Adaptation to Animacy Violations during Listening Comprehension

Megan A. Boudewyn [1,2] · Adam R. Blalock [1] · Debra L. Long [1] · Tamara Y. Swaab [1]

Abstract

The goal of this study was to examine adaptation to various types of animacy violations in cartoon-like stories. We measured the event-related potentials (ERPs) elicited by words at the beginning, middle, and end of four-sentence stories in order to examine adaptation over time to conflicts between stored word knowledge and context-derived meaning (specifically, to inanimate objects serving as main characters, as they might in a cartoon). The fourth and final sentence of each story contained a predicate that required either an animate or an inanimate subject. The results showed that listeners quickly adapted to stories in the Inanimate Noun conditions, consistent with previous research (Filik & Leuthold, 2008; Nieuwland & Van Berkum, 2006). They showed evidence of processing difficulty for animacy-requiring predicates in the Inanimate Noun conditions in Sentence 1, but the effect dissipated in Sentences 2 and 3. In Sentences 2 and 4, we measured ERPs at three critical points where it was possible to observe the influence of both context-based expectations and expectations from prior knowledge on processing. Overall, the pattern of results demonstrates how listeners flexibly adapt to unusual, conflict-ridden input, using previous context to generate expectations about upcoming input, but that current context is weighted appropriately in combination with expectations from background knowledge and prior language experience.

Keywords Discourse · Erp · Semantics

Introduction

Discourse comprehension requires that readers and listeners combine incoming words and phrases with background knowledge in order to generate a representation of linguistic input. This is a more complex process than it might appear, because it involves integrating new information into the developing discourse representation as it is received, suppressing irrelevant information that might be activated, and resolving ambiguities and conflicts that are present in language (Gernsbacher, 1996; Kintsch, 1988; Long, Johns, & Morris, 2006). One example that highlights the complexity of discourse is the distinction between what a word means and how it might be used to reference a specific entity in context. This distinction is not rare in language comprehension. Instead, it is critical to understanding language generally. For example, the stored representation of the word "banana" has a set of semantic features that constitute its meaning (e.g., [inanimate] [edible], etc.). In context, however, a word can have characteristics that are only true of a particular instance of that word (e.g., the context might specify that the banana is ripe, or that the banana is sliced). Thus, it is necessary to combine stored features of words with information pertaining to their use in context. What happens, however, when the way a word is used in context conflicts with stored word knowledge? For example, a child dressed in a banana costume might be referred to as "the banana," requiring the reader or listener to apply animate characteristics to a word normally considered to be inanimate for comprehension to be successful.

Previous studies have examined how listeners and readers interpret an incoming word when characteristics of the word's meaning are in conflict with characteristics that have been applied to a specific instance of that word in context. Nieuwland and Van Berkum (2006) presented participants with short narratives in which some stories portrayed inanimate entities (e.g., peanuts) behaving as if they were animate. Although this may seem to be an unusual circumstance contrived for experimental purposes, this type of conflict occurs frequently in natural language and does not seem to lead to comprehension difficulties. For example, cartoons regularly feature animacy violations but are easily understood (Filik &

✉ Megan A. Boudewyn
maboudewyn@ucdavis.edu

[1] University of California, Davis, Sacramento, CA, USA

[2] Imaging Research Center, UC Davis Medical Center, 4701 X Street, Sacramento, CA 95817, USA

Leuthold, 2008; Nieuwland & Van Berkum, 2006). In their study, Nieuwland and Van Berkum (2006) measured event-related potentials (ERPs) to critical words that did or did not violate animacy constraints (e.g., "The psychotherapist consoled the yacht/sailor," in a story about a distraught yacht/sailor). The most relevant ERP in this study was the N400, which is a negative-going deflection that peaks at around 400 ms after the onset of a word or similar meaningful stimulus and is modulated by the relative ease of semantic processing (Kuperberg, 2016; Kutas & Federmeier, 2011; Kutas & Van Petten, 1994; Kutas, Van Petten, & Kluender, 2006; Swaab, Ledoux, Camblin, & Boudewyn, 2012). Their results showed larger N400 amplitudes to critical words in the inanimate (e.g., yacht) relative to the animate condition (e.g., sailor) in the first sentence of the stories, reflecting that, at this point, words in this condition seem to be processed as animacy violations. However, the effect was attenuated as the stories continued, suggesting that processing difficulties related to what would be considered animacy violations outside of context were reduced when the context licensed the violations. Furthermore, in a second experiment, Nieuwland and van Berkum (2006) demonstrated that context functioned to "neutralize" these animacy violations and also reversed expectations. Specifically, they demonstrated that attributing an animate descriptor to an inanimate object was preferred to an inanimate descriptor, if that entity was serving as a character in the story context (e.g. preferring a peanut to be *in love* (smaller N400) rather than *salted* (larger N400) after several sentences in which the peanut was described as a cartoon character). As the authors stated, "It sometimes may be easier to integrate a lexical-semantic anomaly that is globally coherent than a "correct" phrase that is plausible with regard to real-world knowledge but that has nothing to do with the prior discourse" (Nieuwland & Van Berkum, 2006, p. 1109).

The results of this study and other similar studies were pivotal in demonstrating the power of context to influence how the brain processes incoming words (Filik & Leuthold, 2008; Nieuwland & Van Berkum, 2006; van Berkum, Zwitserlood, Hagoort, & Brown, 2003). The rapid use of context has been replicated in numerous studies, and it is now commonly accepted that expectations generated as a function of discourse exert a powerful and immediate influence on the processing of incoming words. Context may even modulate activation of the semantic representation of a word when that meaning is irrelevant in the context (Kutas et al., 2006; Swaab et al., 2012; Van Berkum, 2010). Many ERP studies have started to focus on disentangling the way in which specific elements of discourse are combined and weighted when interpreting words. For example, Xiang and Kuperberg (2015) demonstrated that expectations based on context can be quickly (within a sentence) reversed with a single semantic indicator that foreshadows unexpected input (e.g., a concessive connective like "even so…"), but that there may be some

processing costs to this reversal. Similarly, Boudewyn, Long, and Swaab (2015) showed that listeners quickly (within a couple of words) adjust expectations based on global context after encountering even a single word that does not fit with those expectations. Again, however, this flexibility has limits and comes with some processing costs. Focusing on the processing of emotional information in context, Delaney-Busch and Kuperberg (2013) found small N400 amplitudes to emotional words whether or not the emotional valence (pleasant or unpleasant) fit the context in which they appeared, suggesting that the influence of context on emotional word processing may be more limited than on nonemotional word processing (Delaney-Busch & Kuperberg, 2013). These examples serve to illustrate that the influence of context on processing of an incoming word depends on the specifics of both the information contained in that context and the word itself.

