

An electrophysiological investigation of co-referential processes in visual narrative comprehension



Cas W. Coopmans ^{a,b,*}, Neil Cohn ^c

^a Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, the Netherlands

^b Centre for Language Studies, Radboud University, P.O. Box 9103, 6500 HD Nijmegen, the Netherlands

^c Tilburg University, P.O. Box 90153, 5000 LE Tilburg, the Netherlands

ARTICLE INFO

Keywords:
 Comics
 Anaphora
 ERPs
 Oscillations
 Nref
 Alpha

ABSTRACT

Visual narratives make use of various means to convey referential and co-referential meaning, so comprehenders must recognize that different depictions across sequential images represent the same character(s). In this study, we investigated how the order in which different types of panels in visual sequences are presented affects how the unfolding narrative is comprehended. Participants viewed short comic strips while their electroencephalogram (EEG) was recorded. We analyzed evoked and induced EEG activity elicited by both full panels (showing a full character) and refiner panels (showing only a zoom of that full panel), and took into account whether they preceded or followed the panel to which they were co-referentially related (i.e., were cataphoric or anaphoric). We found that full panels elicited both larger N300 amplitude and increased gamma-band power compared to refiner panels. Anaphoric panels elicited a sustained negativity compared to cataphoric panels, which appeared to be sensitive to the referential status of the anaphoric panel. In the time-frequency domain, anaphoric panels elicited reduced 8–12 Hz alpha power and increased 45–65 Hz gamma-band power compared to cataphoric panels. These findings are consistent with models in which the processes involved in visual narrative comprehension partially overlap with those in language comprehension.

1. Introduction

A growing literature on the cognition of visual narratives such as comics has shown their similarity to language in terms of both structural properties and processing demands. For instance, sequential visual narratives share basic features with language (Cohn, 2013, 2014; Cohn et al., 2012), models of visual narrative comprehension show strong similarities to theories of the comprehension of text-based narratives (Cohn, 2020a; Loschky et al., 2020), and manipulations of visual sequencing elicit neurocognitive responses similar to those found in sentence processing (Cohn et al., 2012; Cohn et al., 2014; Cohn and Kutas, 2015, 2017; see Cohn, 2020a for a recent overview). In particular, event-related potential (ERP) research has consistently shown that electrophysiological responses to linguistic manipulations of both semantics and grammar are also elicited by visual narratives. These ERP findings therefore suggest the possibility of domain-general mechanisms operating across the sequential meaning-making of both pictorial and verbal modalities (Coderre et al., 2020; Cohn, 2020a; Cohn et al., 2012; Sitnikova et al., 2008).

While ERPs have proven a useful measurement of neurocognitive processing across domains, they are a relatively coarse-grained measure, potentially masking more subtle differences between the processes involved. A complementary view of the same electrophysiological signals can be provided by time-frequency analysis of oscillatory power, which allows for a decomposition of the involved processes via modulations in time as well as frequency (Bastiaansen et al., 2013). This method therefore allows us to make more fine-grained comparisons between the functional properties of the neurocognitive systems responsible for comprehending language and visual narratives. In the present study, we ask (1) whether time-frequency analysis of oscillatory power provides insights into the neural correlates of visual narrative comprehension complementary to the ERP literature, and (2) whether and how oscillatory brain activity elicited by manipulations of visual narratives is similar to that elicited by analogous manipulations in the language domain.

* Corresponding author. Max Planck Institute for Psycholinguistics Wundtlaan 1, 6525 XD Nijmegen, the Netherlands.

E-mail address: cas.coopmans@mpi.nl (C.W. Coopmans).

1.1. A cognitive introduction to visual narratives

The basic unit of a visual narrative sequence is a panel, which is an encapsulated image unit that usually depicts referential and event-based information. At the most basic level, visual sequences require a comprehenders to connect the semantic and referential information across panels. The need for referential continuity in visual narratives has been termed the “continuity constraint”, which pushes comprehenders to recognize that representations across panels—despite potentially using physically different arrangements of lines—depict the same character(s) and objects repeatedly across frames, thus establishing co-referential links across panels (Cohn, 2020b; Magliano et al., 2016). To accomplish continuity, the processing of visual narratives involves assessing the basic semantics of images, whose new information is incorporated into a growing situation model of the unfolding narrative (van Dijk and Kintsch, 1983). The situation model can be used to make inferences about the relationships between events depicted in (non-)adjacent panels (Cohn, 2020a), and is updated with the addition of referential and event-based information. This access-updating cycle iteratively occurs at each unit of a sequence (see Brouwer and Hoeks, 2013; Cohn and Foulsham, 2020).

To illustrate which aspects of visual sequences trigger these cognitive processes, consider the structure of the visual narrative in Fig. 1. This comic strip shows an event sequence where Lucy reaches into a bag and tosses a piece of candy towards Charlie Brown, only to have Snoopy dive between them and grab it out of the air. These events are segmented into panels in a clear narrative structure: it begins with a set-up (Establisher), before initiating an event (Initial), which climaxes (Peak) and then has an aftermath (Release). Within the Initial, modifiers allow for further complexity. Both Lucy and Charlie here belong to their own panels, structurally conjoining two Initials into a single Initial phrase, while leaving the overall environment to be inferred (“e”). Moreover, a panel zooming in on Lucy’s hand and the candy appears in a “refiner”, which is separated at a distance from its “head” in the full depiction of Lucy in the first panel of the Initial.

Refiners have been argued to operate analogously to anaphoric elements in verbal language, indexing a semantically richer antecedent (Cohn et al., in prep.). Like the more general co-referential continuity constraint, refiners involve a more focal form of co-reference by connecting to a previously presented antecedent with the exact same features, only with a broader representation (i.e., framing of a whole character or scene, rather than the zoom of the refiner).

Considering these functional properties, one can think about the contrast between full panels and refiners as being similar to the contrast

between proper names and pronouns. Both full panels and proper names are referentially independent; they can stand on their own and need not be associated with another element in order to be appropriately interpreted. Both refiner panels and pronouns, instead, are semantically poor and become associated with another element in the discourse to receive full interpretation (i.e., they are referentially dependent). Moreover, refiners and pronouns can both precede and follow the element to which they are co-referentially related. In the latter case (i.e., proper name_i – pronoun_j) they are called “anaphoric”, in the former case (i.e., pronoun_j – proper name_i) they are called “cataphoric”. It appears that people’s ordering preferences with respect to these elements are similar in sentences and visual sequences: people prefer refiners to follow rather than precede full panels (Cohn et al., in prep.), and they prefer anaphoric over cataphoric co-referential pronouns (Filik and Sanford, 2008; Kennison et al., 2009). These preferences align with the results of recent corpus analyses. Initial analyses of a corpus of 90 comics showed differences in the use of anaphoric refiners across cultures (Cohn, 2019), and a subsequent unpublished analysis of an expanded corpus of 300+ comics has suggested greater frequency of anaphoric to cataphoric refiners. Nevertheless, the mapping between the (co-)referential units in verbal and visual language is not one-to-one. An important difference is that refiners focus attention on one aspect of the full panel, so they can shift the local topic of the narrative (Cohn, 2013; Foulsham and Cohn, 2021). Pronouns, instead, are mostly used when the referent is already in discourse focus; they sustain the topic rather than shifting it (Gordon and Hendrick, 1998; Marslen-Wilson et al., 1982; Vonk et al., 1992).

In the present study, we investigate referential and co-referential processes in visual narrative comprehension via their effects on event-related EEG activity. We manipulate panel type and sequence order in a crossed two-by-two design, in which we look at semantic and referential processing by comparing full panels to refiner panels, and at co-referential processing by comparing anaphoric panels to cataphoric panels.

