effects were observed most strongly in the early part of the hemodynamic response (see Supplementary Materials for more detail). We used the canonical HRF approach for analysis of the localizer scan because these stimuli were longer in duration (4 s) and the activity of interest was not time-locked to a particular point in the stimulus.

Functional ROIs for evaluating the contrasts of interest were extracted from the lexical contrast (scrambled sentences > consonant strings) and phrasal contrast (sentences > scrambled sentences) using an uncorrected threshold of $p < .0005$ and a cluster size threshold of 20 contiguous voxels.

Type III repeated-measures ANOVAs (condition × region) were conducted separately for predictability and congruity manipulations. Significant condition × region interactions were followed up with ANOVAs evaluating the effect of condition in each region separately.

We also conducted an exploratory whole-brain analysis of the predictability and congruity manipulations. We used a first-level threshold of $p < .01$ and a Monte Carlo procedure using AFNI's AlphaSim function to determine corrected cluster-level significance with a second-level threshold of $p < .05$. We found no clusters that reached significance in this analysis.

Acknowledgements

We thank Allison Fogel, Tania Delgado, Caitlin Richter, Jennifer Stark, Wang Zhan, and William Matchin for assistance with materials creation, data collection, and data analysis. This research was supported in part by NSF grant BCS-1749407 to Ellen Lau.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2019.104697.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.
Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal of Comparative Neurology, 412*(2), 319–341.
Ashby, F. G. (2011). *Statistical analysis of fMRI data.* MIT press.
Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15*(11), 527–536.
Brázdil, M., Dobšík, M., Mikl, M., Hluštík, P., Daniel, P., Pažourková, M., ... Rektor, I. (2005). Combined event-related fMRI and intracerebral ERP study of an auditory oddball task. *Neuroimage, 26*(1), 285–293.
Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W. M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language, 157*, 81–94.
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.
Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research, 29*(3), 162–173.
Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine, 42*(6), 1014–1018.
Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*(2), 159–190.
Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. (2019). Neural evidence of rational adaptation of semantic predictions: a Bayesian model of trial-by-trial N400 amplitude during semantic priming. *Cognition*.
DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61*, 150–162.
Dronkers, N. F., Wilkins, D. P., Van Valin Jr, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*(1–2), 145–177.
Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of memory and Language, 41*(4), 469–495.
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.
Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of neurophysiology, 104*(2), 1177–1194.
Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., ... Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PloS One, 5*(12), e15710.
Friederici, A. D., Rueschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170–177.
Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience, 26*(24), 6523–6532.
Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences*.
Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438–441.
Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage, 132*, 293–300.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393.
Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272–293.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163.
Lau, E., Namyst, A., Fogel, A., & Delgado, T. (2016). A direct comparison of N400 effects of predictability and incongruity in adjective-noun combination. *Collabra: Psychology, 2*(1).
Lau, E. F., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2013a). Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *Journal of Neuroscience, 33*(43), 17174–17181.
Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013b). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*(3), 484–502.
Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920.
Lau, E. F., Weber, K., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2014). Spatiotemporal signatures of lexical–semantic prediction. *Cerebral Cortex, 26*(4), 1377–1387.
Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., & Friederici, A. D. (2006). Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Research, 1096*(1), 163–172.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.
Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976.

Roehm, D., Schlesewsky, M., Bornkessel, I., Frisch, S., & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *Neuroreport, 15*(3), 409–412.
Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex, 19*(4), 786–796.
Todd, N., Moeller, S., Auerbach, E. J., Yacoub, E., Flandin, G., & Weiskopf, N. (2016). Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: Sensitivity and slice leakage artifacts. *Neuroimage, 124*, 32–42.
Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 443–467.
Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279–293.
Weber, K., Lau, E. F., Stillerman, B., & Kuperberg, G. R. (2016). The yin and the yang of prediction: An fMRI study of semantic predictive processing. *PloS one, 11*(3), e0148637.
Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288.
Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex, 26*(6), 2506–2516.
Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage, 171*, 62–74.