In the current study, our goal was to examine multiple aspects of discourse processing during the establishment of a complex context representation. To do this, we built on the paradigm developed by Nieuwland and Van Berkum (2006) that we discussed above, which featured cartoon-style stories in which inanimate entities (e.g., peppers) behaved as if they were animate. We widened the scope of the investigation to examine a variety of discourse processing elements, including thematic role assignment, pronominal reference, and lexical-semantics. Specifically, we examined the ERPs elicited by words at the beginning, middle, and end of short passages. ERPs provide a temporally precise window into the neural processing of incoming words; here, we focused on the N400 and P600, two well-studied ERP components that have been linked to language processing. As noted above, the N400 appears to index the relative ease of semantic processing, such that facilitation of lexico-semantic processing reduces the amplitude of the N400. Reduced N400 amplitude reflects facilitated word processing, which can occur as a function of the lexical properties of words, repetition and semantic priming, and how expected a word is given its meaning in the context (for thorough reviews of the N400, see Kuperberg, 2016; Kutas & Federmeier, 2011; Kutas & Van Petten, 1994; Kutas et al., 2006; Swaab et al., 2012). In contrast, the P600 is a positive-going deflection with a slightly later latency, usually approximately 600 ms after word onset. It is typically found when a word or similar meaningful stimulus requires some additional processing, such as with syntactically anomalous or difficult structures, or with conflict between syntax-dictated sentence meaning and meaning derived in error from other cues (e.g., the restructuring / conflict resolution / updating processes that might be required upon receiving "eat" in "at breakfast the eggs would eat" (Kim & Osterhout, 2005; Kolk & Chwilla, 2007; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Nakano, Saron, & Swaab, 2010). We used N400 amplitude to track the degree to which context

facilitated semantic processing, and the P600 as an indicator of the detection and/or resolution of conflict between structural expectations based on stored word meaning and those based on context. Below, we detail the experimental manipulation, hypotheses, and expected pattern of results.

We created fictional four-sentence passages in which we manipulated character type by using either inanimate or animate entities as main characters (see Table 1 for an example. The full set of stimuli can be found at the following website: swaab.faculty.ucdavis.edu). The first three sentences contained modifiers, predicates and pronouns that typically refer to animate nouns (e.g., "friendly," "loved," "he"). For the inanimate character conditions, this served to create conflict between the words' meaning (e.g., a pepper is, by definition, inanimate) and the context-appropriate meaning (e.g., the pepper takes on the thematic role of agent in a cartoon). Based on previous research, we expected that an accurate discourse representation in which the character was assigned the thematic role of agent would be well-established by the fourth and final sentence. At this point, we manipulated the predicates so that they required either animate or inanimate arguments, creating four story conditions (Table 1): Animate noun / Animate Predicate, Animate noun / Inanimate Predicate, Inanimate noun / Animate Predicate, Inanimate noun / Inanimate Predicate. Critically, up until this point at the end of the stories (i.e., in the first three sentences), all characters were paired with animate predicates.

Table 1. Sample stimulus set representing all four story types

Animate Noun Condition

S1: My friendly uncle <u>loved</u> taking trips to his favorite amusement park.

S2: My <u>uncle</u> was <u>riding</u> a roller coaster when his cart got stuck on the tracks.

S3: After hanging upside-down on a loop for several hours, <u>he</u> was rescued by park staff.

S4: My <u>uncle</u> was <u>panicked/spicy</u> and bought some cotton candy afterwards to feel better.

True/False: My uncle rode a horse.

Inanimate Noun Condition

S1: My friendly pepper <u>loved</u> taking trips to his favorite amusement park.

S2: My <u>pepper</u> was <u>riding</u> a roller coaster when his cart got stuck on the tracks.

S3: After hanging upside-down on a loop for several hours, <u>he</u> was rescued by park staff.

S4: My <u>pepper</u> was <u>panicked/spicy</u> and bought some cotton candy afterwards to feel better.

True/False: My pepper rode a horse.

S1-4 refer to the four sentences of the story; all stories were spoken at a natural pace and with natural pauses between sentences. In the first two sentences, an animate or inanimate character was introduced, always followed by an animate-preferring predicate. Sentence 3 always contained an animate-preferring pronoun. In sentence 4, the predicate was inanimate in half of the stories and animate in the other half of the stories. Underlined words indicate ERP time-lock points

Our overarching hypothesis was that the processing of incoming words, while initially guided by stored lexico-semantic information, would increasingly be guided by context as the stories unfolded. We examined ERPs across all four sentences of the spoken stories: (1) Predicates in Sentences 1 and 2; (2) Sentence 3 Pronouns; (3) Initial NPs in Sentences 2 and 4; and (4) Sentence 4 Predicates. We summarize our predicted results for each contrast below.

(1) Predicates in Sentences 1 and 2: We hypothesized that inanimate characters would be treated as animacy violations in Sentence 1, but that listeners would quickly adapt such that difficulties with thematic role assignment and lexical-semantic processing would be observed for the inanimate character condition in Sentence 1 but would dissipate by Sentence 2. Based on previous findings (Nakano et al., 2010), we expected that this would be indicated by larger N400 and/or P600 amplitudes in the inanimate condition compared to the animate condition for Sentence 1 Predicates. For Sentence 2, we predicted that adaptation to the inanimate characters would reduce the effect of animacy violation on the Predicates, relative to the first sentence.

(2) Sentence 3 Pronouns: We examined Sentence 3 pronouns in order to assess adaptation to animacy violations during the establishment of pronominal reference (e.g., "he" referring to a pepper). Previous studies have shown a larger P600 to pronouns that mismatch the gender of the antecedent (Hagoort & Brown, 1999; Osterhout, Bersick, & McLaughlin, 1997; Schmitt, Lamers, & Münte, 2002). Based on these findings, we expected that difficulties in assigning an inanimate noun to the pronoun (compared with an animate noun) would be reflected by increased P600 amplitude.

(3) Initial NPs in Sentences 2 and 4: Previous work has found reduced N400 amplitude for animate compared to inanimate sentence-initial NPs (Muralikrishnan, Schlesewsky, & Bornkessel-Schlesewsky, 2015; Nakano et al., 2010; Weckerly & Kutas, 1999). This has been interpreted as a preference (in English) for animate nouns in subject position (i.e., the combination of an expectation for subject-verb-object order and an expectation for sentence initial NPs to be assigned the thematic role of agent). Based on these previous findings, we expected that an animacy effect would be found on the N400 in Sentence 2 but may be reduced in Sentence 4 as a function of adaptation to story context.

(4) Sentence 4 Predicates: The predicates in Sentence 4 offered the opportunity to replicate Nieuwland and Van Berkum (2006) by comparing animate to inanimate predicates in the inanimate noun condition (e.g., the pepper was panicked vs. spicy). We predicted that by Sentence 4, comprehenders would have adapted to the inanimate

characters such that a larger N400 amplitude would be found to *inanimate* predicates compared to animate predicates in the inanimate noun condition (despite this being the reverse of what might be expected outside of context). In addition, we included stories in the animate noun condition to examine the interaction of prestored semantic knowledge (about animacy) and context-derived meaning. We have observed such an interaction on N400 amplitude in previous work (Boudewyn, Gordon, Long, Polse, & Swaab, 2012; Boudewyn et al., 2015; Camblin, Gordon, & Swaab, 2007). In the current study, however, adaptation to the inanimate characters over the course of the stories may lead to a similar pattern of N400 effects in the inanimate and animate noun conditions (i.e., similar N400 effect (*spicy>panicked*) following both *pepper* and *uncle*). Alternatively, to the extent that these factors interact to influence the processing of incoming words, we may see an interaction on the N400 amplitude that is time-locked to Sentence 4 predicates.