1.2. ERP correlates of visual narrative comprehension

Manipulations of visual narrative sequences have been associated with several ERP effects, most notably the N300, N400, Late Positive Component (LPC) and Nref. The N300 is a frontally distributed negative potential which is elicited by pictorial stimuli and shows sensitivity to semantic congruity: semantically incongruent or unrelated pictures in both pictorial contexts (McPherson and Holcomb, 1999; West and Holcomb, 2002) and written sentence contexts (Federmeier and Kutas, 2001) elicit a larger N300 than semantically congruent or related

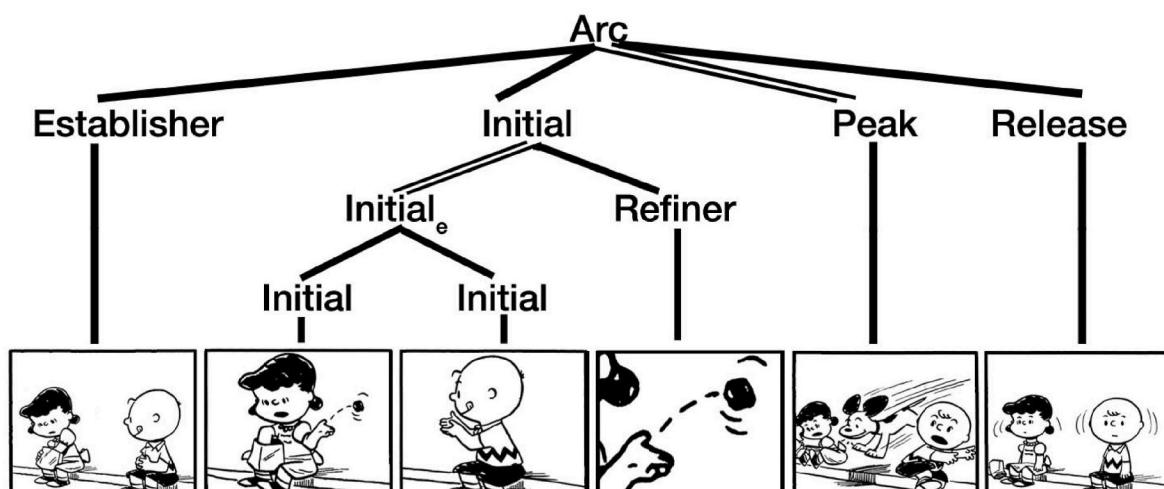


Fig. 1. Visual narrative sequence using a distance dependency between a fully depicted character (Lucy in an Initial) and a “refiner”, which highlights the informative aspects of its corresponding full panel. Peanuts is © Peanuts Worldwide LLC.

pictures. Semantic incongruity is not a necessary prerequisite, however, as full panels elicit a larger N300 than refiner panels even in contexts that are fully congruous (Cohn and Foulsham, 2020). As the N300 component is often associated with semantic identification or categorization of visual objects (Draschkow et al., 2018; Hamm et al., 2002; McPherson and Holcomb, 1999), the N300 difference between refiners and full panels suggests that these panel types differ in the ease with which their content can be identified and categorized. The N300 is not commonly found in response to linguistic manipulations, arguably because words do not have to be ‘identified’ in a similar way (see West and Holcomb, 2002 for discussion).

The N400, in contrast, does respond quite similarly to experimental manipulations in verbal and visual contexts. As in language, the pictorially elicited N400 shows strong sensitivity to expectancy and semantic congruity, with unexpected and incongruous visual events eliciting an increased N400 (Coderre et al., 2020; Cohn et al., 2012; Federmeier and Kutas, 2001; McPherson and Holcomb, 1999; Reid and Striano, 2008; West and Holcomb, 2002; Willems et al., 2008). In both the verbal and the visual domain, the N400 component is a negative deflection that peaks around 400 ms and is viewed as indexing access of information in semantic memory (Kutas and Federmeier, 2011). Moreover, it has been argued that N400 modulations in visual sequences reflect comprehenders’ predictions about the way an incoming image will relate to the prior narrative context (Coderre et al., 2020; Cohn, 2020a). Some of these comprehension processes might be shared with the neurocognitive processes underlying (predictive) language processing (Cohn, 2020a). For instance, a recent study found that visual narrative sequences modulate the N400 effect to written words that replace Peak panels (Manfredi et al., 2017). While this similarity suggests that language and visual narrative comprehension rely on cross-modal semantic resources, the spatio-temporal characteristics of these N400 effects do show notable differences. The N400 in language usually peaks between 300 and 500 ms, but the N400 elicited by pictorial stimuli is characterized by a more prolonged time course (Cohn et al., 2012; West and Holcomb, 2002). And while the language-related N400 effect has a centro-posterior topography, pictorial N400 effects are more frontally distributed (Cohn et al., 2012; Federmeier and Kutas, 2001; Ganis et al., 1996; West and Holcomb, 2002), indicating that the semantic networks involved are at least partially non-overlapping (see Sitnikova et al., 2008 for discussion).

When the semantic information corresponding to an image has been accessed, it must be integrated into the mental representation of the preceding context. This ‘updating’ process involves integration, reanalysis, and/or reorganization of prior information established by the preceding context, depending on the strength of required update (Cohn, 2020a). Updating is an ongoing process, often associated with the positive-going LPC in ERP studies on language processing (Brouwer and Hoeks, 2013; Burkhardt, 2006; Coopmans and Nieuwland, 2020; Delogu et al., 2019). In visual narrative studies, it has been shown that the LPC is sensitive to the amount of updating required to integrate incoming information with the discourse model. For instance, when panels are embedded in a sequence of zoom panels, they elicit a stronger LPC than when they are embedded in a sequence of full panels (Cohn and Foulsham, 2020). When the information in the context has been restricted in its framing, as in the case of a sequence of zoom panels, any incoming information requires a strong update to the discourse model (Cohn and Foulsham, 2020). LPC modulations also occur when the contextual framing of discourse information is completely coherent. Any shift in character or event state, whether congruous or incongruous, triggers an increase in LPC amplitude, suggesting that updating is an ongoing process that does not require anomalies (Cohn and Kutas, 2015, 2017).

Of particular relevance to the current study are ERP effects associated with (co-)referential processing. Referentially ambiguous anaphors elicit a frontally distributed sustained negativity compared to non-ambiguous anaphors (for reviews, see Nieuwland and van Berkum, 2008; van Berkum et al., 2007). This ERP effect, called the Nref, can be

elicited by different types of anaphoric expressions, including noun phrases (e.g., van Berkum et al., 1999; Nieuwland et al., 2007), pronouns (e.g., Nieuwland and van Berkum, 2006; Nieuwland, 2014), and proper names (e.g., Coopmans and Nieuwland, 2020). It has been argued to reflect ‘true’ referential ambiguity at the discourse level (Nieuwland et al., 2007; van Berkum, 2009), but referential ambiguity is not a necessary requirement to elicit an Nref effect (Coopmans and Nieuwland, 2020; Karimi et al., 2018). In visual narrative sequences, (unambiguous) panels following an event that is omitted from a scene also elicit frontally distributed negativities (Cohn and Kutas, 2015), but it is unclear whether these are related to the Nref effect or instead reflect more general inference processes. In the remainder of the text, we will refer to the sustained frontal negativity as Nref, but we remain agnostic about whether they reflect the same effect as the Nref effects found in linguistic studies of co-referential processing.

1.3. A time-frequency perspective on visual narratives

ERPs have proven a useful measure of neurocognitive processing, both of language and visual narratives. However, as ERPs are calculated by averaging the time-locked EEG signal across a large number of trials, they contain only ‘evoked’ activity that is strictly phase-locked to the external event of interest. Non-stationary activity, whose phase varies from trial to trial, will be reduced by this procedure and can therefore not be captured by ERP analysis (Bastiaansen et al., 2013; Tallon-Baudry and Bertrand, 1999). The modulation of this ongoing ‘oscillatory’ activity reflects patterns of (de)synchronization of neuronal networks, which are thought to be related to the dynamic coupling or uncoupling of functional systems (Buzsáki and Draguhn, 2004; Engel et al., 2001; Singer, 2011; Varela et al., 2001). Analysis of oscillatory activity is potentially fruitful for the study of visual narrative comprehension because of its constantly changing demands on integration processes. Here, we study this activity via time-frequency analysis of power, which can be used to assess synchronization changes in local neuronal assemblies. As this method can quantify event-related activity that is not strictly phase-locked to the event of interest, it provides a view on event-related electrophysiological activity that is complementary to the ERP approach.

To date, there are no M/EEG studies that have used time-frequency analysis to study the neural correlates of visual narrative comprehension. In the following sections we will therefore provide a brief overview of psycholinguistic studies that have used experimental manipulations involving semantic congruity and referential ambiguity. We consider the results of this literature relevant for visual narrative comprehension because of the aforementioned literature on ERPs across modalities, and because the oscillatory effects likely reflect domain-general processes underlying the brain’s attempt to deal with these manipulations rather than specific linguistic processes per se (Bastiaansen et al., 2013; Prys-tauka and Lewis, 2019).

In a recent EEG study on the comprehension of anaphoric proper names, Coopmans and Nieuwland (2020) presented participants with two-sentence discourse stories in which the critical proper name in the second sentence was either repeated or new, and either semantically coherent or incoherent with the first sentence. In line with the view that the successful comprehension of an anaphoric expression requires the initial activation and the subsequent integration of the antecedent into the discourse representation (Nieuwland and Martin, 2017), time-frequency analysis showed two patterns of oscillatory synchronization. Repeated names, whose antecedent had to be activated in working memory, elicited stronger theta power than new names, suggesting that theta activity indexed referent activation. Integration, on the other hand, was associated with gamma-band activity: discourse-coherent proper names elicited stronger gamma-band synchronization than discourse-incoherent proper names (see also Nieuwland and Martin, 2017). Both theta and gamma effects are commonly reported in studies of language comprehension, in particular in response

to manipulations of memory retrieval, predictability and semantic integration.

