



Published in final edited form as:

Lang Cogn Neurosci. 2020 ; 35(6): 797–812. doi:10.1080/23273798.2019.1630654.

Talker-specific predictions during language processing

Rachel Ryskin^{a,b}, Shukhan Ng^c, Katherine Mimnaugh^d, Sarah Brown-Schmidt^e, Kara D. Federmeier^{c,d}

^aDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA;

^bDepartment of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts, USA;

^cBeckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Champaign, IL, USA;

^dDepartment of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA;

^eDepartment of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA

Abstract

Language comprehension is shaped by world knowledge. After hearing about “a farm animal,” meanings of typical (“cow”) versus atypical exemplars (“ox”) are more accessible, as evidenced by N400 responses. Moreover, atypical exemplars elicit a larger post-N400 frontal positivity than typical and incongruous (“ivy”) exemplars, indexing the integration of unexpected information. Do listeners adapt this category knowledge to specific talkers? We first replicated typicality effects in the auditory modality. Then, we extended the design to a two-talker context: talkers alternated cueing (Bob: “Susan, name a farm animal”) and answering (Susan: “cow”). Critically, participants first heard interviews in which one talker revealed strong associations with atypical exemplars (Susan works on an ox farm). We observed increased frontal positivity to a typical exemplar (“cow”) said by Susan compared to Bob, indicating participants appreciated that the typical exemplar was atypical for Susan. These results suggest that comprehenders can tailor their expectations to the talker.

Keywords

Comprehension; semantics; category typicality; prediction; ERP

A substantial literature suggests that as language comprehension unfolds, listeners and readers can probabilistically use context to prepare for semantic, morphosyntactic, phonological, and even orthographic features of likely upcoming words (Altmann & Kamide, 1999; Federmeier, 2007; Kuperberg & Jaeger, 2016; though see Nieuwland et al.,

[✉]CONTACT Rachel Ryskin ryskin@mit.edu.

Supplemental data for this article can be accessed <https://doi.org/10.1080/23273798.2019.1630654>.

Disclosure statement

No potential conflict of interest was reported by the authors.

2018 for some limits on these effects). This expectation facilitates word processing - for example, as evidenced by a reduction in the N400 component of the event related potential (ERP), which has been linked to semantic access (Federmeier & Kutas, 1999; Federmeier & Laszlo, 2009). Such predictions are filtered through the structure of the networks with which they make contact; semantic predictions, for example, are sensitive to category structure (e.g. Federmeier & Kutas, 1999). In Federmeier, Kutas, and Schul (2010), participants read category cues (e.g. a type of tree) followed by potential exemplars of that category. Relative to unrelated words (e.g. tin), which elicited the largest N400s, category members elicited facilitated (reduced) N400s, graded by typicality (and hence predictability): N400 responses to typical exemplars (e.g. oak) were smaller than those to atypical exemplars (e.g. ash).

ERPs have further revealed that effects of context are multifaceted. Following the N400, a frontally-distributed positivity indexes the consequences associated with integrating an unexpected, prediction-violating word that is nevertheless a plausible continuation (e.g. “The groom took the bride’s hand and placed the ring on her dresser”. [“finger” is predicted]; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Similarly, in young adults (who are making predictions), Federmeier et al. (2010) observed a larger frontal positivity for the unpredictable category exemplars (e.g. ash) than the highly predictable ones (e.g. oak). Unrelated words (e.g. tin) patterned with the highly predictable words in this time window. Thus, fine-grained information about category membership guides listeners’ on-line predictions about upcoming words as well the subsequent integration of their meaning with the context.

It is noteworthy that the measure of predictability of a word in response to a category is derived from norms averaged across large numbers of people. Any individual responder may have an idiosyncratic representation of what is a more or less prototypical category exemplar. For example, while most responders say “oak” when asked to name a kind of tree, someone who grew up near a forest of ash trees may think of an “ash” first. The same is likely to be true in everyday language production: talkers are presumably more likely to name category exemplars with which they have close personal experience because those will be more readily retrieved from their mental lexicon (see Johnson & Mervis, 1997; Tanaka & Taylor, 1991). A key open question is whether listeners take such talker-specific information into account when rapidly generating online predictions during language comprehension.

Listeners attend to information about the talker at a number of levels. For example, listeners can learn the fine-grained acoustic features of a talker’s dialect (e.g. *tag* and *tack* have different vowels in certain North American dialects) and use that information to anticipate unfolding words (Dahan, Drucker, & Scarborough, 2008; Trude & Brown-Schmidt, 2012). Similarly, listeners learn associations between talkers and the words they tend to say from exposure (e.g. male talker says *penguin* and female talker says *pencil* repeatedly) and access those lexical items more readily when the talker-word pairing is consistent with recent experience (Creel, 2012, 2014; Creel & Bregman, 2011; Creel, Aslin, & Tanenhaus, 2008). At a higher level, listeners use information about a talker’s spatial perspective to disambiguate spatial terms (Ryskin, Wang, & Brown-Schmidt, 2016) and integrate social identity information - inferred from the talker’s voice/accents - into online language interpretation (Hanulíková, van Alphen, van Goch, & Weber, 2012; Van Berkum, van den

Brink, Tesink, Kos, & Hagoort, 2008). However, whether listeners can override their long-term knowledge of category structure to accommodate specific knowledge about a talker's experiences or preferences is unknown. Do listeners always predict the most likely category exemplar given language-wide and world knowledge, or do they account for talkers' idiosyncratic category knowledge?

In the present work, we first adapt the paradigm used by Federmeier et al. (2010) to study online category exemplar predictions in the auditory modality. In the critical, second experiment, we then test whether listeners' category predictions can be tuned to the specific talker. Background knowledge about one of the talkers is provided via interviews that indicate that they have some idiosyncratic experience (e.g. they worked on an ox farm) and therefore should be more likely to produce an exemplar that is not common in the general population (e.g. a farm animal: ox). On the other hand, no background information is provided about the other talker, so category predictions should be generated based on population-wide probabilities (e.g. cow). The ERP signal associated with processing the category exemplar can reveal whether or not listeners use talker-specific information. In particular, if talker-specific knowledge can impact initial semantic access, we might see effects on the N400, and if this information elicits novel predictions, then the disconfirmation of talker specific expectations might manifest on the later frontal positivity.

Experiment 1

The goal of Experiment 1 was to replicate the findings in Federmeier et al. (2010) and extend them to the auditory modality, in order to determine the appropriate time windows and electrodes to be used when addressing the primary research question in Experiment 2.

Method

Participants—Sixteen participants (12 female, age range: 18–22 years) were recruited from the undergraduate participant pool at the University of Illinois at Urbana-Champaign. They either received course credit or cash as compensation.

Materials & design—The stimuli were modelled after Federmeier et al. (2010). They consisted of 72 category cues (e.g. “A farm animal”; see osf.io/72p9m/ for the full set of stimuli) paired with a target word for each of three predictability conditions: high-predictability (HP, e.g. “cow”), low-predictability (LP, e.g. “ox”), and incongruous (IN, e.g. “ivy”). HP and LP targets were selected based on existing category norms (Battig & Montague, 1969; Hunt & Hodge, 1971; McEvoy & Nelson, 1982; Shapiro & Palermo, 1970). The 72 category cues were split into 2 lists of 36 category cues. Targets were matched for length and number of phonemes ($p > .4$) but differed in print frequency (based on the HAL written corpus of US English; Lund & Burgess, 1996). HP items, on average, had higher frequency counts than LP and IN items ($p < .004$), and the latter two target types were not reliably different ($p = .308$) (see osf.io/72p9m/ for details). For the auditory stimuli, both the category cues and targets were spoken by a female American English speaker and digitally recorded and normalised in amplitude.

The materials were divided into two lists, each of which was presented visually or auditorily. Every participant completed two blocks of 108 trials each (36 trials per predictability condition), one in each modality. The order in which these blocks were presented as well as the list of stimuli that was paired with each modality were counterbalanced across participants.

Procedures—Participants first filled out demographics and handedness questionnaires. They were then seated in a sound-proof, electrically-shielded chamber 40 in. from a CRT monitor. They were given in-ear headphones for auditory stimuli and a button box to record their overt responses, while electrodes positioned on the scalp recorded neural responses. The volume of the in-ear headphones was adjusted to a comfortable level for each participant before the experiment began. They were instructed to decide whether the target word (e.g. *oak*) that was presented following a category cue (e.g. *A kind of tree*) belonged to the category. The response button for “yes” or “no” was counterbalanced across participants. No feedback was provided.

Both blocks (auditory and visual) started with five practice items, which were identical across participants. Each trial began with the prompt “READY?” in the centre of the screen. After participants pressed a button, a fixation cross, “+”, was displayed on the screen for 500 ms, followed by a category cue.

For auditory trials, when the category cue was played, a dot appeared in the centre of the screen. Participants were instructed to look at the dot so as to limit eye-movements. Category cue sound files ranged in duration from 603 to 2134 ms. An inter-stimulus interval (ISI) of 1000 ms followed the offset of the category cue. The auditory target then started playing while a dot was presented on the screen until the participant’s response was recorded. The auditory target sound files ranged from 401 to 1061 ms in duration.