Methods

Participants

Twenty-four undergraduates from the University of California, Davis (18 females, 6 males; mean age 20.2 years, SD 2.2) gave informed consent and were compensated with course credit. All were right-handed, native speakers of English, with no reported hearing loss or psychiatric/neurological disorders.

Materials

Experimental materials consisted of 160 naturally spoken English short stories (divided over four lists, see below) in four conditions: Animate noun / Animate Predicate, Animate noun / Inanimate Predicate, Inanimate noun / Animate Predicate, Inanimate noun / Inanimate Predicate (see Table 1 for an examples. The full set of materials can be found on the following website: swaab.faculty.ucdavis.edu). An additional 40 filler stories were included to provide variation in the content and nature of the stories and to make the structure of the stories less predictable. Filler stories were four sentences long and described human characters engaged in various activities.

Stimuli were divided into 4 lists of 160 experimental and 40 filler stories, and every participant listened to 1 list (8 participants per list). Each list contained an equal number of stories per condition (40 per condition per list). For presentation purposes, each list was divided into 10 blocks with an equal number of trials, so that participants could have breaks between blocks. As noted in the *Introduction*, character type

was manipulated by using either animate or inanimate entities as main characters. Modifiers, predicates and nouns that typically refer to animate nouns were used throughout the stories to refer to both character types.

ERPs were measured to words at five critical points in the stories (see underlined words in Table 1 for examples): Sentence 1 Predicate, Sentence 2 Noun Phrase (NP), Sentence 2 Predicate, Sentence 3 Pronoun, Sentence 4 NP and Sentence 4 Predicate. When critical words varied across conditions, they were matched for duration and lexical frequency. Specifically, animate and inanimate nouns were matched for word duration (ms): Animate average = 525; Inanimate average = 511; $p = 0.263$) and lexical frequency (Kucera-Francis frequency: Animate average = 23.74; Inanimate average = 26.46; $p = 0.59$) (Francis & Kucera, 1982). Sentence 4 Predicates were also matched on word duration (ms): Animate average = 651 ms; Inanimate average = 653 ms; $p = 0.753$) and lexical frequency (Kucera-Francis frequency: Animate average = 5.26; Inanimate average = 3.76 ms; $p = 0.136$).

Stories were recorded with neutral intonation and speaking rate by a female speaker who was very experienced in producing stimuli for experiments. The speaker was instructed before the recording that the stories could contain semantic anomalies and that it was important that she produce all sentences with the same neutral intonation.[1] The first three sentences of each story were recorded together, one token for the animate stories and one token for the inanimate stories. The fourth sentence of each story was recorded separately for each condition. A subset of the recorded stories is available on the following website: swaab.faculty.ucdavis.edu. The stories were digitally recorded using a Schoeps MK2 microphone and Sound Devices USBPre A/D (44100 Hz, 32 bit). Speech onsets and offsets of each critical word were determined by visual inspection of the speech waveform and by listening to the words using speech editing software (Audacity, by Soundforge). For each Character Type condition (Animate noun and Inanimate noun), the first three sentences of each story were saved as one recording, so that the story context would be identical across Predicate conditions. The fourth sentence of each story was recorded separately for each condition, and spliced to the context recordings with a 1 second pause inserted between them (the average length of the natural pauses between sentences 1 and 2, and 2 and 3).

---

[1] We used a trained speaker who was instructed to match her prosody across conditions, but we cannot completely rule out the presence of subtle prosodic differences that could have provided an additional contextual cue to generate expectations about the upcoming input However, we do not think that prosodic differences can account for the pattern of results that we found in the present experiment, as we observed different effects of the animacy manipulation in the first, second, third and final sentences of the short passages. If the effects had been induced as a function of prosodic differences between animate and inanimate conditions, then we would have expected the effect of animacy to be very similar across the discourse.

Procedure

Participants sat in a comfortable chair, 100 cm from a computer monitor (Dell) in an electrically shielded sound-attenuating booth. Their task was to listen to the stories for comprehension and answer true/false comprehension questions that followed each story. Stimuli were presented through Beyer dynamic headphones (BeyerDynamic) using Presentation software (Neurobehavioral Systems). Trials began with a white fixation cross (1.5 cm x 1.5 cm) that appeared in the middle of a computer screen against a black background. The cross was presented 1,000 ms before the onset of the first word of each story and was present for the duration of the story. Participants were asked to keep their eyes fixated on the cross and to refrain from making eye movements and other body movements while the cross was present. This was done to minimize ocular and movement-related artifacts in the EEG signal. Condition-specific stimulus codes were sent at the onset of critical words to be used for later off-line averaging. One second after the offset of story final words, the fixation cross disappeared and was replaced by a true/false comprehension question (presented in Georgia font, size 12) about the content of the story that the participant had just heard (Table 1; "True or False: My uncle/pepper rode a horse"). Across items, comprehension questions focused on information that had been presented in one of the three preceding sentences. Participants were told that they could move their eyes while the true/false questions were on the screen. Responses were given via keyboard press: "true" ("z" on the keyboard) or "false" ("m" on the keyboard). After participants answered the true/false question, the fixation cross appeared on the screen again and the next story began playing after 1,500 ms.

EEG Recording and Data Processing

EEG was recorded from 29 tin electrodes mounted in an elastic cap following the 10-20 system (ElectroCap International). Additional electrodes were placed on the outer canthi and below the left eye to monitor for eye movements and blinks. The outer canthi electrodes were referenced to each other and the electrode under the left eye was referenced to channel FP1 (above the left eye). The right mastoid electrode was used as the recording reference. The left mastoid was also recorded for later off-line algebraic re-referencing. The EEG signal was amplified with band pass cutoffs at 0.01 and 30 Hz, and digitized online at a sampling rate of 250 Hz (Neuroscan Synamp II). Impedances were kept below 5 kΩ.

Data processing and analyses were performed offline using EEGLAB toolbox (Delorme & Makeig, 2004) with the ERPLAB plugin (Lopez-Calderon & Luck, 2014). ICA-based artifact correction was used to correct for eye blinks (Delorme & Makeig, 2004). After ICA correction was performed and prior to off-line averaging, all single-trial waveforms were screened for amplifier blocking, muscle artifacts, horizontal eye movements, and any remaining blinks over epochs of 1,200 ms, starting 200 ms before the onset of the critical words. Baseline correction was performed using the −200 to 0 ms prestimulus period. Artifact correction and rejection were performed using an automatic script-based process in combination with manual data inspection. Our threshold for inclusion in the experiment was a minimum of 20 artifact-free trials in all conditions and word positions. On average, 21.49% of trials were excluded following artifact rejection (range: 0-44%). This translates to an average of 32 or 64 artifact-free trials included per condition, depending on the word position. Average ERPs were computed over all artifact-free trials for each condition and participant. All ERPs were filtered with a Gaussian low-pass filter with a 25 Hz half-amplitude cutoff. Statistical analyses were conducted on the filtered data.