Theta effects (3–7 Hz) in language comprehension research are often associated with the retrieval of lexical-semantic information from long-term memory (Bastiaansen et al., 2005, 2008; Bastiaansen and Hagoort, 2015; Piai et al., 2016). Bastiaansen et al. (2005), for instance, found stronger theta power for content words than for function words, plausibly because content words are semantically richer and therefore activate the (lexico-)semantic network more strongly. In the context of anaphor processing, theta activity has been linked to the reactivation of linguistic information from working memory (Coopmans and Nieuwland, 2020; Heine et al., 2006; Meyer et al., 2015; Nieuwland et al., 2019). The direction of this effect, however, is not consistent across studies, which could be due to differences in both the types of anaphoric expressions and the experimental manipulations used to probe the working memory system (i.e., repeated vs. new, ambiguous vs. unambiguous, difficult vs. easy to reactivate). Despite the mixed pattern of results, these effects are relevant for the current study, particularly in light of the idea that the working memory resources recruited for language and visual narrative comprehension are partially shared (Magliano et al., 2016).

Gamma-band activity (>30 Hz) in response to linguistic manipulations is often associated with semantic integration (Bastiaansen and Hagoort, 2003; Coopmans and Nieuwland, 2020; Fedorenko et al., 2016; Nieuwland and Martin, 2017; Peña and Melloni, 2012; Rommers et al., 2013) and predictability (Lewis et al., 2015; Lewis and Bastiaansen, 2015; Wang et al., 2012; Wang et al., 2018). Semantically coherent and predictable words elicit stronger gamma-band power than words that are unpredictable and/or difficult to integrate into the semantic representation of the sentence (e.g., Coopmans and Nieuwland, 2020; Lewis et al., 2015; Rommers et al., 2013; Wang et al., 2012). These studies used purely linguistic input, but semantic integration of audio-visually presented information is also reflected in modulations in gamma-band activity. Willems et al. (2008) found decreased gamma power in response to pictures that were semantically incongruous with the preceding linguistic context, possibly indexing the involvement of an amodal semantic system, in line with the ERP literature (Coderre et al., 2020; Sitnikova et al., 2008).

While Coopmans and Nieuwland (2020) reported no modulations in the alpha band (8–12 Hz), alpha effects were reported in another EEG study on referential processing. Boudewyn et al. (2015) showed that the size of the Nref effect in response to ambiguous vs. unambiguous anaphors was predicted by oscillatory activity that occurred earlier, when possible antecedents of the anaphor were introduced into the discourse. Specifically, alpha power during the presentation of these discourse referents was negatively correlated with the amplitude of the Nref elicited by the ambiguous anaphor. Boudewyn et al. (2015) argued that if participants do not pay close attention to the introduction of referents in the discourse (indexed by increased alpha power), they will be less sensitive to an ambiguity manipulation that depends on having encoded that referential information (i.e., smaller Nref for ambiguous anaphors). This proposed link between alpha activity and attentional (dis)engagement during discourse-referential processing aligns well with the results of a recent study about narrative comprehension, which relied on cartoons as experimental manipulation. Guan et al. (2018) presented participants stories consisting of five vignettes (picture-sound pairs). The story narrative was presented in the first three vignettes. At the fourth vignette, participants were asked a question that probed them to think about the story protagonist's mental state, which could either be true or false compared to reality (i.e., a false belief task). Time-frequency analysis of EEG activity elicited at the fourth vignette showed a sustained increase in parietal-occipital alpha power for 'false' conditions, in which the protagonist's belief conflicted with reality. In line with the finding that alpha power is increased for more effortful internal cognitive activity when external information is actively suppressed (Bonnefond and Jensen, 2012; Foxe and Snyder, 2011), this alpha effect was

interpreted as an index of effortful internal processing related to the evaluation of two conflicting views (i.e., the perspective of the protagonist vs. reality).

1.4. The present study

The current study seeks to explore the electrophysiological signatures of referential and co-referential processes in visual narrative comprehension. To this end, we compare both evoked and induced activity elicited by different panels in semantically coherent narrative structures. These narrative structures contained six sequential images, which included refiner panels that zoom in on the contents of another full panel. Refiner panels and full panels, which differ in referential content, are co-referentially related to one another and are therefore presented within the same sequence: cataphoric panels served as antecedent for anaphoric panels (see Fig. 2 for an example). We test for the effect of panel type by comparing full panels to refiner panels (i.e., panels with blue vs. red outline in Fig. 2), and for the effect of anaphoricity by comparing anaphoric to cataphoric panels (i.e., panels with solid vs. dashed outline in Fig. 2). As cataphoric panels always precede anaphoric panels, we additionally test for the effect of ordinal sequence position on the processing of co-reference by comparing co-referential panels to non-co-referential panels. In Fig. 2, this corresponds to the co-referential (anaphoric) panels in position 4 and the non-co-referential panels in position 3 (see Section 3.3 for discussion).

We hypothesize that the neurocognitive processes triggered by these manipulations can be dissociated in event-related EEG activity. First of all, as refiners frame crucial information, they facilitate categorization processes. This would lead to a reduced N300 for refiners compared to full panels, which might or might not be followed by a modulation of the N400 component (see Cohn and Foulsham, 2020). We expect anaphoricity to be associated with the Nref effect: anaphoric panels are expected to elicit sustained negative potential compared to cataphoric panels, indexing co-referential processing initiated in response to the relational properties of anaphoric panels (e.g., Nieuwland and van Berkum, 2008). We also consider the possibility that panel type and anaphoricity interact, in particular with respect to cataphoric refiner panels and anaphoric full panels. That is, refiner panels are unexpected and infelicitous when used cataphorically (Cohn et al., in prep.), which leads to the prediction that they elicit an increased N400 compared to anaphoric refiners (Coderre et al., 2020; Cohn et al., 2012). While no such coherence effect is expected for full panels, which can be used both anaphorically and cataphorically, these panels might modulate LPC amplitude. We expect anaphoric full panels (which follow cataphoric refiners) to elicit a stronger LPC, indexing greater updating cost for panels that must be linked to a constrained situation model (Cohn and Foulsham, 2020).

These ERP effects might be accompanied by time-frequency effects in the theta band, which has been associated with the retrieval of lexical-semantic information from long-term memory (Bastiaansen et al., 2005, 2008). As full panels are semantically richer than refiners, we predict an increase in theta power for full panels compared to refiner panels (e.g., the content word-function word contrast in Bastiaansen et al., 2005). Moreover, as theta has been linked to reactivation from working memory as well, we expect anaphoric panels to elicit stronger theta power (e.g., the repeated-new contrast in Coopmans and Nieuwland, 2020) and stronger gamma power than cataphoric panels, reflecting the fact that anaphoric panels complete a co-referential dependency (Coopmans and Nieuwland, 2020; Nieuwland and Martin, 2017).

2. Methods

2.1. Participants, materials and procedure

Thirty-two right-handed participants (16 female, mean age = 22.3,

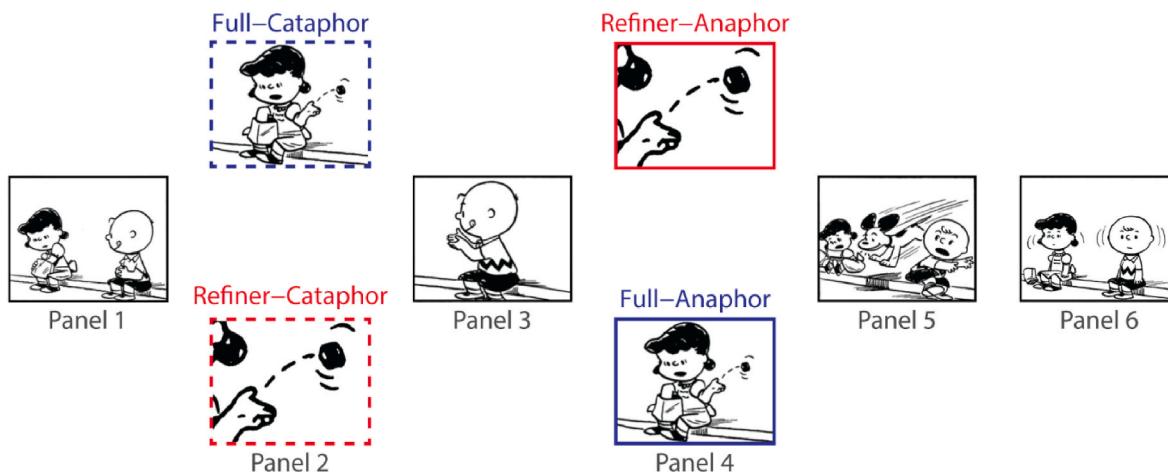


Fig. 2. Example narrative with two different sequences, in which the factors panel type and anaphoricity are crossed. In all sequences, the cataphoric panel served as the antecedent of the anaphoric panel. *Peanuts* is © Peanuts Worldwide LLC.