For visual trials, the category cue display duration was matched to the duration of the corresponding auditory sound file, with the constraint that all cues were displayed for a minimum of 1000 ms. After an ISI of 1000 ms, the visual target appeared for 1000 ms, followed by a blank screen until a response was recorded. Participants were asked to sit still and control their blinking while the target word was on the screen. All words appeared in black text on a white background. The target words were in an Arial 48-point font but the category cues were in an Arial 40-point font. This font size difference was needed to accommodate more characters in the cues.

EEG recording and analysis—The electroencephalogram (EEG) was recorded using a 32-electrode BrainVision system, referenced online to the left mastoid. A schematic of the electrode arrangement can be seen in Figure 1. Blinks and eye movements were monitored via electrodes placed on the outer canthus of each eye and infraorbital ridge of the left eye. Electrode impedances were kept below 5 k Ω . EEG was processed through BrainVision amplifiers set at a bandpass of 0.02–250 Hz and sampled at 1000 Hz.

Data were re-referenced off-line to the average of the left and right mastoids. Trials containing blink, lateral eye-movements, excessive muscle activity, electrode drift, or

amplifier blocking were rejected off-line before averaging. This affected an average of 7.9% of trials per participant, per modality (range 0–19.2%). ERPs were computed for epochs extending from 100 ms before target word onset to 900 ms after target onset. Averages were computed for artefact-free trials for each predictability condition, after subtraction of the 100 ms presti-mulus baseline.

Results

Behavioural

Accuracy: For visually presented stimuli, participants responded accurately on 99% of HP, 92% of LP, and 99% of IN trials. For aurally presented stimuli, participants responded accurately on 99% of HP, 91% of LP, and 98% of IN trials. All analyses were conducted using the lme4 (Bates, Mächler, Bolker, & Walker, 2014) package in R (R Core Team, 2016). Accuracy was analysed in a mixed effects logistic regression. The fixed predictor was predictability condition entered as a pair of orthogonal contrast codes: IN vs. HP & LP (Contrast 1: IN = -0.66, HP = 0.33, LP = 0.33) and HP vs. LP (Contrast 2: IN = 0, HP = 0.5, LP = -0.5). The model also included random intercepts for participant and target word as well as random slopes for condition by participant. Model estimates are summarised in Table 1 (see Supplemental Material 1 for random effects structure). Participants were significantly less accurate on LP trials than on HP trials in both modalities.

Response times: Response times shorter than 50 milliseconds and longer than 3000 milliseconds were excluded from analyses. This affected 0.87% of data points in the visual modality and 0.98% of data points in the auditory modality. For visually presented stimuli, average response times were 711 milliseconds for HP words, 852 milliseconds for LP words, and 850 milliseconds for IN words. For aurally presented stimuli, average response times were 931 milliseconds for HP words, 1086 milliseconds for LP words, and 1158 milliseconds for IN words. Log-transformed response times were analysed in a mixed effects linear regression. The fixed predictor was predictability condition entered as a pair of orthogonal contrast codes: IN vs. HP & LP (Contrast 1: IN = -0.66, LP = 0.33, HP = 0.33) and HP vs. LP (Contrast 2: IN = 0, HP = 0.5, LP = -0.5). The model also included random intercepts for participant and target word as well as random slopes for condition by participant. Model estimates are summarised in Table 1 (see Supplemental Material 1 for random effects structure).

Participants were significantly slower on IN trials relative to LP and HP trials, and slower on LP trials than HP trials in both modalities.

ERPs—Average ERPs for each predictability condition in the visual modality can be seen in Figure 1(a). Average ERPs for each predictability condition in the auditory modality can be seen in Figure 1(b). Analyses of ERP data only included trials on which participants responded accurately.¹ The data were analysed in two time windows: N400 and Frontal positivity (FP).

¹Data loss due to artefacts and incorrect responses was uniformly distributed across conditions ($M = 11\%$, $\min = 0\%$, $\max = 30\%$)

N400: For the visual stimuli, the dependent variable was the mean amplitude across trials for a subject² in a time window ranging between 300 and 500 milliseconds (as in Federmeier et al., 2010), since visual N400s typically peak just before 400 ms in college populations. For the auditory stimuli, the dependent variable was the mean amplitude between 300 and 600 milliseconds, because the auditory N400 is typically longer-lasting (Kutas & Van Petten, 1994), given that sounds unfold over time. Mean EEG amplitude from eight centro-posterior electrodes (see Figure 1), where the N400 tends to be most prominent (Kutas & Federmeier, 2011), was analysed in a linear mixed effects regression. Predictability condition was entered as a fixed effect, as a pair of orthogonal contrasts: IN vs. LP and HP (Contrast 1: IN = -0.66, LP = 0.33, HP = 0.33) and LP vs. HP (Contrast 2: IN = 0, LP = -0.5, HP = 0.5). Participant and electrode were entered as random intercepts. Condition was also allowed to vary by participant and by electrode. Model estimates for both modalities are summarised in Table 2. *P*-values were obtained using the Satterthwaite approximation to degrees of freedom in the lmerTest R package (Kuznetsova, Brockhoff, & Christensen, 2017). We observed significant effects of predictability condition in both modalities: IN (visual Mean [M] = -0.50 μ V, standard deviation [SD] = 4.76; auditory *M* = -5.61 μ V, SD = 3.28) elicited a significantly larger negativity than HP (visual *M* = 4.44 μ V, SD = 5.06; auditory *M* = -0.62 μ V, SD = 4.71) and LP (visual *M* = 3.33 μ V, SD = 4.71; auditory *M* = -2.62 μ V, SD = 2.70) whereas LP elicited more negative-going EEG than HP but only marginally in the visual modality³ ($BF_{01} = 12.47$ in favour of the null hypothesis [from model in Supplemental Material 3 using the brms package, Bürkner, 2018]; note the Bayes Factor calculation is highly dependent on the prior distribution: in this case, a normal distribution centred on 0 with a SD of 10).

Post-N400 frontal positivity: The frontal positivity was analysed in the time window immediately subsequent to the N400 in both modalities, continuing to the end of the epoch. Thus, for the visually presented stimuli, the analysis time window was between 500 and 900 milliseconds (as in Federmeier et al., 2010) and for the aurally presented stimuli it was between 600 and 900 milliseconds (because the N400 window ended at 600 milliseconds). For the visually presented stimuli, EEG amplitudes from five prefrontal electrodes (MiPf, LLPf, RLPf, LMPf, RMPf; see Figure 1) were analysed, as in Federmeier et al. (2010). For auditory stimuli, because the presence of an auditorily-elicited frontal positivity has not yet been characterised in the literature, we conducted an analysis on the full set of frontal electrodes (MiPf, LLPf, RLPf, LMPf, RMPf, LDFr, RDFr, LMFr, RMFr, LLFr, RLFr), as well as on the smaller subset of prefrontal electrodes where the frontal positivity is typically maximal in the visual modality.

For all analyses, mean amplitudes over trials⁴ were analysed in a linear mixed effects regression with predictability condition as a fixed effect: LP vs. HP & IN (Contrast 1: LP = -0.66, HP = 0.33, IN = 0.33) and HP vs. IN (Contrast 2: LP = 0, HP = 0.5, IN = -0.5). Participant and electrode were entered as random intercepts and condition was allowed to

².See Supplemental Material 3 for analyses with trial-level data. Note that we observe the same patterns in both analyses.

³.The effect size was estimated to be approximately 2 μ V in previous work (Federmeier et al., 2010)

⁴.See Supplemental Material 3 for analyses with trial-level data. We found evidence for the same contrasts in these analyses, unless noted otherwise in the main text.

vary by participant and by electrode. Model estimates for both modalities are summarised in Table 3. In both modalities, LP (visual $M = -3.04 \mu\text{V}$, $SD = 4.78$; auditory-frontal $M = -3.96 \mu\text{V}$, $SD = 5.19$; auditory-prefrontal $M = -6.15 \mu\text{V}$, $SD = 5.11$) elicited a significantly more positive response than the HP (visual $M = -4.02 \mu\text{V}$, $SD = 4.58$; auditory-frontal $M = -4.95 \mu\text{V}$, $SD = 5.42$; auditory-prefrontal $M = -7.47 \mu\text{V}$, $SD = 5.04$) and IN (visual $M = -4.75 \mu\text{V}$, $SD = 5.12$; auditory-frontal $M = -7.65 \mu\text{V}$, $SD = 5.03$; auditory-prefrontal $M = -9.42 \mu\text{V}$, $SD = 5.14$) conditions. The HP and IN conditions did not differ significantly in the visual modality (though note that in the trial-level analyses, 92% of the posterior for this effect was above 0, Table S3.3). HP elicited a more positive response than IN in the auditory modality (although only marginally so in the analysis using the subset of prefrontal channels).