ERP Analysis

Our hypotheses focused on four comparisons of interest that included critical words in all four sentences of the spoken stories: (1) Predicates in Sentences 1 and 2; (2) Sentence 3 Pronouns; (3) Initial NPs in Sentences 2 and 4; and (4) Sentence 4 Predicates. For each comparison, we conducted separate repeated measures ANOVAs for the two ERP components of interest, the N400 (dependent measure: mean amplitude 300-500 ms post word onset) and the P600 (dependent measure: mean amplitude 500-800 ms post word onset). To capture the topographic distribution of each effect, we conducted both a Midline analysis that included Electrode as a variable (5 levels: Afz, Fz, Cz, Pz, POz), and a Lateral analysis that included the Hemisphere (2 levels: left, right) and Anteriority (2 levels: anterior, posterior) as variables to cover four scalp regions, representing the averaged data from the following electrode groups: left anterior: F7, F3, FC5, FC1; left posterior (CP5, CP1, T5, P3; right anterior: F4, F8, FC2, FC6; right posterior: CP2, CP6, P4, T6. All analyses included the within-participants variable of Character Type (Animate, Inanimate). For comparisons (1) and (2), we included the additional within-participants variable of Story Position (Earlier Sentence, Later Sentence), and for comparison (4) we included the additional within-participants variable of Predicate Type (Animate, Inanimate). Significant interactions were followed-up by simple effects comparisons. A Greenhouse-Geisser correction was used for all F tests with more than one degree of freedom in the numerator.

Results

Behavior

Participants answered an average of 93.6% of the comprehension questions correctly. A 2 Character Type (Animate, Inanimate) x 2 Predicate Type (Animate, Inanimate) repeated

measures ANOVA was performed on the comprehension data, which yielded no significant differences in performance across conditions: 95.6% (Inanimate/Animate), 93.9% (Inanimate/Inanimate), 93.7% (Animate/Animate), 90.0% (Animate/Inanimate).

ERPs

Predicates in Sentences 1 and 2: Early Establishment of Story Context

We observed a larger positive-going deflection for Sentence 1 Predicates in the Inanimate noun condition than in the Animate noun condition, which began in the N400 time window and lasted throughout the P600 time window; this effect was not present in Sentence 2. Results for this comparison are plotted in Figure 1 and summarized below.

N400 window. The omnibus rANOVA showed a significant main effect of Story Position (midline: $F(1,23) = 88.48$; $p < 0.0001$; lateral: $F(1,23) = 78.18$; $p < 0.0001$), which was characterized by a significant interaction with Character Type (midline: $F(1,23) = 4.49$; $p = 0.045$; lateral: $F(1,23) = 3.85$; $p = 0.062$). Simple effects comparisons revealed a significant effect of Character Type for Sentence 1 Predicates (midline: $F(1,23) = 4.72$; $p = 0.04$; lateral: $F(1,23) = 4.36$; $p = 0.048$). No significant effects of Character Type were found for Sentence 2 Predicates.

P600 window. The omnibus rANOVA showed significant main effects of Story Position (midline: $F(1,23) = 50.24$; $p < 0.0001$; lateral: $F(1,23) = 48.83$; $p < 0.0001$), and Character Type (midline: $F(1,23) = 4.46$; $p = 0.046$; lateral: $F(1,23) = 4.84$; $p = 0.038$). These effects were modulated by a significant interaction of Story Position by Character Type (midline: $F(1,23) = 4.41$; $p = 0.047$; lateral: $F(1,23) = 2.17$; $p = 0.154$).
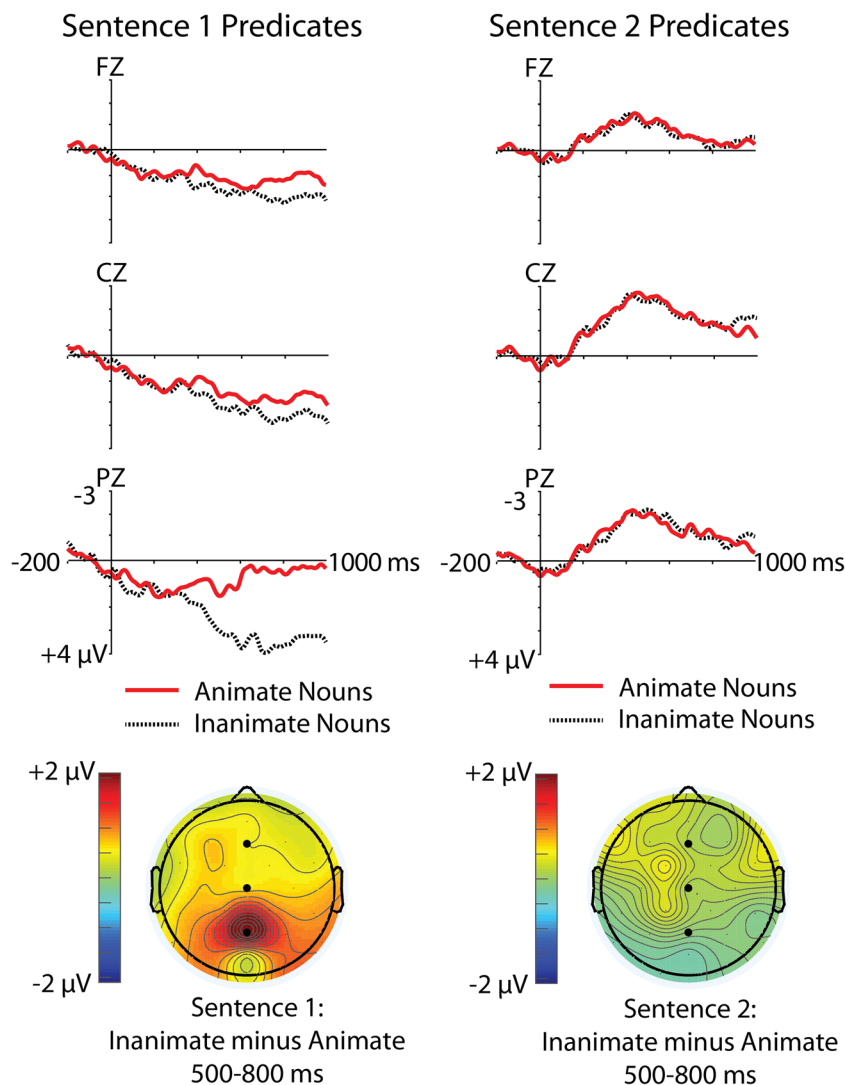

Fig. 1 ERPs to Sentence 1 and 2 Predicates. Waveforms are plotted for three midline electrodes (Fz, Cz, Pz). For reference, locations are shown as black dots on the topographic map at the bottom of this Figure, which shows the difference in mean amplitude between the Inanimate Noun and Animate Noun conditions in the P600 time window for all electrodes

Simple effects comparisons yielded a significant effect of Character Type for Sentence 1 Predicates (midline: $F(1,23)$ = 10.83; p = 0.003; lateral: $F(1,23)$ = 8.1; p = 0.009). No significant effects of Character Type were found for Sentence 2 Predicates.