SD = 3.6) watched short visual narratives while their EEG was measured. All participants gave written informed consent to take part in the EEG experiment, which was approved by the Tilburg University School of Humanities and Digital Sciences Research Ethics and Data Management Committee. A total of 96 comic strips was used, each consisting of six panels, taken from a corpus of novel sequences constructed using panels from *The Complete Peanuts* by Charles Schulz. Of each comic strip, there were four versions: in the first two versions the full panel preceded the refiner, in the other two the refiner preceded the full panel (see Fig. 2). Within each order, the full panel and the refiner were either adjacent or they were separated by another panel which occupied position 3. The critical panels in the sequences were those in position 2, 3, and 4. Cataphoric panels always occupied position 2, whereas anaphoric panels were either in position 3 or 4. This was part of an anaphor-distance manipulation that was discussed in Cohn et al. (in prep.) but not analyzed here. The four different versions of each comic strip were equally distributed over four lists, with each list containing one version of each strip, such that participants never saw more than one version of the same strip. Each list also contained 72 filler sequences. For each strip we looked at the EEG activity elicited by both the cataphoric panel in position 2 (both refiner and full) and the anaphoric panel in position 3 or 4 (both refiner and full).

Participants were individually tested in a soundproof booth. At the start of each trial, they had to press a button, after which a visual narrative sequence appeared one panel at a time at the center of the screen. Each panel remained on the screen for 1350 ms. Successive panels were separated by 300 ms. To make sure participants paid attention to the sequences, they had to rate the comprehensibility of each sequence (1–7 Likert scale, ranging from “hard to understand” to “easy to understand”) immediately after its offset.

2.2. EEG recording and preprocessing

EEG was measured using a Brain Products actiCHamp system via 32 electrodes mounted in a Standard actiCAP. The data were acquired at a sampling rate of 250 Hz with a high cut-off filter of 70 Hz, and online referenced to electrode Fz. Ocular activity was monitored using electrodes beside the left eye and beneath the right eye.

For preprocessing, the data were high-pass filtered at 0.1 Hz (36 db/oct), re-referenced to the average of the left and right mastoids and segmented into epochs ranging from -1000 to 2000 ms relative to panel onset. We then visually inspected the data and excluded bad segments containing movement artifacts or multiple-channel muscle activity. We used Independent Component Analysis (ICA; using ICA weights from a 1 Hz high-pass filtered version of the data) to filter artifacts resulting from

eye movements. Last, we excluded trials in which the difference between the maximum and minimum voltage exceeded 200 μ V. This procedure excluded on average 8 segments per participant ($M = 4.2\%$ ($SD = 8.0\%$) of overall data; range across participants = 0–24.4%). As no participants were excluded after preprocessing, the analyses reported are based on the full sample of 32 participants.

2.3. ERP analysis

Before statistical analysis of the ERPs, we low-pass filtered the EEG signal at 35 Hz (36 db/oct) to remove high-frequency activity (Luck, 2014) and performed baseline correction using a 200 ms interval preceding the critical panel. ERPs were then analyzed with linear mixed-effects models (Baayen et al., 2008), using the lme4-package (Bates et al., 2014) in R (version 4.0.3, R Core Team, 2021). Separate models were applied to the N300, N400, LPC and Nref spatio-temporal regions of interest (ROI). For the N400, LPC and Nref, we adopted the temporal parameters used by Coopmans and Nieuwland (2020). For the LPC, the dependent variable was the average voltage value for each trial across the eight centro-parietal electrodes C3, Cz, C4, CP1, CP2, P3, Pz, P4 in a 500–1000 ms time window after panel onset. At the Nref ROI, the voltage values for each trial were averaged across the ten fronto-central electrodes Fp1, Fp2, F7, F3, F4, F8, FC5, FC1, FC2, FC6 in a 300–1500 ms time window. The N300 ROI was composed of the same electrodes in a 200–400 ms time window. These fronto-central electrodes were also used for the N400 ROI (300–500 ms), because the N400 for pictorial stimuli is more frontally distributed than the N400 elicited by linguistic stimuli (Cohn et al., 2012; Federmeier and Kutas, 2001; Ganis et al., 1996; West and Holcomb, 2002).

We applied separate models to all four ROIs to assess the effects of anaphoricity and panel type. For each analysis, we started with a full model, which had anaphoricity and panel type (both deviation coded) as well as their interaction as main effects. We included participant as random effect, which had a random intercept and the interaction between anaphoricity and panel type as random slope. In case of non-convergence, we removed the interaction term and retained anaphoricity and panel type separately as random slope for participant. To assess the effect of each factor, we compared the model with that factor to the model without it using the anova function in R ($\alpha = 0.05$).

2.4. Time-frequency analysis

Time-frequency analysis of oscillatory power was performed using the Fieldtrip toolbox (Oostenveld et al., 2011), following the same procedure as used by Coopmans and Nieuwland (2020). We performed

time-frequency analysis in two different but partially overlapping frequency ranges. For the low (2–30 Hz) frequency range, power was extracted from each individual frequency using a 400-ms sliding Hanning window in time steps of 10 ms. For the high (25–70 Hz) frequency range, we used a multitaper approach (Mitra and Pesaran, 1999) with Slepian tapers, with a 400-ms time-smoothing and a 5-Hz frequency-smoothing window, in frequency steps of 2.5 Hz and time steps of 10 ms. On each individual trial, power in the event-related interval was computed as a relative change from a baseline period ranging from −500 to −250 ms relative to panel onset. We computed average power changes for each condition and participant separately.

Differences in power across conditions were compared using non-parametric cluster-based random permutation tests (Maris and Oostenveld, 2007). First, by means of a two-sided dependent samples *t*-test we performed the comparisons described below, yielding uncorrected *p*-values. Neighboring data triplets of electrode, time and frequency band whose *p*-values exceeded a critical α -level of .05 were clustered. Clusters of activity were then evaluated by comparing their cluster-level test statistic (sum of individual *t*-values) to a reference distribution that was created by computing the largest cluster-level *t*-value on each of 1000 permutations of the same dataset. Clusters falling in the highest or lowest 2.5th percentile were considered significant. Using the correct-tail option to correct *p*-values for doing a two-sided test, we evaluated *p*-values at $\alpha = .05$.

All time (0–1500 ms) and frequency points (2–30 Hz; 30–70 Hz) were submitted to the cluster-based permutation tests, with which we tested the following comparisons: anaphoric panels vs. cataphoric panels (collapsed over panel type), full panels vs. refiner panels (collapsed over anaphoricity), and the interaction between panel type and anaphoricity. We tested for this interaction by comparing the effect of panel type in the anaphoric condition to the same effect in the cataphoric condition (i.e., difference between anaphor-full and anaphor-refiner vs. the difference between cataphor-full and cataphor-refiner).

To ensure that the reported time-frequency effects in the 2–30 Hz frequency band provide information above and beyond the information found in the ERPs, we also performed time-frequency analysis on the EEG signal after subtracting, for each participant, the average ERP for each condition (Cohen, 2014). As time-frequency data contains both phase-locked and non-phase-locked activity, any observed difference

might (at least partially) be caused by such phase-locked activity, which is the only activity contained in the ERP signal. Subtracting the average ERP removes much phase-locked activity from the signal, such that time-frequency analysis of the resulting signal provides a measure of non-phase-locked activity.

3. Results

3.1. Event-related potentials

The average ERPs of all conditions in the fronto-central ROI are presented in Fig. 3A. In the 200–400 ms N300 time window, the effect of panel type was similar for anaphoric and cataphoric panels ($\beta = -0.08$, $SE = 0.33$, $\chi^2 = 0.06$, $p = .80$). Moreover, anaphoric and cataphoric panels elicited similar ERP amplitude ($\beta = -0.14$, $SE = 0.16$, $\chi^2 = 0.71$, $p = .40$). Refiner panels elicited a less negative ERP than full panels ($\beta = 1.04$, $SE = 0.16$, $\chi^2 = 27.47$, $p < .001$ (see Fig. 3B for ERPs and Fig. 3D (top) for the scalp topography of this effect of panel type).