Discussion

When participants assess whether a word fits with a category cue, incongruous words elicit more negative N400s than predictable words do. Low-predictability words also elicit more negative N400 responses than high-predictability words, although in the present study the effect was only marginally significant in the visual modality.⁵ Prior work in the visual modality, however, has repeatedly shown this pattern (see, e.g. Federmeier et al., 2010; Heinze, Munte, & Kutas, 1998; Stuss, Picton, & Cerri, 1988). Here, we show that this pattern holds as well in the auditory modality, though it has a broader spatial distribution. In the post-N400 time window, low-predictability words elicit the largest positivity in both modalities. In the visual modality, responses to high-predictability and incongruous words were similar in this time window, as has previously been reported (Federmeier et al., 2010). High-predictability words elicit more positivity than incongruous words in the auditory modality. These patterns extend the results in Federmeier et al. (2010) to the auditory domain, as well as highlight potential differences between modalities (though note that item-level analyses suggest these differences may be more quantitative than qualitative) Though it is not unusual to observe slight differences in spatial and/or temporal qualities of ERP effects across modalities, it will be important for future investigations to determine whether this more gradient response in the auditory domain reflects a potential substantive difference between auditorily- and visually-elicited predictions.

Experiment 2

The goal of Experiment 2 was to test the talker-specificity of predictions in language comprehension. Participants heard a dialogue between an (unnamed) interviewer and one of two talkers - Susan or Bob. The dialogue revealed that this talker had idiosyncratic experiences with certain categories (e.g. Susan works on an ox farm) but it revealed no additional information about the other talker (who did not participate in the dialogue). After each dialogue, the participants heard Susan and Bob taking turns providing category cues (e.g. Bob: “Susan, name a farm animal”.) and responding with category exemplars (e.g. Susan: “ox”). At that point, participants had different knowledge about the two talkers. The talker with the “idiosyncratic” experience was presumably associated with low-predictability

⁵There was not strong evidence of a LP vs. HP effect (40% posterior probability of an effect > 0) in the visual modality in the trial-level analysis (see S3.1).

exemplars (e.g. Susan is likely to respond “ox”) but the other talker was expected to be an “average” person (e.g. Bob is likely to respond “cow”). We predicted that listeners would be sensitive to the differences between the two talkers and anticipate responses from a talker that are compatible with his/her personal experience. Thus, high-predictability words would be less expected and low-predictability words more expected for the idiosyncratic talker than the average talker. Note that how this change in the probability space of listeners’ category associations would manifest in the ERP signal (e.g. if it would impact high-predictability word responses, or low-predictability word responses, or both) is an open question.

Method

Participants—Twenty-four right-handed participants (fifteen female, age range: 19–26 years) were recruited from the undergraduate participant pool at the University of Illinois at Urbana-Champaign. They either received course credit or cash as compensation.

Materials & design—The target trials were an expanded version of the materials in Experiment 1 (see osf.io/72p9m/ for the full set). They consisted of 120 category cues (e.g. “<Susan/Bob>, name a farm animal”) paired with a target word for each of three predictability conditions: high-predictability (HP, e.g. “cow”), low-predictability (LP, e.g. “ox”), and incongruous (IN, e.g. “ivy”). HP and LP targets were selected based on existing category norms (Battig & Montague, 1969; Hunt & Hodge, 1971; McEvoy & Nelson, 1982; Shapiro & Palermo, 1970). Targets were matched for length, and frequency across predictability conditions and lists (see osf.io/72p9m/ for details). Each category cue and the three corresponding target words were recorded by two talkers: one male, N.A., and one female, R.R. (the first author). The three types of target words (HP, LP, and IN) were crossed with two talker conditions: idiosyncratic talker and average talker. Both conditions were manipulated within participants. Talker conditions were counterbalanced between Bob and Susan and across subjects. Two other talkers (A.J. and G.M.) recorded the sentences spoken by an interviewer during the dialogues that preceded the blocks of critical trials (a male talker interviewed Susan and a female talker interviewed Bob).

The experiment consisted of twelve blocks of ten test trials. Each block was preceded by a recorded interview in which an interviewer (A.J. or G.M.) asked Bob or Susan questions. These interviews (available in full at osf.io/72p9m/) lasted approximately 2 min and provided information about Bob or Susan’s idiosyncratic experience. For example, the interviewer might ask, *Susan, you have a very unique job. You work on an ox farm. How do you like that?* Susan would respond: *Oh, I love it. They are such sweet animals and it’s a great working environment. I love everything about it, all the way down to the uniforms.* This exchange was intended to establish that, in response to the category cue “Susan, name a farm animal”, Susan would be less likely to respond with the HP target word, “cow” and more likely to respond with the LP target word “ox” than Bob, an “average” talker the participants know nothing about. Both talkers should be equally unlikely to respond with the IN target word, “ivy”. Only LP target words were mentioned in the interviews and always by the interviewer, not the idiosyncratic talker. Two versions of the interviews were created: (1) one in which Susan was the idiosyncratic talker (recorded by the female talker who recorded target trials spoken by Susan) with a male interviewer (a second male talker), and (2) one in

which Bob was the idiosyncratic talker (recorded by the male talker who recorded target trials spoken by Bob) with a female interviewer (a second female talker).

Thus, each participant heard 120 test trials: 48 HP targets, 48 LP targets, and 24 IN targets (fewer IN trials were included in an attempt to keep the task to a reasonable timeframe, under 3 h including setup). Half of the targets from each type of target word were produced by the idiosyncratic talker and the other half were produced by the average talker. Participants heard a particular category cue (e.g. “farm animal”) only once. The pairing of category cues and predictability condition, as well as which talker was idiosyncratic was counterbalanced across participants using six lists created by Latin Square assignment. The order of blocks and trials within blocks was randomised for each participant.

Procedures—The procedure was identical to that of Experiment 1, but in this experiment, participants were told that they would be hearing short interviews with one of the two talkers, Bob or Susan. After that, they would hear them asking each other to name examples of categories, such as “A kind of tree”, followed by a target word, (e.g. “oak”) and they would have to respond “yes” if the target word fit with the category cue and “no” if it did not. The button corresponding to the response was counterbalanced across participants. The experimental session started with five practice trials. Stimulus presentation timing was identical to the auditory version of Experiment 1.

EEG recording and analysis—EEG recording and data processing were identical to those for Experiment 1. An average of 4.7% of trials per participant (range 0–18.5%) was rejected due to artefacts. Analyses of ERP data only included trials on which participants responded accurately (i.e. said “yes” to HP and LP items and “no” to IN items for both talkers).⁶

Results

Behavioural

Accuracy: Participants responded accurately on 98.6% of HP, 95.1% of LP, and 99.3% of IN trials produced by the average talker. They responded accurately on 99.1% of HP, 94.3% of LP, and 99.0% of IN trials produced by the idiosyncratic talker. Accuracy was analysed in a mixed effects logistic regression with predictability condition (Contrast 1: IN = -0.66 vs. HP = 0.33 & LP = 0.33; Contrast 2: HP = -0.5 vs. LP = 0.5) and talker condition (idiosyncratic = -0.5 vs. average = 0.5) as fixed effects and participant and target word as random intercepts. Models with maximal random effects structure did not converge, so slopes explaining the least variance were iteratively removed. Model estimates are summarised in Table 4 (see Supplemental Material 2 for random effects structure). Participants were marginally less accurate on LP trials than on HP trials but not on HP and LP compared to IN trials. There was no significant effect of talker condition, nor did it interact significantly with predictability condition.

⁶Data loss due to artefacts and incorrect responses was uniform across conditions ($M = 7\%$, min = 0%, max = 29%)

Response times: Response times shorter than 50 milliseconds and longer than 3000 milliseconds were excluded from analyses. This affected 1.35% of data points. Average response times were 1009 milliseconds for HP-average, 991 milliseconds for HP-idiosyncratic, 1064 for LP-average, 1067 milliseconds for LP-idiosyncratic, 1156 for IN-average, and 1154 milliseconds for IN-idiosyncratic. Log-transformed response times were analysed in a mixed effects linear regression. The fixed predictors were word condition entered as a pair of orthogonal contrast codes, IN vs. HP & LP (Contrast 1: IN = -0.66, HP = 0.33, LP = 0.33) and HP vs. LP (Contrast 2: IN = 0, HP = 0.5, LP = -0.5), and talker condition (average: 0.5, idiosyncratic: -0.5). The model also included random intercepts for participant and target word, along with maximal random slopes. Model estimates are summarised in Table 4 (see Supplemental Material 2 for random effects structure). Participants were significantly slower on IN trials relative to LP and HP trials, and slower on LP trials than HP trials. There was no effect of talker condition and no interaction between talker condition and predictability condition.

ERPs—Average ERPs for each predictability condition (collapsing across talker), for a parietal and a prefrontal channel, can be seen in Figure 2. Visual inspection indicates that incongruous words elicit more negative responses than high-predictability words in the N400 time window, as in Experiment 1. However, low-predictability words elicit a more positive response than high-predictability words, reflecting an N400 repetition effect (Rugg, 1985) due to the fact that the low-predictability words are the only ones that have been recently experienced during the interviews. Over frontal channels in the post-N400 window, low-predictability words elicit more positivity than high-predictability words and high-predictability words appear to elicit a more positive response than incongruous words (as in Experiment 1).