Sentence 3 Pronouns: Pronominal Reference in Established Story Context

No significant effects were found in either the N400 or P600 time window. Waveforms are plotted in Figure 2.
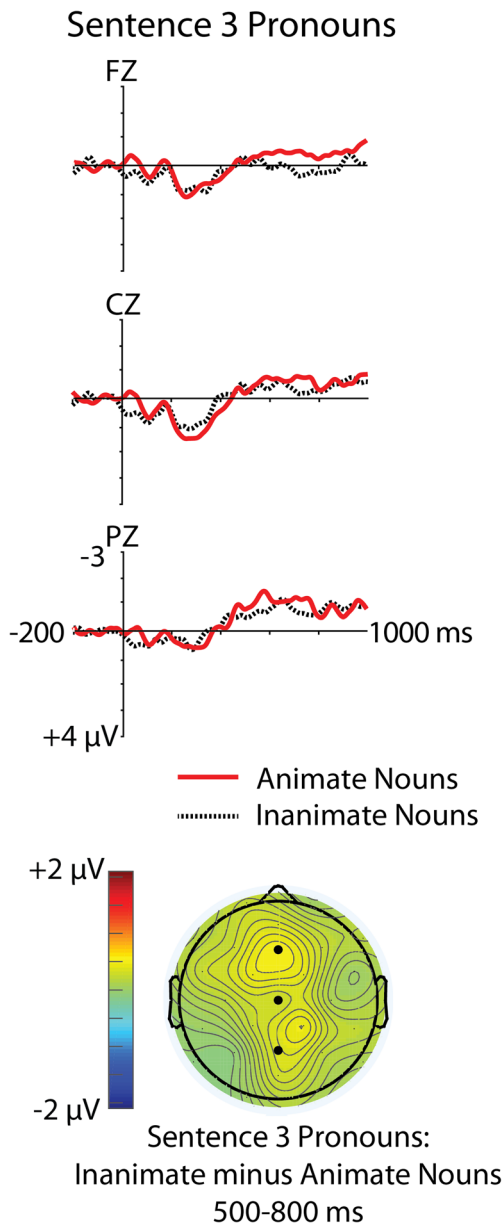

Fig. 2 ERPs to Sentence 3 Pronouns. Waveforms are plotted for three midline electrodes (Fz, Cz, Pz). For reference, locations are shown as black dots on the topographic map at the bottom of this Figure, which shows the difference in mean amplitude between the Inanimate Noun and Animate Noun conditions in the P600 time window for all electrodes

Initial NPs in Sentences 2 and 4: Sentence Initial NP in Established Story Context

We observed a larger N400 for both Sentence 2 NPs and Sentence 4 NPs in the Inanimate condition than in the Animate condition. Results for this comparison are plotted in Figure 3 and summarized below.

N400 window. The omnibus rANOVA showed a significant main effect of Character Type (midline: $F(1,23)$ = 10.16; p = 0.004; lateral: $F(1,23)$ = 6.29; p = 0.02).

P600 window. There were no significant effects for this comparison in this time window.

Sentence 4 Predicates: Conflict in Established Story Context

We observed larger N400 amplitudes for Animate Sentence 4 Predicates than for Inanimate Sentence 4 Predicates, irrespective of Character Type. This effect lingered into the P600 time window and was larger in the Animate noun condition than in the Inanimate noun condition. Importantly, simple effects comparisons of Character Type showed that this difference was driven by Inanimate Predicates: there was no significant difference between Animate Predicates in the Animate noun compared with the Inanimate noun condition. The results are plotted in Figure 4 and summarized below.

N400 window. The omnibus rANOVA showed significant main effects of Character Type (midline: $F(1,23)$ = 7.48; p = 0.012; lateral: $F(1,23)$ = 6.23; p = 0.02) and Predicate Type (midline: $F(1,23)$ = 18.49; p = 0.0003; lateral: $F(1,23)$=14.32; p = 0.001). These effects were modified by a marginally significant interaction (midline: $F(1,23)$ = 3.12; p = 0.091). Based on our *a priori* predictions and the significant interactions found in the later time window, we conducted simple effects comparisons to examine the interaction. This revealed no significant effects of Character Type for Animate Sentence 4 Predicates, and a significant main effect of Character Type for Inanimate Sentence 4 Predicates (midline: $F(1,23)$ = 13.04; p = 0.001; lateral: $F(1,23)$ = 9.1; p = 0.006). In addition, the analyses showed a significant effect of Predicate Type for both Animate Characters (midline: $F(1,23)$ = 17.5; p = 0.0004; lateral: $F(1,23)$ = 14.94; p = 0.001) and Inanimate Characters (midline: $F(1,23)$ = 5.31; p = 0.031; lateral: $F(1,23)$ = 4.89; p = 0.037).