In the 300–500 ms N400 time window, the effect of anaphoricity was similar for refiner and full panels ($\beta = 0.01$, $SE = 0.33$, $\chi^2 = 0.001$, $p = .98$). Cataphoric panels elicited more negative N400 amplitude than anaphoric panels ($\beta = -0.38$, $SE = 0.18$, $\chi^2 = 4.37$, $p = .04$ (Fig. 3C; scalp topography in Fig. 3D (middle)). However, this effect of anaphoricity appeared to be driven by the different ordinal positions of anaphoric and cataphoric panels rather than their co-referential status (see Section 3.3). Refiner and full panels elicited similar N400 amplitude ($\beta = 0.11$, $SE = 0.19$, $\chi^2 = 0.35$, $p = .56$).

At the centro-parietal LPC ROI (500–1000 ms), the effect of panel type was similar for anaphoric and cataphoric panels ($\beta = 0.23$, $SE = 0.35$, $\chi^2 = 0.45$, $p = .50$). Anaphoric panels elicited more negative amplitude than cataphoric panels ($\beta = 0.64$, $SE = 0.18$, $\chi^2 = 10.74$, $p = .001$). Based on the sustained nature of this effect (see Fig. 3C), however, we believe that this is a spill-over of the Nref effect. Visual inspection of the ERPs in Fig. 3B suggests that full panels elicited more positive ERP amplitude than refiner panels in the LPC time window, but this difference was not significant: refiner and full panels elicited similar amplitude ($\beta = -0.18$, $SE = 0.17$, $\chi^2 = 1.34$, $p = .25$).

In the 300–1500 ms Nref time window, the effect of anaphoricity was similar for refiner and full panels ($\beta = 0.27$, $SE = 0.31$, $\chi^2 = 0.75$, $p =$

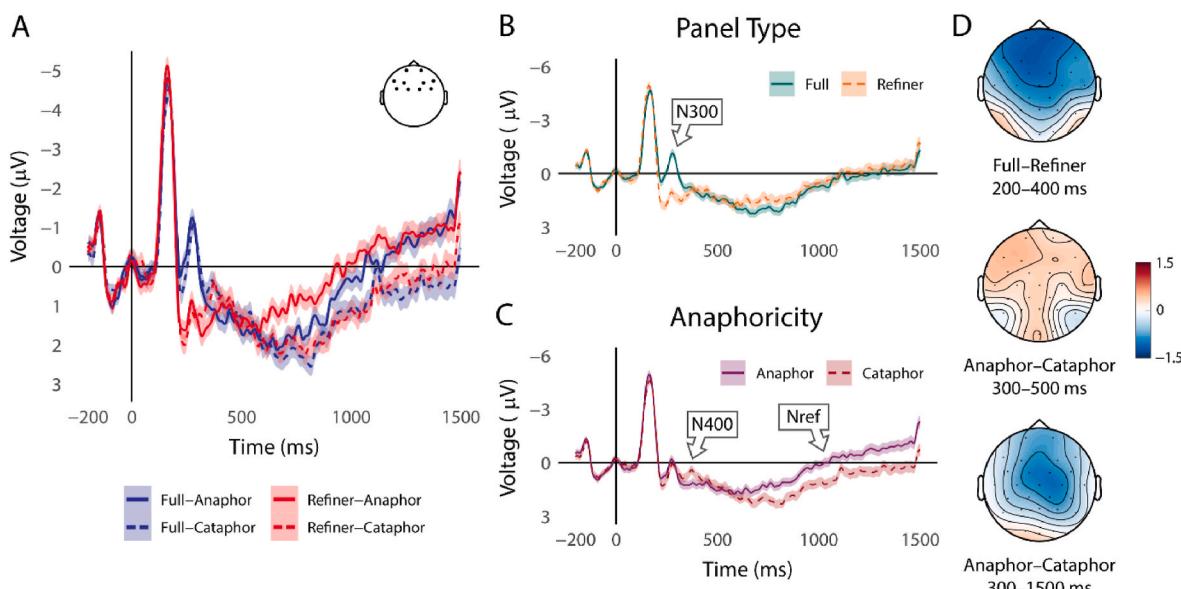


Fig. 3. (A) Event-related potentials (ERPs) for all conditions in the fronto-central region of interest. Negative voltage is plotted upwards. The shaded area around the ERPs represents within-subjects standard error of the mean per time sample. (B) ERPs by panel type, collapsed over anaphoricity. (C) ERPs by anaphoricity, collapsed over panel type. (D) Topographical plots of the difference between full and refiners panels in the 200–400 ms (N300) time window, and between anaphoric and cataphoric panels in the 300–500 ms (N400) and 300–1500 ms (Nref) time windows.

.39. The average ERP for anaphoric panels was more negative than the average ERP for cataphoric panels ($\beta = 0.90$, SE = 0.18), $\chi^2 = 19.11$, $p < .001$ (Fig. 3C; scalp topography in Fig. 3D (bottom)). Refiner and full panels elicited similar ERP amplitude ($\beta = -0.27$, SE = 0.21), $\chi^2 = 1.66$, $p = .20$.

These analyses show that, in the pre-defined Nref ROI, the effect of anaphoricity was similar for refiner and full panels. The ERPs in Fig. 3A, however, suggest that anaphoricity and panel type interact, but that this interaction is not captured in the rather long 300–1500 ms time window. It seems that the interaction has to do with the onset latency of the sustained negativity rather than with its amplitude. To explore the possibility that the Nref effect was delayed for full panels, we split the original ROI up into two time windows: one ranging from 500 ms (the offset of the N400 time window) to 1000 ms and the other from 1000 to 1500 ms (see also Nieuwland, 2014; Nieuwland et al., 2007). We ran two separate linear mixed-effects models on the voltage values in both time windows, both of which were averaged in the fronto-central ROI defined above. The model for the 500–1000 ms effect indeed showed an interaction between panel type and anaphoricity ($\beta = 0.59$, SE = 0.30), $\chi^2 = 3.85$, $p = .050$, while the model for the 1000–1500 ms effect did not ($\beta = -0.05$, SE = 0.38), $\chi^2 = 0.02$, $p = .90$. These exploratory results are based on a partially data-driven decomposition of the sustained negativity and should therefore be interpreted with caution, yet they do suggest that the onset of this effect of anaphoricity is modulated by the type of panel eliciting the co-referential processes.

3.2. Time-frequency data

As shown in Fig. 4, all conditions elicit a visually salient early increase in the theta band. Patterns in higher frequency ranges are less consistent.

Analyses in the low frequency band showed a theta effect as a function of panel type: refiner panels elicit stronger theta power (3–7 Hz) compared to full panels (one significant cluster, $p = .004$), which was most prominent between 100 and 300 ms after panel onset and had a fronto-central distribution (Fig. 5A). As this effect has a similar latency and scalp distribution as the N300 effect elicited by the same contrast (Fig. 3B and D (top)), it might be a time-frequency correlate of ERP activity. To see whether this is the case, we used cluster-based permutation tests on non-phase-locked activity elicited by full and refiner panels. This analysis was applied in the 0–1500 ms time window and in both a broad 2–30 Hz and a more specific 3–7 Hz frequency range. Neither contrasts yielded significant clusters, indicating that the theta effect is indeed driven by phase-locked activity and is therefore not complementary to the ERP results.

In the high frequency band, full panels elicited an increase in gamma

power compared to refiner panels. This effect shows up in two parts of the gamma-band: one cluster peaks around 40 Hz ($p = .032$), the other around 60 Hz ($p = .022$), but both have a similar time course and scalp distribution and might thus be the same effect (Fig. 5B).

Anaphoric panels elicited a decrease in alpha-band (8–12 Hz) power compared to cataphoric panels (one significant cluster, $p = .002$). This effect was most prominent between 400 and 1000 ms after panel onset and had a centro-parietal distribution (Fig. 6B). In the high frequency range, we found that anaphoric panels elicited a strong increase in 45–65 Hz gamma-band activity (one significant cluster, $p = .002$), which was widely distributed in time and space (Fig. 6A).