To explore the interaction between predictability condition and talker condition, average ERPs for each condition are plotted in Figure 3. For clarity, only high-predictability and incongruous words are included because the amplitude of the low-predictability words is not comparable due to the effects of repetition (see Supplemental Figure 1 for figure with all six conditions). The data were analysed to examine two components: N400 and frontal positivity (in all frontal electrodes as well as the subset of prefrontal electrodes).

N400: As in Experiment 1, mean amplitudes across trials⁷ were computed between 300 and 600 milliseconds after the onset of the target word from eight centro-posterior electrodes (see Figure 1) and were modelled by linear mixed effects regression. Predictability, as a pair of contrasts (Contrast 1: IN = -0.66 vs. HP = 0.33 & LP = 0.33; Contrast 2: HP = -0.5 vs. LP = 0.5), talker condition (average = 0.5 vs. idiosyncratic talker = -0.5), and their interaction were entered as fixed effects. Participant and electrode were entered as random intercepts with maximal random slopes (both main effects and the interaction varied by participants and electrodes). Model estimates are summarised in Table 5. We observed significant effects of predictability condition. IN elicited a significantly more negative response than the HP and LP conditions and the HP condition elicited a more negative

⁷See Supplemental Material 3 for analyses with trial-level data. We found evidence for the same contrasts in these analyses, unless noted otherwise in the main text.

response than the LP condition. The latter is likely due to the fact that the LP words were recently mentioned, and repetition tends to reduce the N400 amplitude (Rugg, 1985). There was no significant main effect of talker condition or any significant interactions between talker condition and predictability condition (predictability contrast 1 \times talker $BF_{01} = 4.24$ and predictability contrast 2 \times talker $BF_{01} = 6.45$ in favour of the null hypothesis [from model in Supplemental Material 3]; note the Bayes Factor calculation is highly dependent on the prior distribution: in this case, a normal distribution centred on 0 with a SD of 10).

Post-N400 frontal positivity: Given the broadly frontal (as opposed to exclusively prefrontal) distribution of the late positivity effects observed in Experiment 1, we analysed amplitudes from all frontal electrodes (MiPf, LLPf, RLPf, LMPf, RMPf, LDFr, RDFr, LMFr, RMFr, LLFr, RLFr) as well as the subset of prefrontal electrodes for continuity with the literature (MiPf, LLPf, RLPf, LMPf, RMPf; see Figure 2). These amplitudes were averaged across trials⁸ between 600 and 900 milliseconds (in the same way as in the auditory version of Experiment 1) and were analysed in two linear mixed effects regressions with predictability condition (Contrast 1: LP = -0.66 vs. HP = 0.33 & IN = 0.33; Contrast 2: HP = -0.5 vs. IN = 0.5), talker condition (average = 0.5 vs. idiosyncratic = -0.5), and their interaction entered as fixed effects. Participant and electrode were entered as random intercepts with maximal random slopes. Model estimates are summarised in Table 6. The LP condition elicited a significantly more positive response than the HP and IN conditions and the HP condition elicited a more positive response than the IN condition.⁹ There was no main effect of talker condition, but there was a significant interaction between predictability condition (Contrast 2: HP vs. IN) and talker condition (although only marginally so in the prefrontal subset, $p = 0.050$). Follow-up analyses using the emmeans package (Length, 2018), examining the effect of talker condition within each predictability condition found no significant effects of talker condition in LP targets¹⁰ (average talker-frontal $M = 0.18 \mu\text{V}$, $SD = 4.29$; idiosyncratic talker-frontal $M = 0.36 \mu\text{V}$, $SD = 4.45$; $b = 0.177$, $SE = 0.627$, $t = 0.283$, $p = 0.780$; average talker-prefrontal $M = -2.04 \mu\text{V}$, $SD = 4.17$, idiosyncratic talker-prefrontal $M = -1.76 \mu\text{V}$, $SD = 4.42$, $b = 0.283$, $SE = 0.652$, $t = 0.4346$, $p = 0.669$) or IN targets (average talker-frontal $M = -2.861 \mu\text{V}$, $SD = 4.38$, idiosyncratic talker-frontal $M = -3.644 \mu\text{V}$, $SD = 4.70$, $b = -0.782$, $SE = 0.705$, $t = -1.110$, $p = 0.279$; average talker-prefrontal $M = -4.69 \mu\text{V}$, $SD = 4.22$ idiosyncratic talker-prefrontal $M = -5.38 \mu\text{V}$, $SD = 4.59$, $b = -0.690$, $SE = 0.782$, $t = -0.882$, $p = 0.387$). However, there was an effect of talker condition in HP targets (average talker-frontal $M = -1.575 \mu\text{V}$, $SD = 3.97$, idiosyncratic talker-frontal $M = -0.164 \mu\text{V}$, $SD = 4.72$, $b = 1.411$, $SE = 0.627$, $t = 2.250$, $p = 0.034$; average talker-prefrontal $M = -3.55 \mu\text{V}$, $SD = 3.93$, idiosyncratic talker-prefrontal $M =$

⁸.See Supplemental Material 3 for analyses with trial-level data. We found evidence for the same contrasts in these analyses, unless noted otherwise in the main text.

⁹.As requested by a reviewer, in a post-hoc analysis, we compared the HP and LP conditions directly (Contrast 1: IN = -0.66 vs. HP = 0.33 & LP = 0.33; Contrast 2: HP = -0.5 vs. LP = 0.5). HP and LP words elicited a more positive response than IN words ($b = 2.954$, $SE = 0.615$, $t = 4.804$, $p < 0.001$). LP words elicited a more positive response on average than HP words ($b = 1.142$, $SE = 0.525$, $t = 2.175$, $p = 0.0397$). There was no main effect of talker condition ($b = -0.269$, $SE = 0.391$, $t = -0.687$, $p = 0.499$). Neither interaction between talker condition and predictability condition was significant (Contrast 1: $b = -1.577$, $SE = 0.909$, $t = -1.734$, $p = 0.096$; Contrast 2: $b = 1.234$, $SE = 0.740$, $t = 1.667$, $p = 0.109$).

¹⁰. $BF_{01} = 11.31$ in favor of the null hypothesis (using model in Supplemental Material 3). Note the Bayes Factor calculation is highly dependent on the prior distribution: in this case, a normal distribution centered on 0 with a SD of 10

$-2.21 \mu\text{V}$, $SD = 4.77$, $b = 1.346$, $SE = 0.711$, $t = 1.893$, $p = 0.072$), such that HP words elicited a more positive response in the idiosyncratic talker condition.

Discussion

The results of Experiment 2 conceptually replicated the auditory version of Experiment 1. Incongruous target words elicited a larger N400 than high-predictability and low-predictability words. However, low-predictability words elicited a smaller N400 than high-predictability in this paradigm, because the low-predictability words were recently presented (albeit as different perceptual tokens) during the short interviews preceding each block of test trials and thus induced (well-replicated) N400 repetition effects (see review in Kutas & Federmeier, 2011). The frontal positivity patterns were also broadly consistent with previous work: low-predictability words elicited the largest positivity relative to incongruous and high-predictability target words. Although we cannot rule out the possibility of continued effects of repetition on the LP items, there is an extensive literature on ERP repetition effects showing that, when post-N400 repetition effects occur, they typically have a posterior distribution (e.g. Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). To our knowledge, there are no reported effects of repetition on the frontal positivity. Thus, it seems plausible that this pattern simply replicates the effect observed when there is no repetition, as in Experiment 1.

High-predictability words elicited a significantly larger positive response than incongruous words, as in Experiment 1. Critically, this effect was modulated by an interaction between predictability condition and talker condition: high-predictability target words produced by the idiosyncratic talker elicited more positive responses than high-predictability target words produced by the average talker, suggesting that listeners found those words unexpected when produced by the idiosyncratic talker compared to the average talker.

General discussion

In this work, we examined the interplay between language-wide statistics and talker-specific predictions in online language processing. We first extended prior work (Federmeier et al., 2010) showing that (1) the predictability of a word given a semantic category has a graded effect on semantic access (as indexed by the N400 ERP component), and (2) the process of integrating a word that is not strongly predicted but is congruent with the category context elicits an increased late frontal positivity in the ERP signal. In our data, these effects were most evident in the auditory domain (previous work has repeatedly observed them in the visual domain; Federmeier et al., 2007, 2010). In a second experiment, we manipulated talker-specific category information: one of the two speakers had an idiosyncratic experience or preference (e.g. worked on an ox farm) that should affect what category exemplar they are most likely to produce (e.g. “ox” rather than “cow” for the “farm animal” category). We found no evidence that this talker-specific information had any discernable impact on the initial semantic access to the words, as indexed by the N400 (although note that the nature of the paradigm-i.e. repetition of LP items-may preclude detection of subtler N400 effects and that this is a null result and thus must be interpreted with caution). However, words that were highly predictable given the category and language-wide statistics (e.g. cow) elicited an

increased frontal positivity when spoken by the idiosyncratic compared to the average talker, suggesting that processing of these words was affected by listeners' recently-gained knowledge of that talker's idiosyncratic experience. In other words, talker-specific information interacted with prior world and language knowledge.