P600 window. The omnibus rANOVA showed significant main effects of Character Type (midline: $F(1,23)$ = 13.94; p = 0.001; lateral: $F(1,23)$ = 13.47; p = 0.001) and Predicate Type (midline: $F(1,23)$ = 24.52; $p < 0.0001$; lateral: $F(1,23)$ = 26.74; $p < 0.0001$). The effects were modified by a significant interaction (midline: $F(1,23)$ = 20.21; p = 0.0002; lateral: $F(1,23)$ = 15.51; p = 0.001). As in the N400 time window, simple effects comparisons revealed no significant effects of Character Type for Animate Sentence 4 Predicates, and a significant main effect of Character Type for Inanimate Sentence 4 Predicates (midline: $F(1,23)$ = 30.94; $p < 0.0001$; lateral: $F(1,23)$ = 24.17; $p <$
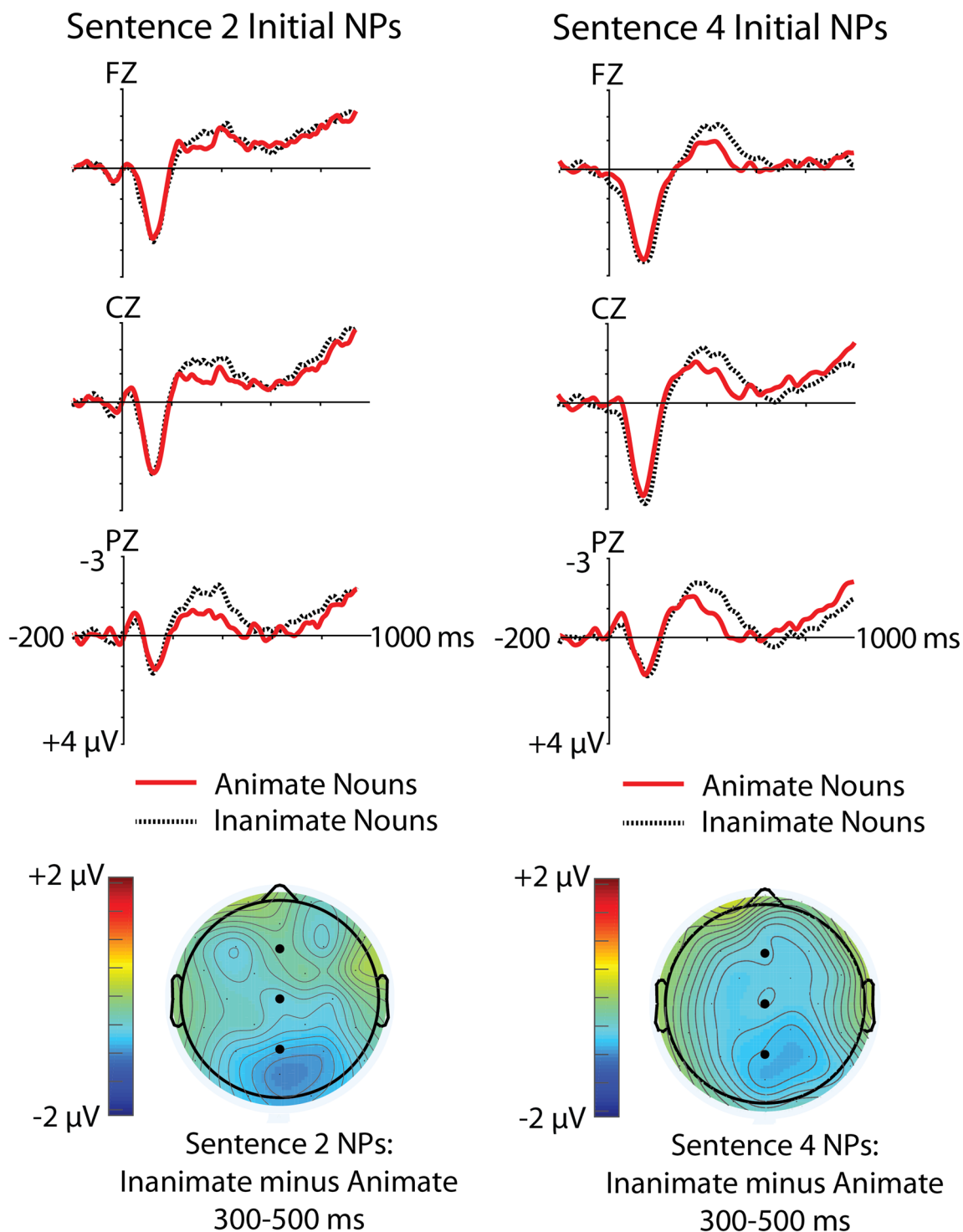
Fig. 3 ERPs to Sentence 2 and 4 Initial Noun Phrases (NPs). Waveforms are plotted for three midline electrodes (Fz, Cz, Pz). For reference, locations are shown as black dots on the topographic map at the bottom of this Figure, which shows the difference in mean amplitude between the Inanimate Noun and Animate Noun conditions in the N400 time window for all electrodes

0.0001). The analyses also showed a significant effect of Predicate Type for both Animate Characters (midline: $F_{(1,23)}$ = 36.04; $p < 0.0001$; lateral: $F_{(1,23)}$ = 36.17; $p < 0.0001$) and Inanimate Characters (midline: $F_{(1,23)}$ = 4.24; $p = 0.051$; lateral: $F_{(1,23)}$ = 4.56; $p = 0.044$).

Discussion

We used spoken fictional stories to investigate the time-course by which listeners adapt to conflict between stored word information and new information that they encounter in context,
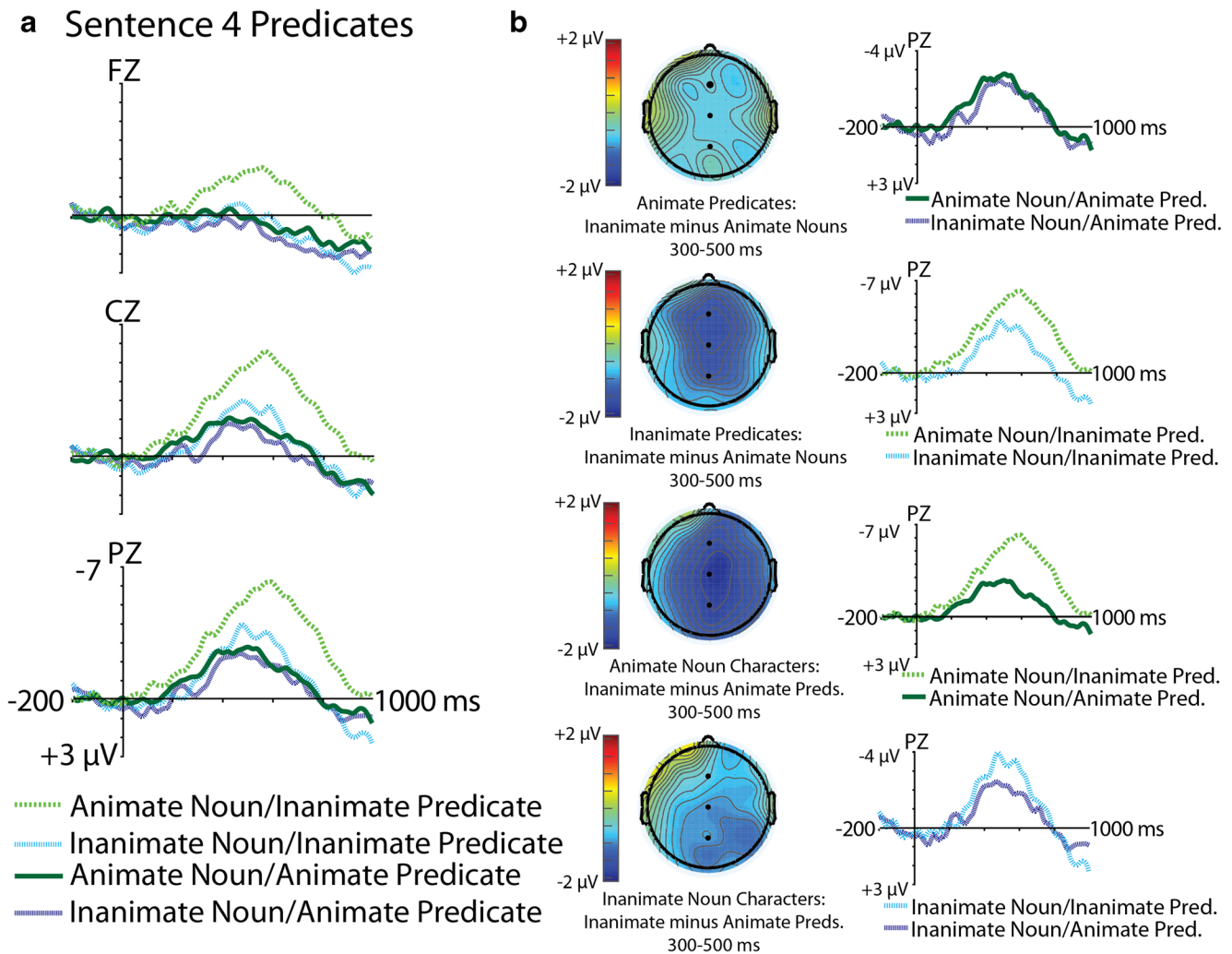
Fig. 4 ERPs to Sentence 4 Predicates. A Waveforms in all four conditions are plotted for three midline electrodes (Fz, Cz, Pz). For reference, locations are shown as black dots on the topographic maps in Panel B. **B** Simple effects comparisons for the Sentence 4 Predicates. Topographic maps of the difference between each set of conditions in the N400 time window and the corresponding waveforms at electrode Pz are shown, in order from top to bottom, for: Inanimate Noun/Animate Predicate vs. Animate Noun/Animate Predicate, Inanimate Noun/Inanimate Predicate vs. Animate Noun/Inanimate Predicate, Animate Noun/Inanimate Predicate vs. Animate Noun/Animate Predicate, and Inanimate Noun/Inanimate Predicate vs. Inanimate Noun/Animate Predicate