3.3. Exploratory analysis of the effects of sequence position

We recognize that the contrast between anaphoric and cataphoric panels is confounded with their ordinal sequence position, because cataphoric panels always precede anaphoric panels (i.e., cataphoric panels are always in position 2, anaphoric panels are in position 3 or 4). This might be particularly relevant for the alpha effect because alpha activity is linked to working memory processes, which plausibly vary across different positions in the comic strip. It is also relevant for the N400 effect, which has been shown to be modulated by the ordinal position of the panel or word it is elicited by (Cohn et al., 2012; Giglio et al., 2013; van Petten and Kutas, 1990). To check whether sequence position can explain these effects, we compared full panels that were co-referential (i.e., anaphoric panels; heads in their dependency) to full panels that were not co-referential (i.e., not heads). For the latter, we used the full panels of the Initial that were preceded by the refiner but were not co-referential with it (i.e., the picture of Charlie Brown preparing to catch the candy in panel 3 of Fig. 2). This contrast between co-referential and non-co-referential full panels is not confounded with sequence position, because both panel types were equally often in positions 3 and 4.

We assessed the effect of co-reference on the N400 amplitude by means of a linear mixed-effects model with condition (co-referential vs. non-co-referential) as fixed effect and participant as random effect, which had a random intercept and condition as random slope. This analysis showed that, in the N400 ROI, co-referential and non-co-referential panels elicited very similar N400 amplitude ($\beta = -0.11$, SE = 0.21), $\chi^2 = 0.29$, $p = .59$. The N400 effect of anaphoricity reported above thus appears to be an effect of sequence position, rather than being driven by the co-referential status of anaphoric panels.

For the time-frequency comparison of the same contrast, we restricted our analysis to centro-parietal activity within the 8–12 Hz alpha range and the 400–1000 ms time window (see the rectangular outlines in Fig. 7). These parameters are based on the spatio-temporal

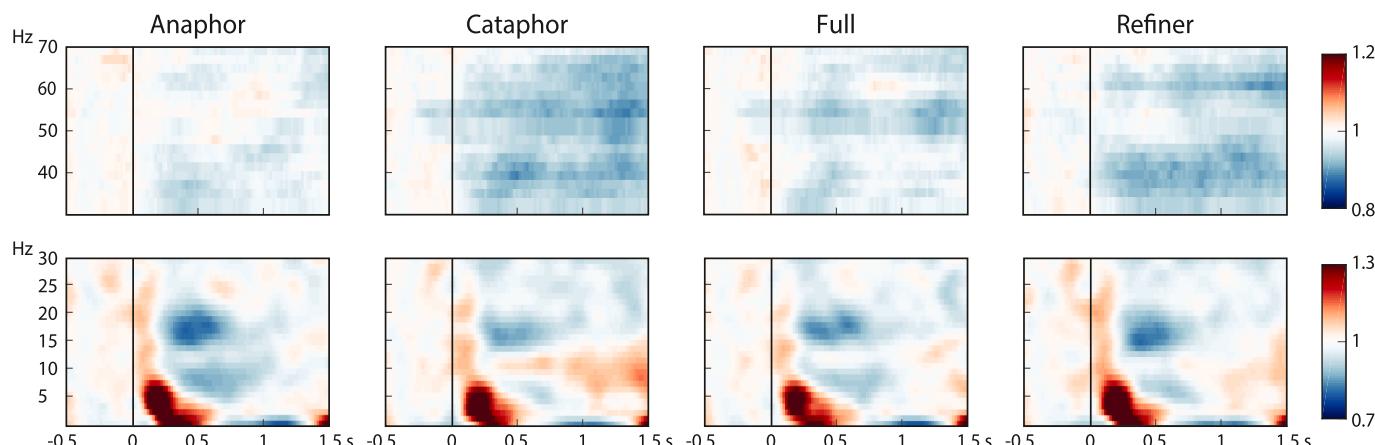


Fig. 4. Time-frequency representations of all conditions at centro-parietal electrode CP2. Power is represented as a relative change from activity in the baseline interval.

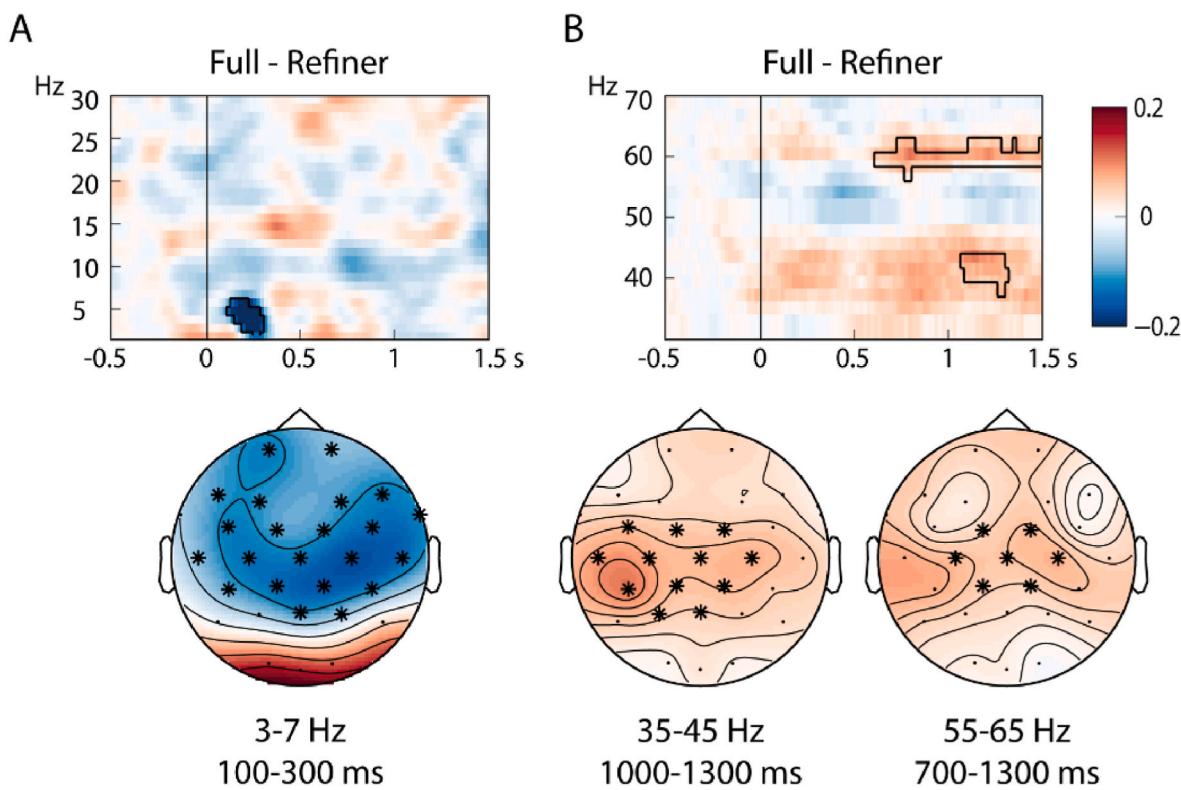


Fig. 5. Time-frequency representations (centro-parietal electrode CP2) for the contrasts between full and refiner panels in the low (A) and high (B) frequency range. The topographical plots show the average distributions of the theta effect (A) as well as both gamma effects (B). The electrodes that showed a significant difference in more than 30% of the selected time points are marked by an asterisk.

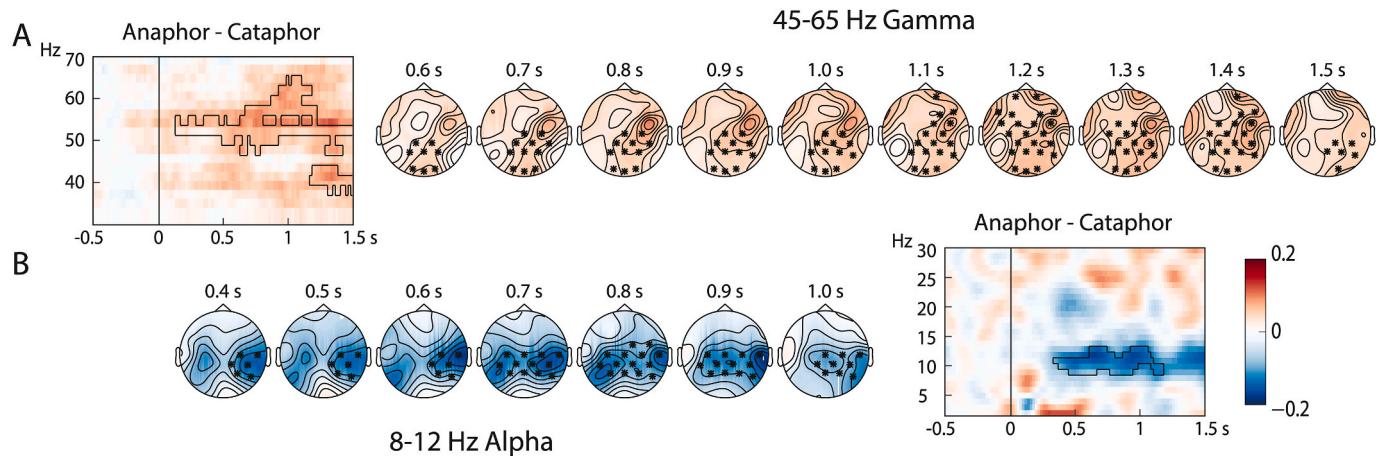


Fig. 6. Time-frequency representations (centro-parietal electrode CP2) for the contrasts between anaphoric and cataphoric panels in the high (A) and low (B) frequency range. The topographical plots show the spatio-temporal distributions of the 45–65 Hz gamma effect (A) and the 8–12 Hz alpha effect (B), with electrodes participating in the significant cluster at each of the time points marked by an asterisk.

properties of the alpha effect (see Fig. 6B). A two-sided paired-samples t -test on average power in this ROI indicated that alpha power was lower for co-referential than for non-co-referential full panels, $t(31) = -2.26$, $p = .031$. This effect is in line with the interpretation that the decrease in alpha power for anaphoric compared to cataphoric panels is related to the co-referentiality of anaphoric panels, not to their sequence position.