The absence of talker-specificity in the N400 time window may seem surprising given previous work showing that the N400 is sensitive to the discourse (Nieuwland & Van Berkum, 2006) and general aspects of talker context (Van Berkum et al., 2008). For instance, the word "wine" in "Every evening I drink some wine before I go to sleep" elicits larger N400s when spoken by a young child than by an adult. However, a key difference between the previous work and our own is that, in the present study, the voices themselves carried no differential a priori semantic or pragmatic content: both critical talkers were young adults with Midwestern-American English accents and gender was counterbalanced across participants. Thus, any differential predictions associated with the voices had to be constructed based on recent exposure to talker-specific associations. We found that the recent exposure to the low predictability words facilitated their processing overall, in the form of a general repetition effect on the N400. However, we observed no differential facilitation for this condition based on talker. It is possible that subtle effects of talker could have been masked by the effect of repetition, although other relatively small effects on the N400, such as those due to word frequency, have been found even in the presence of word repetition (e.g. Rugg, 1990). Investigating how representations of low predictability words are affected by novel talker-specific knowledge (in a situation where this is not confounded with repetition of the exact word), and whether they are more affected than high predictability words, is an important avenue for future research.

The N400 has been argued to reflect activation spreading through multi-modal semantic networks stored in long-term memory (Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011). In this framework, we might not expect talker-specific N400 facilitations for the low predictability words, because the newly-learned associations about the idiosyncratic talker are unlikely to have altered long-term semantic stores. It may be the case that, given more time or repeated exposure to talker-specific knowledge, talker-specific effects of this nature could emerge in the N400 time window.

We also observed no talker-based N400 differences in response to the high predictability items, a condition that was not subject to repetition effects. Thus, it seems that knowledge that a particular talker may not share average experience and/or preferences does not override the normal facilitation obtained for words that are likely to follow a given cue. This pattern is consistent with findings that quantification and negation also often do not affect the N400 - i.e. one sees similar levels of facilitation for "crops" across "No/some/all farmers grow crops" (Fischler, Bloom, Childers, Roucos, & Perry, 1983; Urbach & Kutas, 2010). Similarly, Moreno, Casado, and Martín-Loeches (2016) found that N400s are facilitated for words that are conceptually primed by a given discourse but that are socially inappropriate to express, and thus unlikely to actually be realised in that discourse. For example, given a story like, "Ines has put on too much makeup and she looks older. When she asks her friend how the makeup looks on her, her friend says, 'The makeup highlights your ...'" N400 facilitation was observed both for "cheeks" (the expected "white lie") as well as for

“wrinkles”, even though comprehenders find such “blunt truth” completions unexpected. Thus, our findings are consistent with a body of literature suggesting that linguistic terms (like quantification markers) and social or talker-specific knowledge that may eventually “override” information typically associated with particular cues and events do not have their effects on early stages of semantic access, as indexed by the N400.

However, effects on the frontal positivity make clear that comprehenders can and do use recently-obtained talker-specific knowledge to shape their expectations about what people will say, affecting processing in a time window after the N400. The frontal positivity has been seen in both cueing (Federmeier et al., 2010) and sentence processing paradigms (Federmeier et al., 2007) wherein people are making predictions and those predictions are violated because an unexpected - but nonetheless plausible - word is presented instead. It contrasts with effects over posterior sites in this time window (on what is typically termed the late positive complex [LPC]), which can be observed for words that are not only unexpected but wholly incongruent in context (see review by Van Petten & Luka, 2012). Thus, frontal positivities would not be expected in this paradigm for the incongruous words (which also have not generally been shown to elicit LPC effects in this kind of paradigm; Federmeier et al., 2010). Because of this specificity to plausible words, the frontal positivity has been associated with efforts to integrate unexpected information into the context. It is most readily observed when a word is in the focus of attention (Wlotko & Federmeier, 2007) and is subject to individual differences (Federmeier et al., 2010) and task strategies (Payne & Federmeier, 2017), suggesting the processing it indexes is more attentionally demanding/controlled than that manifesting in the N400.

In the current experiment, talker-specific information appears to be brought to bear at this later processing stage (though it may be the case that we simply failed to detect analogous effects at earlier stages). We found that low predictability words elicited a frontal positivity (compared to the incongruous condition), similar to that seen in Experiment 1 and in prior work using the category cueing paradigm (Federmeier et al., 2010), and that this pattern was not modulated by talker. This would seem to suggest, although caution is always warranted in interpreting a null effect, that low predictability words continue to be considered unexpected (but plausible), even for the idiosyncratic talker. Repetition effects are not typically seen over frontal sites in this time window (e.g. Van Petten et al., 1991), although we cannot rule out the possibility that the effects we observe here might be influenced by the fact that comprehenders had heard these words (spoken by a different talker) recently. The lack of talker effect on this condition is consistent with the hypothesis that comprehenders do not come to have a specific expectation for the idiosyncratic talker to produce the low predictability word, or at least do not do so quickly enough to modulate their brain responses in the first second of processing. An important avenue for future investigation will be to characterise the timecourse of these talker-specific effects.

However, we do find talker effects on the responses to high predictability words. In particular, comprehenders elicited a larger frontal positivity to these items when they were spoken by the idiosyncratic talker than by the average talker. Thus, comprehenders appear to appreciate that it is unexpected for the idiosyncratic talker to provide a high predictability completion to the category cue. Taken together, the pattern of results on the frontal positivity

suggests that comprehenders can use newly learned information about a talker to determine that they are unlikely to respond typically to a particular cue, even if they cannot necessarily recall the specific information that talker is likely to provide in lieu of the high predictability word.

In prior work, frontal positivities have most typically been observed when an unexpected but plausible item is encountered in the face of a strong prediction for a different word (e.g. Federmeier et al., 2007; Federmeier et al., 2010; see review by Van Petten & Luka, 2012). Under those circumstances, then, a word that was not already active (as indexed by large N400 responses) must be integrated into the message level representation in the face of possible competition from a different, highly active word (the predicted but never obtained completion). Here, however, we observe the frontal positivity to a word that had already been activated, as evidenced by facilitated N400 responses. This pattern is similar to that seen by Moreno et al. (2016), who found enhanced frontal positivity to “blunt truth” completions (“wrinkles” in the previously described example), which had been activated by the discourse meaning - and showed facilitated N400 responses - but which, because of social norms, were considered to be unlikely to be realised in the discourse. Thus, the frontal positivity seems to reflect mechanisms recruited to deal with plausible yet unexpected words, regardless of the source of the unexpectedness - i.e. that were not activated to begin with or that had been activated but were not selected or were even suppressed due to social norms or, in the present case, talker-specific expectations.

The finding of talker-specific expectations during online language processing is consistent with an emerging literature indicating that listeners take into account the identity of the talker when interpreting linguistic input, across many levels of language representation (Brehm, Jackson, & Miller, 2019; Cai et al., 2017; Creel, 2012; 2014; Hanulíková et al., 2012; Trude & Brown-Schmidt, 2012; Van Berkum et al., 2008; see also Brown-Schmidt, Yoon, & Ryskin, 2015). Despite the tentative consensus that talker-specific information plays a role in comprehension, little is known about the nature of the representations that are brought to bear in this process. On one hand, talker-specific information can be learned from repeated exposure to co-occurrence of a talker and a particular lexical item or phonetic realisation of a word (Kleinschmidt & Jaeger, 2015). With experience, recognition of the word becomes facilitated when produced by the talker it has been paired with and not when produced by a different talker (e.g. Creel, 2012; 2014; Trude & Brown-Schmidt, 2012). This type of bottom-up information has been proposed to rely on implicit learning mechanisms (e.g. Trude, Duff, & Brown-Schmidt, 2014) and therefore the talker-specific association may emerge over time, as evidence accumulates.

On the other hand, talker-specific expectations can reflect a top-down attribution about the talker’s tendencies. For example, listeners process grammatical violations differently when spoken by a foreign-accented speaker than a native speaker (Hanulíková et al., 2012). Similarly, lexical access appears to be faster for word meanings associated with the accent of the talker (e.g. *hat* meaning is more facilitated for “bonnet” when spoken with an American than a British English accent; Cai et al., 2017). Such talker-specific processing presumably reflects generalisation from past experiences with similar talkers and thus can happen swiftly, but only when the listener is aware of the relevant differences between talkers.

Whether and how bottom-up and top-down sources of information about a talker interact in comprehension is an open question.

In the present work, we provide a first glimpse into the neural timecourse of talker-specific semantic processing. Insights from this work, in particular the contrast between effects in the N400 and frontal positivity windows, may help to circumscribe the underlying mechanisms. For instance, the current paradigm primarily manipulated top-down information about the talkers. During comprehension of the categories and target words, listeners recognised the particular talker as one who had an explicit idiosyncratic preference, or perhaps more generally as one who had many idiosyncratic preferences, and adjusted their predictions accordingly. Consistent with a view of this as an attention-demanding process, this adjustment was only detectable during the post-N400 period. As this is, to our knowledge, the first study to examine talker-specific semantic category prediction, replications and extensions will be necessary for a thorough grasp of the magnitude of this effect and the circumstances under which it manifests. A potential avenue for future research would be to explore how repeated exposure to bottom-up information about a talker might affect online semantic category predictions and whether those effects might eventually manifest at earlier time windows by virtue of changing the core semantic networks.