namely, conflict when inanimate nouns serve as main characters. We measured ERPs at multiple points in the stories (Figure 5), and found that listeners adapted quickly to the Inanimate noun condition, consistent with previous research (Filik & Leuthold, 2008; Nieuwland & Van Berkum, 2006). In Sentence 1, we found a larger P600 deflection in response to animate-agent-preferring predicates in the Inanimate noun condition compared with the animate noun condition. This effect dissipated by Sentence 2, and in Sentence 3, no difference was found when an animate-preferring pronoun (he or she) was used to refer to Inanimate nouns compared with Animate nouns. Finally, the pattern of N400 effects at the Sentence 4 predicates showed that for both the Inanimate noun and Animate noun conditions, predicates requiring an animate agent were preferred to inanimate-requiring predicates. Interestingly, this difference was driven by Inanimate

Predicates; there was no significant difference between Animate Predicates in the Animate noun compared with the Inanimate noun condition. Although we found evidence of adaptation to the conflict in stories featuring Inanimate characters, a larger N400 was elicited by sentence initial noun phrases in the Inanimate noun condition compared to the Animate noun condition.

Relevance for Accounts of the N400 and the Dynamic Generative Framework of Event Comprehension

Our results can be readily interpreted in the context of the dynamic generative framework (DGF) of event comprehension proposed by Kuperberg (2016). This framework draws on principles of constraint satisfaction approaches to language processing, extending them beyond lexical processing and
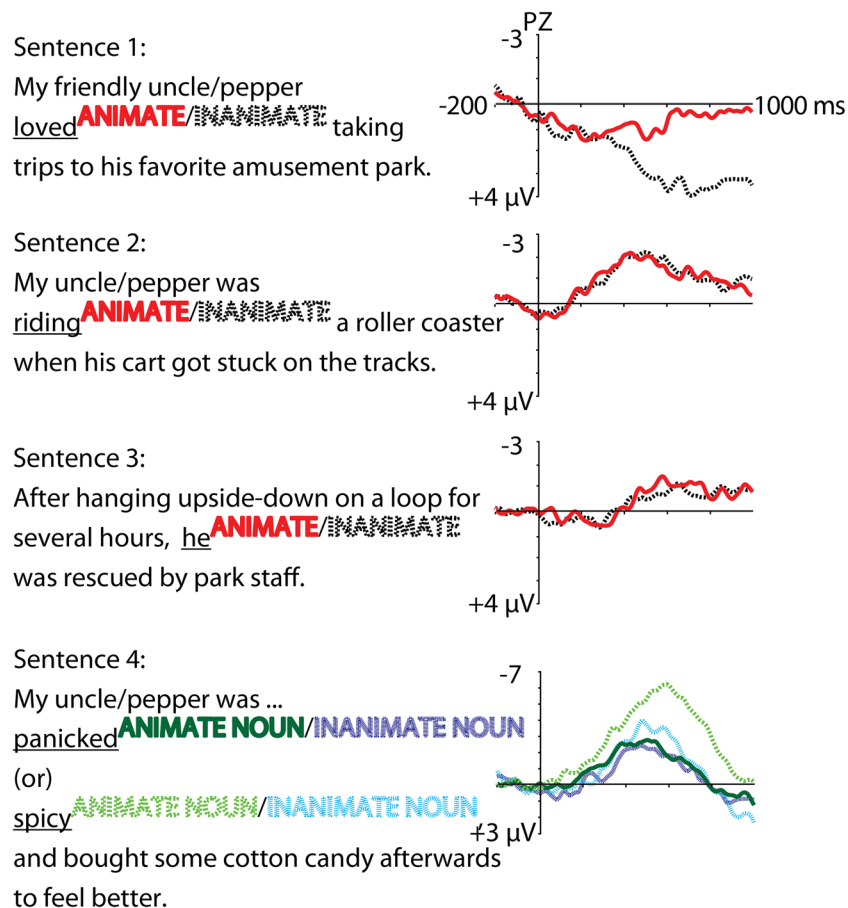
Fig. 5 Time-course of ERP context effects during the establishment of discourse context. A sample story is shown on the left side, and the critical words to which the ERPs at right were time-locked are underlined

incorporating elements of Bayesian computational modeling. The framework proposes that readers/listeners apply weighted combinations of inferences based on all sources that are available to them (background knowledge, lexical semantic information, syntactic structure, linear argument order, speaker reliability, environmental context, etc.) in order to anticipate and process incoming words. Critically, the framework specifies that comprehenders adapt and update their expectations based on the degree to which their expectations match the input that is actually received.

In the current study, we examined processing at multiple time-points in the stories, as "evidence" about the events being described accumulated. To adopt the terminology of the DGF, early in the stories, one might expect there to be maximal conflict between expectations based on a number of sources (e.g., background knowledge, word meaning) and the received input of an inanimate noun being assigned to the thematic role of *agent* and treated as a main character. This is reflected in the P600 effect observed in Sentence 1 Predicates, which might be interpreted to reflect the conflict and/or resolution between the expectations generated by different sources of information. However, by Sentence 2, substantial evidence has accumulated to suggest an event in which a cartoon-style

character is assigned to the thematic role of agent; thus, in Sentence 2, we saw no difference between conditions. Similarly, in Sentence 3, we found no difference between conditions at the pronouns. This suggests that nouns in both the animate and inanimate conditions served equally well as antecedents for the pronouns at this point in the stories.