We aimed to make the same comparison for the Nref, which was more negative for anaphoric compared to cataphoric panels, but this effect of anaphoricity was also sensitive to panel type. We therefore compared non-co-referential full panels to both co-referential full panels and co-referential refiner panels. As the interaction between

anaphoricity and panel type was time-dependent, we again split the Nref ROI up into two time windows, ranging from 500 to 1000 ms and from 1000 to 1500 ms. In both time windows we compared two mixed-effects models, one with and one without the factor condition, which had the levels 'co-referent refiner panel', 'co-referent full panel', and 'non-co-referent full panel'.

In the 500–1000 ms window, the factor condition was associated with modulations of the Nref, $\chi^2 = 16.47$, $p < .001$. Pair-wise comparisons (Holm-corrected) showed that this effect was driven co-referential refiners (see Fig. 8), which elicited more negative potential than both co-referential full panels ($\beta = -0.63$, $SE = 0.21$, $z = -2.97$, $p = .006$) and

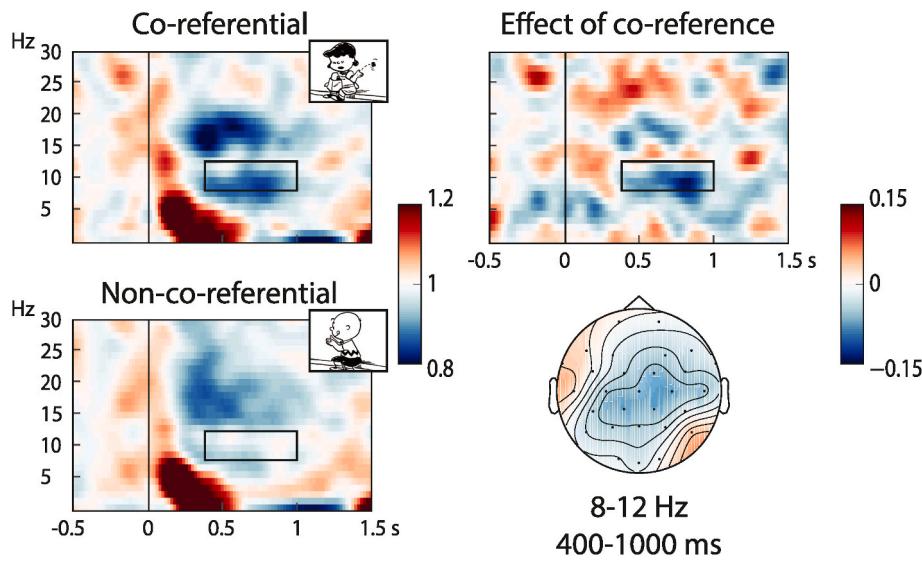


Fig. 7. Time-frequency representations (centro-parietal electrode CP1) for full panels that are co-referential, those that are not co-referential, and their difference. The black outline reflects the alpha region of interest on which the topographical plot is based.

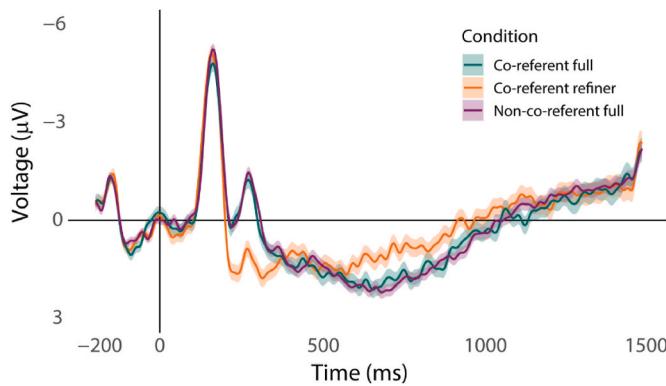


Fig. 8. Event-related potentials for co-referential refiner panels, co-referential full panels, and non-co-referential full panels in the fronto-central region of interest. Negative voltage is plotted upwards. The shaded area around the ERPs represents within-subjects standard error of the mean per time sample.

non-co-referential full panels ($\beta = -0.73$, SE = 0.18, $z = -3.98$, $p < .001$). Critically, co-referential and non-co-referential full panels did not differ from one another ($\beta = -0.10$, SE = 0.18, $z = -0.55$, $p = .58$), as can be seen in Fig. 8. In the 1000–1500 ms window, condition did not predict modulations of the Nref, $\chi^2 = 0.39$, $p = .82$. These findings suggest that the Nref effect can only partially be attributed to the different ordinal positions of anaphoric and cataphoric panels in the comic strip. That is, when we controlled for sequence position, the early time window shows an effect of co-reference for refiners only, in line with the interaction between anaphoricity and panel type in that window (see Section 3.1). The absence of a difference between the three conditions in the later time window suggests that the part of the Nref effect we observed for anaphoric panels in that later time window (i.e., main effect of anaphoricity; see Section 3.1) is not related to co-referentiality but rather reflects processes specifically related to panels later in the sequence.

4. Discussion

In this EEG study, we used ERP and time-frequency analyses to study referential and co-referential processes involved in the comprehension of visual narratives. Participants viewed short comic strips with refiner

panels and full panels that were co-referentially related to another. We analyzed EEG activity elicited by both panel types, and took into account whether they preceded (i.e., cataphoric) or followed (i.e., anaphoric) the panel to which they were co-referentially related. ERP analyses revealed that refiner panels elicit a reduced N300 compared to full panels, while anaphoric panels elicit a reduced N400 as well as a sustained negativity compared to cataphoric panels. The N400 effect appeared to be an effect of sequence position rather than anaphoricity, while the sustained negativity was sensitive to both the sequence position and the co-referential status of the critical panel.

Time-frequency analyses of power showed effects in several frequency bands. Full panels elicited a broad increase in gamma-band power compared to refiner panels, while anaphoric panels elicited reduced 8–12 Hz alpha power and increased 45–65 Hz gamma power compared to cataphoric panels. These effects reveal both similarities to and differences with the electrophysiological correlates of (co-)referential processing in language, suggesting that comprehending structured sequences in both domains relies on neural resources that are partially overlapping, and showing that time-frequency analysis provides a completer picture on the neurocognition of visual narrative comprehension.

4.1. Electrophysiological signatures of reference: refiners vs. full panels

Refiner panels elicited an attenuated negativity compared to full panels, which peaked between 200 and 400 ms after panel onset (see also Cohn and Foulsham, 2020). We interpret this component as an N300, which has been related to object categorization or identification (Draschkow et al., 2018; Hamm et al., 2002; McPherson and Holcomb, 1999). That is, the difference between refiners and full panels in this early time window can reflect the neural processes involved in extracting information about the content of the panel in order to recognize what is being looked at. These processes are differently engaged for refiner and full panels because they differ in complexity: as refiner panels zoom in on specific visual features of one element in the full panel, they attenuate demands on categorization (Cohn and Foulsham, 2020).

Modulations of the N300 often co-occur with a subsequent N400 effect (Cohn et al., 2012; Federmeier and Kutas, 2001; McPherson and Holcomb, 1999; West and Holcomb, 2002), but the N400 in our study was not modulated by panel type. In contrast to previous studies, we did not have a strong manipulation of expectancy or semantic congruency,

which could explain why our ERP results do not index differential demands on semantic processing. Indeed, while refiners and full panels might have differed in categorization and identification processes, as indexed by the N300 effect, they were presented in coherently structured visual sequences, so they plausibly required comparable access to semantic memory (Cohn and Foulsham, 2020).