Conclusion

The pre-activation of upcoming linguistic material is guided by world and linguistic knowledge but also tailored to the individual talker. These two sources of information appear to have their initial effects on different aspects of unfolding comprehension. An important avenue for future work will be to explore the memory and attention mechanisms that enable talker-specific predictions and the circumstances under which they override prior knowledge.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Nathaniel Anderson, Ariel James, and Geoffrey McKinley for help with recording audio stimuli.

References

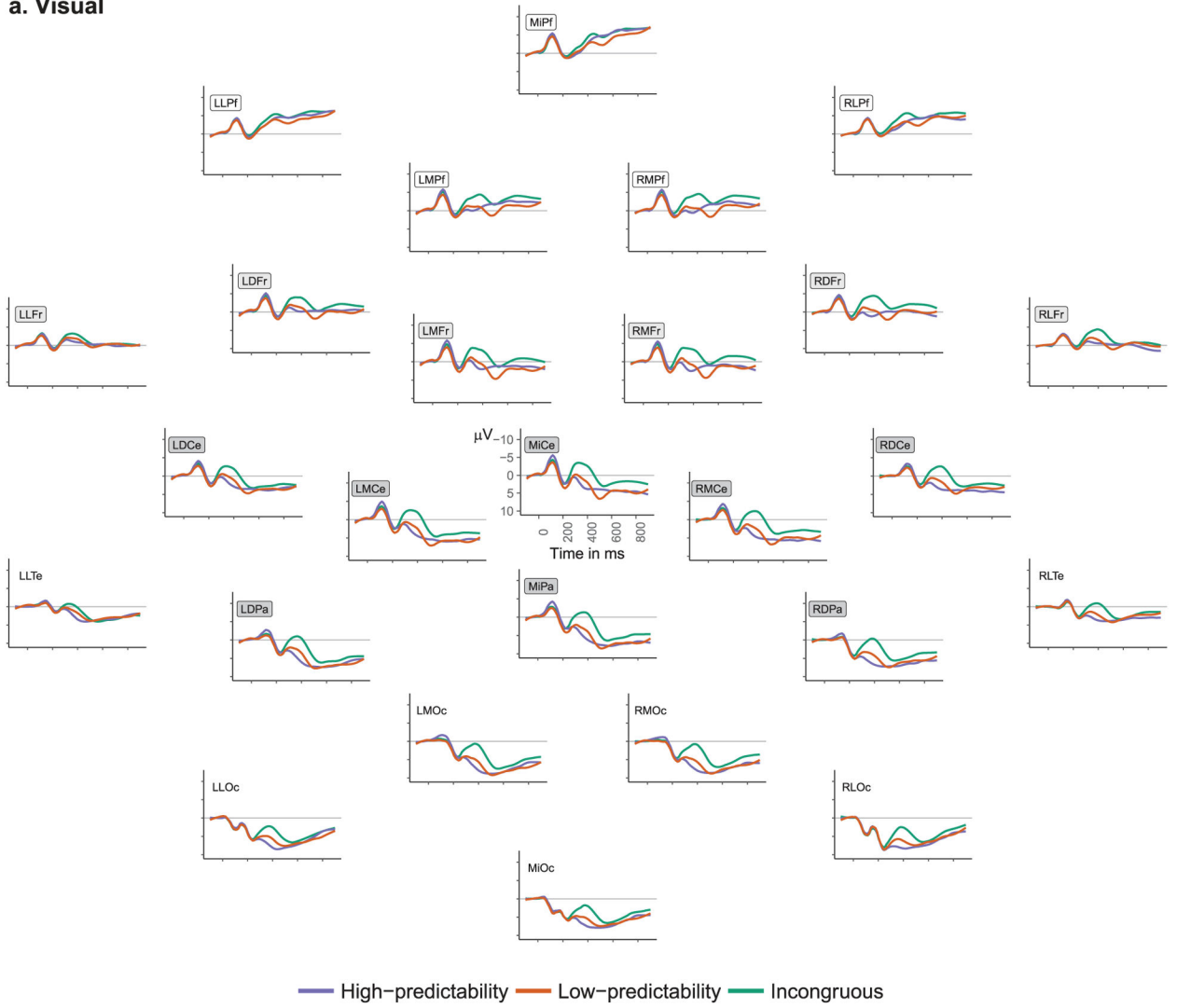
- Altmann GTM, & Kamide Y (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. doi:10.1016/S0010-0277(99)00059-1 [PubMed: 10585516]
- Bates D, Mächler M, Bolker B, & Walker S (2014). Fitting Linear Mixed-Effects Models using lme4. 67(1). doi:10.18637/jss.v067.i01
- Battig WF, & Montague WE (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 80(3. Pt.2), 1–46. doi:10.1037/h0027577
- Brehm L, Jackson CN, & Miller KL (2019). Speaker-specific processing of anomalous utterances. *Quarterly Journal of Experimental Psychology*, 72(4), 764–778. doi:10.1177/1747021818765547

- Brown-Schmidt S, Yoon SO, & Ryskin RA (2015). People as contexts in conversation. In Ross BH (Ed.), *Psychology of learning and motivation* (Vol. 62, pp. 59–99). Academic Press. doi:10.1016/bs.plm.2014.09.003
- Bürkner P-C (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395–411. doi:10.32614/RJ-2018-017
- Cai ZG, Gilbert RA, Davis MH, Gaskell MG, Farrar L, Adler S, & Rodd JM (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, 98, 73–101. [PubMed: 28881224]
- Creel SC (2012). Preschoolers' use of talker information in online comprehension. *Child Development*, 83(6), 2042–2056. doi:10.1111/j.1467-8624.2012.01816.x [PubMed: 22803622]
- Creel SC (2014). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language*, 73(1), 81–98. doi:10.1016/j.jml.2014.03.001
- Creel SC, Aslin RN, & Tanenhaus MK (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 106(2), 633–664. doi:10.1016/j.cognition.2007.03.013 [PubMed: 17507006]
- Creel SC, & Bregman MR (2011). How talker identity relates to language processing. *Linguistics and Language Compass*, 5(5), 190–204. doi:10.1111/j.1749-818X.2011.00276.x
- Dahan D, Drucker SJ, & Scarborough RA (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108(3), 710–718. doi:10.1016/j.cognition.2008.06.003 [PubMed: 18653175]
- Federmeier KD (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505. doi:10.1111/j.1469-8986.2007.00531.x [PubMed: 17521377]
- Federmeier KD, & 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. doi:10.1006/jmla.1999.2660
- Federmeier KD, Kutas M, & Schul R (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149–161. doi:10.1016/j.bandl.2010.07.006 [PubMed: 20728207]
- Federmeier KD, & Laszlo S (2009). Chapter 1 time for meaning. Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. In Ross BH (Ed.), *Psychology of learning and motivation* (Vol. 51, pp. 1–44). Academic Press. doi:10.1016/S0079-7421(09)51001-8
- Federmeier KD, Wlotko EW, De Ochoa-Dewald E, & Kutas M (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146(1), 75–84. doi:10.1016/j.brainres.2006.06.101 [PubMed: 16901469]
- Fischler I, Bloom PA, Childers DG, Roucos SE, & Perry NW (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20(4), 400–409. doi:10.1111/j.1469-8986.1983.tb00920.x [PubMed: 6356204]
- Hanulíková A, van Alphen PM, van Goch MM, & Weber A (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878–887. doi:10.1162/jocn_a_00103 [PubMed: 21812565]
- Heinze H-J, Muentz TF, & Kutas M (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology*, 47(2), 121–135. [PubMed: 9554184]
- Hunt KP, & Hodge MH (1971). Category-item frequency and category-name meaningfulness (m'): Taxonomic norms for 84 categories. *Psychonomic Monograph Supplements*, 4(6), 97–121.
- Johnson KE, & Mervis CB (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, 126(3), 248–277. doi:10.1037/0096-3445.126.3.248 [PubMed: 9281832]
- Kleinschmidt DF, & Jaeger TF (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. doi:10.1037/a0038695 [PubMed: 25844873]
- Kruschke J (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed). Academic Press.