Interestingly, despite the pattern of data in Sentences 2 and 3 being consistent with a fully-established context in which inanimate nouns were easily integrated into the representation as main characters, we found that sentence-initial NPs in the Inanimate noun condition elicited a larger N400 than in the Animate noun condition in both Sentence 2 and Sentence 4. As noted above, English's preference for a subject-verb-object (SVO) word order combined with the preference for animate nouns to be assigned the thematic role of agent have led to a preference for animate nouns over inanimate nouns in the subject position of the sentence (although it should be noted that inanimate nouns in subject position are grammatically licensed and do not constitute violations). A similar N400 effect has been found in previous studies, such that amplitude was increased for inanimate compared to animate nouns in subject positions (Muralikrishnan et al., 2015; Nakano et al., 2010; Weckerly & Kutas, 1999). Thus, it is not surprising that

we observed such an effect in the current study. However, it is of note because, in other contrasts, we observed adaptation to the inanimate noun condition such that it became more similar or even equivalent to the animate noun condition over time. That this was not the case for the sentence initial NP effect may suggest that, although the context provided by the short stories in this paradigm was enough to facilitate the processing of other discourse elements in the inanimate condition (as evidenced by the ERP results for the other contrasts of interest), it did not outweigh the preference in English for animate nouns in the subject position (as we did not observe a change in the animacy effect to the sentence-initial NPs over the course of the story in the current study). Although it is possible that such an adaptation effect might be observed given suitable context (for example, task conditions in which listeners encounter repeated examples of passive voice constructions containing sentence-initial NPs that are inanimate), the effect at the Sentence 2 and 4 NPs demonstrates an important point. Adaptation to language context is a powerful mechanism that gives comprehenders the flexibility to understand and even expect conflicts between stored word meanings and context, but comprehenders do not abandon all of their preexisting expectations about language based on a given context. Instead, some aspects of discourse processing may be more resistant to the influence of context than others (more on this below).

Finally, the Sentence 4 predicates reveal the interplay of expectations about incoming words in the stories that come from different sources. As the DGF would predict, incoming words fit with the preceding context to different degrees. For example, four sentences of story context appears to have established Inanimate noun characters sufficiently well as to render Animate predicates that modify them to be no different than Animate predicates that modify Animate noun characters. However, for Inanimate predicates, the difference in N400 amplitude between the Animate noun and Inanimate noun conditions suggests that at least some degree of expectation for inanimate-compatible features must have been generated in the Inanimate noun condition, as the processing of Inanimate predicates in this condition was facilitated compared to the Animate noun condition (which presumably generated no expectations for an inanimate predicate). In other words, this contrast shows a small effect of "local" context, such that "spicy" benefits from recently encountering "pepper" compared to "uncle." Despite this, the preponderance of the expectations in the Inanimate noun condition appear to have been in favor of an animate predicate, as the processing of Animate predicates was equally facilitated in both conditions. This pattern of results serves to illustrate how processing is quickly influenced by a mix of expectations coming from different cues, in real-time.

Discourse is Dynamic

In ERP experiments, the focus is, by definition, on processing at a given moment in time. Most ERP experiments in psycholinguistics involve conditions in which the context before a critical moment in time (usually a certain word) is manipulated. This creates experimental designs in which it is meaningful to discuss words that are "discourse congruent" or "discourse incongruent" with respect to prior information. Although this can be a convenient and useful way of discussing these manipulations, "real" context unfolds and changes over time, such that readers/listeners must continually adapt to incoming information. New information must be integrated into the developing representation, and the updated representation must be used in integrating and updating incoming information, and so on. Although we discuss the "establishment" of context and demonstrate its rapid influence on word processing, we wish to emphasize that our conclusion is not that it takes approximately one sentence to construct a representation of context that will guide all future processing. Instead, the development of a discourse representation is a dynamic process. The specific amount of time that it requires to process enough context to generate accurate expectations will vary as a function of factors like background knowledge (e.g. evoking a restaurant scenario may require less context than evoking an unusual or novel scenario) (Kuperberg, 2016; Sanford & Garrod, 1998). Most importantly, expectations will change as new information is accommodated.

Finally, our results highlight the time-course of context-based adaptation and demonstrate that adaptation occurs quickly even for relatively extreme examples of conflict, as in cartoon stories featuring inanimate objects as characters. This kind of flexibility is necessary in order to understand conflict in everyday language: language users apply known words to new situations, use context to resolve lexical ambiguities and to understand metaphors, all of which might be characterized as conflicts to varying degrees. Our findings also indicate that brief exposure to an unusual context does not suffice to overturn expectations based on strongly established expectations, such as expectations for animate over inanimate nouns in subject positions in a sentence. Although listeners immediately adjust their expectations about upcoming events and words based on context, this does not necessarily extend to the adjustment of expectations about how their language might structure this content.

Compliance with ethical standards

References

Boudewyn, M. A., Gordon, P. C., Long, D., Polse, L., & Swaab, T. Y. (2012). Does Discourse Congruence Influence Spoken Language Comprehension before Lexical Association? Evidence from Event-Related Potentials. *Lang Cogn Process, 27*(5), 698-733. doi: https://doi.org/10.1080/01690965.2011.577980

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 1-18.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *J Mem Lang, 56*(1), 103-128. doi: https://doi.org/10.1016/j.jml.2006.07.005

Delaney-Busch, N., & Kuperberg, G. (2013). Friendly drug-dealers and terrifying puppies: Affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive, Affective, & Behavioral Neuroscience*, 1-18. doi: https://doi.org/10.3758/s13415-013-0159-5

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods, 134*(1), 9-21. doi: https://doi.org/10.1016/j.jneumeth.2003.10.009

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: evidence from the N400. *Psychophysiology, 45*(4), 554-558. doi: https://doi.org/10.1111/j.1469-8986.2008.00656.x

Francis, W., & Kucera, H. (1982). Frequency analysis of English usage.

Gernsbacher, M. A. (1996). The structure-building framework: What it is, what it might also be, and why. *Models of understanding text*, 289-311.

Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of psycholinguistic research, 28*(6), 715-728.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of memory and language, 52*(2), 205-225. doi: https://doi.org/10.1016/j.jml.2004.10.002

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review, 95*(2), 163.

Kolk, H. H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain Lang, 100*(3), 257-261. doi: https://doi.org/10.1016/j.bandl.2006.07.006

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and language, 85*(1), 1-36.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience, 31*(5), 602-616.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117-129. doi: https://doi.org/10.1016/s0926-6410(03)00086-7

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol, 62*, 621-647. doi: https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. *Handbook of psycholinguistics*, 83-143.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). Handbook of psycholinguistics , ed. by M. Traxler and MA Gernsbacher, 659–724: London, UK: Elsevier.

Long, D. L., Johns, C., & Morris, P. (2006). Comprehension ability in mature readers. *Handbook of psycholinguistics, 2*, 801-833.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front Hum Neurosci, 8*, 213. doi: https://doi.org/10.3389/fnhum.2014.00213

Muralikrishnan, R., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). Animacy-based predictions in language comprehension are robust: contextual cues modulate but do not nullify them. *Brain research, 1608*, 108-137.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience, 22*(12), 2886-2898.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience, 18*(7), 1098-1111.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition, 25*(3), 273-285.

Sanford, A. J., & Garrod, S. C. (1998). The role of scenario mapping in text comprehension. *Discourse processes, 26*(2-3), 159-190.

Schmitt, B. M., Lamers, M., & Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research, 14*(3), 333-346.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press.

Van Berkum, J. J. A. (2010). The neuropragmatics of "simple" utterance comprehension. An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantic and Pragmatics: From Experiment to theory*.

van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*(3), 701-718. doi: https://doi.org/10.1016/s0926-6410(03)00196-4

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology, 36*(5), 559-570.

Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience, 30*(6), 648-672.