- Kuperberg GR, & Jaeger TF (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. doi:10.1080/23273798.2015.1102299
- Kutas M, & Federmeier KD (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. doi:10.1146/annurev.psych.093008.131123
- Kutas M, & Van Petten CK (1994). Psycholinguistics electrified: Event-related brain potential investigations. In *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.
- Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). Lmertest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. doi:10.18637/jss.v082.i13
- Lenth R (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.2. <https://CRAN.Rproject.org/package=emmeans>.
- Lund K, & Burgess C (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2), 203–208.
- McEvoy CL, & Nelson DL (1982). Category name and instance norms for 106 categories of various sizes. *The American Journal of Psychology*, 95(4), 581–634. doi:10.2307/1422189
- Moreno EM, Casado P, & Martín-Loeches M (2016). Tell me sweet little lies: An event-related potentials study on the processing of social lies. *Cognitive, Affective and Behavioral Neuroscience*, 16(4), 616–625. doi:10.3758/s13415-016-0418-3
- Nieuwland MS, Politzer-Ahles S, Heyselaar E, Segaert K, Darley E, Kazanina N, ... Mézière D (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. [PubMed: 29631695]
- Nieuwland MS, & Van Berkum JJA (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18(7), 1098–1111. doi:10.1162/jocn.2006.18.7.1098 [PubMed: 16839284]
- Payne BR, & Federmeier KD (2017). Pace yourself: Intraindividual variability in context use revealed by self-paced event-related brain potentials. *Journal of Cognitive Neuroscience*, 29(5), 837–854. doi:10.1162/jocn [PubMed: 28129064]
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rugg MD (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22(6), 642–647. doi:10.1111/j.1469-8986.1985.tb01661.x [PubMed: 4089090]
- Rugg MD (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory Cognition*, 18(4), 367–379. [PubMed: 2381316]
- Ryskin RA, Wang RF, & Brown-Schmidt S (2016). Listeners use speaker identity to access representations of spatial perspective during online language comprehension. *Cognition*, 147, 75–84. doi:10.1016/j.cognition.2015.11.011 [PubMed: 26638050]
- Shapiro SI, & Palermo DS (1970). Conceptual organization and class membership: Normative data for representatives of 100 categories. *Psychonomic Monograph Supplements*, 3(11), 107–127.
- Stuss D, Picton T, & Cerri A (1988). Electrophysiological manifestations of typicality judgment. *Brain and Language*, 33(2), 260–272. [PubMed: 3359171]
- Tanaka JW, & Taylor M (1991). Object categories and expertise: Is the basic-level in the eye of the beholder? *Cognitive Psychology*, 23, 457–482. doi:10.1016/0010-0285(91)90016-H
- Trude AM, & Brown-Schmidt S (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27(7–8), 979–1001. doi:10.1080/01690965.2011.597153
- Trude AM, Duff MC, & Brown-Schmidt S (2014). Talker-specific learning in amnesia: Insight into mechanisms of adaptive speech perception. *Cortex*, 54(1), 117–123. doi:10.1016/j.cortex.2014.01.015 [PubMed: 24657480]
- Urbach TP, & Kutas M (2010). Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63(2), 158–179. doi:10.1016/j.jml.2010.03.008 [PubMed: 20640044]

- Van Berkum JJA, van den Brink D, Tesink CMJY, Kos M, & Hagoort P (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–591. doi:10.1162/jocn.2008.20054 [PubMed: 18052777]
- Van Petten C, Kutas M, Kluender R, Mitchiner M, & McIsaac H (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, 3, 131–150. [PubMed: 23972089]
- Van Petten C, & Luka BJ (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190. doi:10.1016/j.ijpsycho.2011.09.015 [PubMed: 22019481]
- Wlotko EW, & Federmeier KD (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. *Neuropsychologia*, 45(13), 3001–3014. doi:10.1016/j.neuropsychologia.2007.05.013 [PubMed: 17659309]

a. Visual



b. Auditory

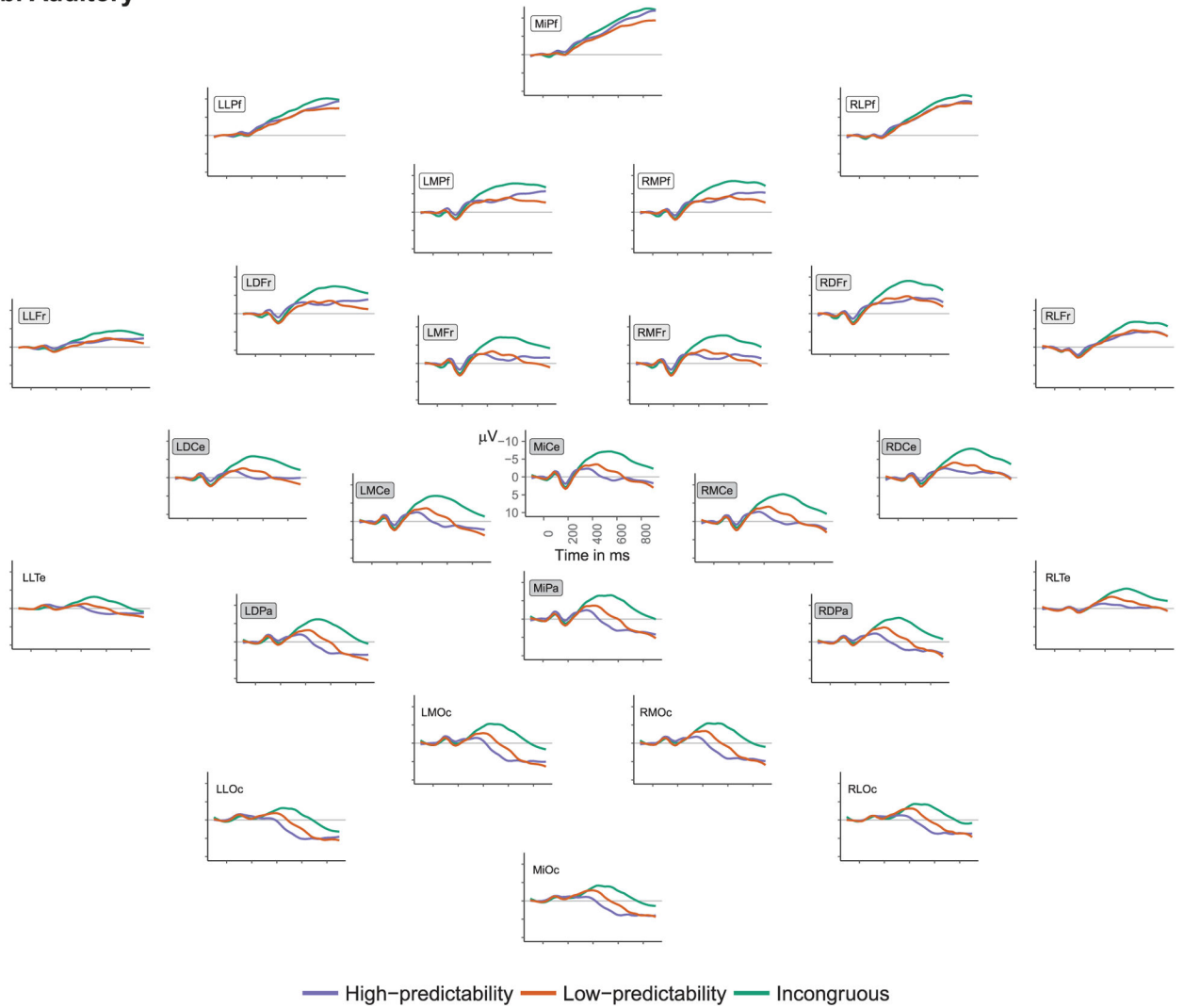


Figure 1. Average ERPs in response to target words by predictability condition for the (a) visual and (b) auditory modalities. Channels included in analyses (either in the N400 window or later frontal positivity window) are indicated by the presence of rectangles around their labels (N400 electrodes = gray fill, Frontal electrodes = light gray and white fill, Prefrontal electrodes = white fill).

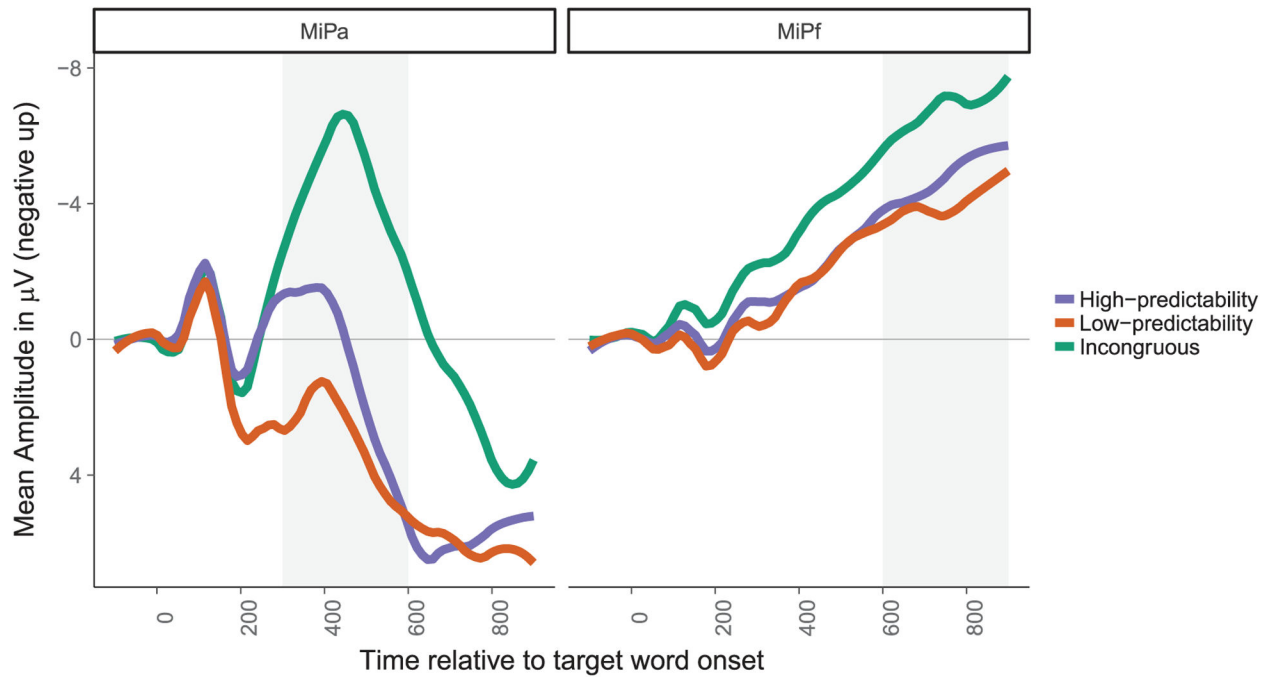


Figure 2.

Average ERPs in response to target words by predictability condition (collapsing over talker condition). One channel (MiPa) included in analysis of the N400 window is shown on the left and one channel (MiPf) included in analysis of the late frontal positivity window is shown on the right.

Table 1.

Fixed effect estimates and standard errors from mixed effects regression analyses of response accuracy and log-transformed response time in both Visual and Auditory modalities by Condition (HP = High-Predictability; LP = Low-Predictability; IN = Incongruous).

	Accuracy		Response Time	
	Visual	Auditory	Visual	Auditory
Intercept	6.774 (1.307)	6.031 (0.814)	6.615 (0.055)	6.919 (0.035)
	$p = 0.000$ ***	$p = 0.000$ ***	$p = 0.000$ ***	$p = 0.000$ ***
IN vs. HP and LP	-1.870 (1.453)	0.306 (0.847)	-0.100 (0.027)	-0.146 (0.023)
	$p = 0.198$	$p = 0.718$	$p = 0.0002$ ***	$p = 0.001$ ***
HP vs. LP	2.004 (0.870)	3.125 (1.050)	-0.169 (0.022)	-0.147 (0.024)
	$p = 0.021$ *	$p = 0.003$ **	$p = 0.000$ ***	$p = 0.000$ ***
Observations	1,728	1,728	1,713	1,711
Log Likelihood	-207.552	-222.896	-375.681	-66.055
Akaike Inf. Crit.	435.103	465.791	773.362	154.110
Bayesian Inf. Crit.	489.651	520.339	833.268	214.004

Note:

* $p < 0.05$;

** $p < 0.01$;

*** $p < 0.001$.

Table 2.

Fixed effects estimates and standard errors from mixed effects linear regression analyses of mean amplitude in the N400 time window in both Visual (300–500 ms) and Auditory (300–600 ms) modalities by Condition (HP = High-Predictability; LP = Low-Predictability; IN = Incongruous).

	Mean Amplitude in the N400 window	
	Visual	Auditory
Intercept	2.423 (1.181)	-2.951 (0.837)
	$p = 0.041^*$	$p = 0.003^{**}$
IN vs. LP and HP	4.433 (0.540)	4.026 (0.545)
	$p = 0.000^{***}$	$p = 0.000^{***}$
LP vs. HP	1.104 (0.606)	1.996 (0.657)
	$p = 0.088^{\dagger}$	$p = 0.008^{**}$
Observations	384	384
Log Likelihood	-785.705	-661.672
Akaike Inf. Crit.	1,603.410	1,355.345
Bayesian Inf. Crit.	1,666.620	1,418.555

Note:

$^{\dagger} p < 0.1$;

$*$ $p < .05$;

$**$ $p < 0.01$;

$***$ $p < 0.001$.

Random effects estimates available in Supplemental Material S1.

Table 3.

Fixed effects estimates and standard errors from mixed effects linear regression analyses of EEG amplitudes in the post-N400 time window in both Visual (500–900 ms) and Auditory (600–900 ms) modalities (for all frontal and prefrontal channels and prefrontal channels only) by Condition (HP = High-Predictability; LP = Low-Predictability; IN = Incongruous).

	Mean Amplitudes in the post-N400 time window		
	Visual	Auditory-Frontal	Auditory-Prefrontal
Intercept	-3.934 (1.248)	-5.520 (1.247)	-7.682 (1.389)
	$p = 0.006^{**}$	$p = 0.0002^{***}$	$p = 0.000^{***}$
LP vs. HP and IN	-1.361 (0.487)	-2.369 (0.638)	-2.320 (0.759)
	$p = 0.015^*$	$p = 0.002^{**}$	$p = 0.008^{**}$
HP vs. IN	0.731 (0.840)	2.695 (0.741)	1.951 (0.962)
	$p = 0.398$	$p = 0.002^{**}$	$p = 0.059^{\dagger}$
Observations	240	528	240
Log Likelihood	-571.150	-1,245.328	-571.124
Akaike Inf. Crit.	1,174.299	2,522.656	1,174.249
Bayesian Inf. Crit.	1,229.989	2,590.961	1,229.939

Note:

$^{\dagger} p < 0.1$;

$*$ $p < .05$;

$**$ $p < 0.01$;

$***$ $p < 0.001$.

Random effects estimates available in Supplemental Material S1.

Table 4.

Fixed effect estimates and standard errors from mixed effects regression analyses of response accuracy and log-transformed response times in Experiment 2 by Word (HP = High-Predictability; LP = Low-Predictability; IN = Incongruous) and Talker Conditions.

	Accuracy	Response Time
Intercept	9.401 (0.812)	6.922 (0.037)
	$p = 0.000^{***}$	$p = 0.000^{***}$
IN vs. HP and LP	-1.138 (1.121)	-0.132 (0.025)
	$p = 0.310$	$p = 0.00000^{***}$
HP vs. LP	-1.233 (0.713)	0.051 (0.018)
	$p = 0.084$	$p = 0.006^{**}$
Average vs. Idiosyncratic	0.524 (0.680)	0.010 (0.020)
	$p = 0.442$	$p = 0.625$
IN vs. HP and LP \times Average vs. Idiosyncratic	-1.488 (1.928)	0.012 (0.029)
	$p = 0.441$	$p = 0.681$
HP vs. LP \times Average vs. Idiosyncratic	1.024 (0.856)	-0.014 (0.024)
	$p = 0.232$	$p = 0.559$
Observations	2,880	2,841
Log Likelihood	-294.713	-510.482
Akaike Inf. Crit.	605.427	1,082.965
Bayesian Inf. Crit.	653.151	1,267.474

Note:

* $p < .05$;

** $p < 0.01$;

*** $p < 0.001$.

Table 5.

Fixed effects estimates and standard errors from mixed effects linear regression analyses of mean amplitude in the N400 time window by Target Word (HP = High-Predictability; LP = Low-Predictability; IN = Incongruous) and Talker condition in Experiment 2.

	Mean N400 Amplitude
Intercept	-0.360 (0.425)
	$p = 0.404$
IN vs. LP and HP	5.547 (0.525)
	$p = 0.000$ ***
HP vs. LP	1.959 (0.554)
	$p = 0.002$ ***
Average vs. Idiosyncratic	-0.111 (0.375)
	$p = 0.769$
IN vs. LP and HP \times Average vs. Idiosyncratic	-1.340 (0.764)
	$p = 0.092$
HP vs. LP \times Average vs. Idiosyncratic	0.970 (1.003)
	$p = 0.344$
Observations	1,152
Log Likelihood	-2,171.476
Akaike Inf. Crit.	4,440.951
Bayesian Inf. Crit.	4,688.365

Note:

* $p < .05$;

** $p < 0.01$;

*** $p < 0.001$.

Random effects estimates available in Supplemental Material S2.

Table 6.

Fixed effects estimates and standard errors from mixed effects linear regression analyses of mean EEG amplitudes in the FP time window by Word (HP = High-Predictability; LP = Low-Predictability; IN = Incongruous) and Talker Condition.

	Mean FP Amplitude	
	Frontal	Prefrontal
Intercept	-1.283 (0.890)	-3.273 (1.173)
	$p = 0.168$	$p = 0.033^{**}$
LP vs. HP and IN	-2.333 (0.562)	-2.062 (0.594)
	$p = 0.0003^{***}$	$p = 0.002^{**}$
HP vs. IN	-2.383 (0.598)	-2.159 (0.679)
	$p = 0.0005^{***}$	$p = 0.004^{**}$
Typical vs. Atypical	-0.269 (0.391)	-0.313 (0.466)
	$p = 0.499$	$p = 0.503$
LP vs. HP and IN \times Typical vs. Atypical	-0.137 (0.763)	-0.046 (0.777)
	$p = 0.859$	$p = 0.954$
HP vs. IN \times Typical vs. Atypical	2.193 (0.935)	2.036 (0.987)
	$p = 0.028^*$	$p = 0.050^{\dagger}$
Observations	1,584	720
Log Likelihood	-3,624.200	-1,628.706
Akaike Inf. Crit.	7,346.399	3,355.411
Bayesian Inf. Crit.	7,609.417	3,579.794

Note:

$^{\dagger} p < 0.1;$

$* p < .05;$

$** p < 0.01;$

$*** p < 0.001.$

Random effects estimates available in Supplemental Material S2.