The manipulation of panel type also affected gamma-band activity: full panels elicited stronger gamma power than refiner panels. This effect appeared in two frequency ranges (i.e., 35–45 Hz and 55–65 Hz), but both clusters showed a similar latency and scalp distribution, so we cannot exclude the possibility that they index one effect. This gamma-band difference might reflect a difference in object recognition for both panel types. Traditionally, gamma activity has been implicated in successful recognition of objects: gamma power is increased when sensory (and cognitive) properties of an object can be bound together into a coherent percept (Tallon-Baudry, 2003; Tallon-Baudry and Bertrand, 1999). However, object recognition happens quite early, as indicated by the early onset of the N300, and leads to a transient gamma response (Martinovic et al., 2007; Schneider et al., 2008; Tallon-Baudry, 2003). The late onset and prolonged time course of this effect (see Fig. 5B) therefore speak to an integratory interpretation. This need not be similar to the type of discourse-level anaphoric integration discussed in the next section; instead, it could index more general binding processes related to the featural and semantic richness of full panels (e.g., Clarke et al., 2011; Herrmann et al., 2004).

4.2. Electrophysiological signatures of co-reference: anaphoric vs. cataphoric panels

The comparison between anaphoric and cataphoric panels revealed two ERP effects, the first of which was a reduction in N400 amplitude for anaphoric compared to cataphoric panels. This effect was not driven by the co-referential status of anaphoric panels, because it disappeared when we compared co-referential to non-co-referential full panels while controlling for differences in sequence position. The reduced N400 therefore seems to be driven by the different ordinal positions of anaphoric and cataphoric panels in the comic strips (i.e., cataphoric panels always precede anaphoric panels). Panels at the start of sequences typically elicit a larger N400 than panels in subsequent sequence positions (Cohn et al., 2012; Giglio et al., 2013), and similar effects are found for words in early versus later sentence positions (van Petten and Kutas, 1990). This reduction in N400 amplitude for later panels likely reflects the facilitation of semantic access and/or integration due to the preceding coherent narrative structure.

The second ERP effect was a sustained negativity for anaphoric compared to cataphoric panels, which could only partially be attributed to their different sequence positions. We predicted an Nref effect, which would index the processing cost associated with establishing a co-referential dependency for anaphoric panels (Nieuwland and van Berkum, 2008). The negativity we observed is less frontally distributed than the Nref effect commonly associated with referential processing (Nieuwland and van Berkum, 2006; van Berkum et al., 1999; van Berkum et al., 2003; though see Coopmans and Nieuwland, 2020), which might have to do with the pictorial nature of the stimuli. Moreover, the onset of this effect was somewhat later than the Nref elicited by linguistic stimuli, which usually starts around 300 ms after the onset of the ambiguous critical word (Nieuwland and van Berkum, 2008). However, we do not think that this necessarily reflects a later onset of referential processing. Rather, it could be related to the fact that the contrast between anaphoric and cataphoric panels also elicited an N400 effect, which overlaps with the Nref effect in topography and (partially) in timing, but which had an opposite polarity. The onset of the Nref effect is right at the offset of the N400 effect (see Fig. 3C), which supports the possibility that these effects canceled each other out in the time window in which they overlap.

Two additional analyses show that the sustained negativity was not

identical for full and refiner panels. First, in the early half of the Nref time window, between 500 and 1000 ms after panel onset, the effect for anaphoric vs. cataphoric refiners was larger than the effect for anaphoric vs. cataphoric full panels (see Fig. 3A). This interaction disappeared in the subsequent 1000–1500 ms time window, which only contained a main effect of anaphoricity. Second, when we looked at the effect of co-reference while controlling for sequence position, we again found a time-dependent difference between anaphoric full and refiner panels. Specifically, the contrast between co-referential refiner panels, co-referential full panels and non-co-referential full panels showed an increased negativity for co-referential refiners compared to both co-referential and non-co-referential full panels in the 500–1000 ms time window, but no difference between the three conditions in the 1000–1500 ms time window (Fig. 8). This indicates that the effect of anaphoricity for refiner panels reflects their anaphoric properties, while the same effect for full panels instead reflects their later sequence position.

We suggest that this difference between co-referential (anaphoric) full and refiner panels is related to their different referential status (not unlike the contrast between pronouns and proper names; Gordon and Hendrick, 1998). As noted in the introduction, refiners are referentially dependent, whereas full panels are not. The reduced N300 for refiners suggests that comprehenders are quickly able to identify and categorize the semantic content of the refiner. It is likely that refiners are also quickly identified as being referentially dependent, hence automatically triggering an attempt to establish a co-referential relationship, indexed by the Nref effect. Full panels, instead, are semantically rich enough to establish a discourse representation themselves. As they need not be linked to another panel, they do not elicit co-referential processing right away (for a similar account of proper names, see Barkley et al., 2015; though see also Coopmans and Nieuwland, 2020). Eventually they do elicit increased negative potential compared to cataphoric full panels (Fig. 3A), but a similar effect was elicited by full panels that were not co-referential (Fig. 8). This later effect is thus not related to anaphoric processing, but might index the more general integration of referents into the event structure. That is, while non-co-referential full panels are not co-referential in the ‘anaphoric’ sense (i.e., linking two representations of the same character), they are still referentially related to the discourse model, as they play an active role in the narrated event (i.e., in the form of an interaction between characters, see the narrative structure in Fig. 1). The negativity for full panels in later sequence positions, observed for both co-referential and non-co-referential full panels, might thus reflect the participants’ attempt to link this panel to the event structure of the visual sequence (Wittenberg et al., 2014). This process could be a more general event structure integration process, related to the need for co-referential continuity (Cohn, 2020b), and possibly different from anaphoric processing. The potentially different nature of these effects could explain why the sustained negativity has a different time course for refiners compared to full panels.

In the time-frequency domain, the contrast between anaphoric and cataphoric panels also yielded two effects: compared to cataphoric panels, anaphoric panels elicited reduced 8–12 Hz alpha power (~400–1000 ms) as well as increased 45–65 Hz gamma power (~600–1500 ms). This particular pattern of results, in which a reduction in alpha power is accompanied by an increase in gamma power, has been reported in the working memory literature, where it is linked to the control of working memory storage (for reviews, see Miller et al., 2018; Roux and Uhlhaas, 2014). A reduction in alpha synchronization can be linked to activation of the underlying neural sources because alpha oscillations are thought to regulate the flow of information in the cortex via active inhibition of task-irrelevant brain regions (Jensen and Mazaheri, 2010; Klimesch, 2012). Event-related alpha desynchronization, called ‘alpha suppression’, thus reflects a release from inhibition and therefore an increase in the engagement of the relevant brain regions (Klimesch, 2012).

The alpha effect in our study appeared to be specific to co-referential

panels, which have to be linked to a previously presented panel, rather than being an effect of sequence position (see Fig. 7). We therefore interpret it as alpha suppression for anaphoric panels, reflecting reduced inhibition of the neuronal populations involved in establishing a co-referential relationship between an anaphoric panel and its antecedent. This co-referential process requires active engagement of the working memory system, both for reactivating (a memory trace of) the antecedent and for linking it with the representation of the anaphoric panel. On this account, these specific processes might be supported by activity in different frequency bands (e.g., theta, gamma), but their regulation is controlled by alpha oscillations. Consistent with this view, in language comprehension research, alpha suppression has been observed in response to modulations of reactivation (e.g., previously seen vs. unseen words, Rommers and Federmeier, 2018) as well as integration (e.g., increased load in speech-gesture integration, Drijvers et al., 2018). In addition, the centro-parietal topography of the alpha effect could indicate the involvement of the parietal lobe, whose medial regions have been associated with (modality-independent) reference resolution (Brodbeck et al., 2016), tentatively supporting the link between alpha suppression and co-reference processing.

The broad increase in gamma-band power for anaphoric compared to cataphoric elements might be related to their differential predictability: cataphoric panels, which occur early in the sequence, are less predictable than anaphoric panels. More specifically, refiner panels are preferentially used anaphorically, so cataphoric refiners are less expected than anaphoric refiners. Full panels, instead, can be used both cataphorically and anaphorically. However, the anaphoric full panels in our visual sequences were preceded by a cataphoric refiner. The cataphoric refiner triggers the expectation for a full panel, making anaphoric full panels more expected than cataphoric ones. Overall, then, anaphoric panels were more expected than cataphoric panels, even though these panel types are unlikely to differ in overall demands on semantic integration processes (i.e., all sequences were semantically coherent).

Gamma-band modulations induced by linguistic manipulations are often linked to semantic processing (Bastiaansen and Hagoort, 2015; Coopmans and Nieuwland, 2020; Fedorenko et al., 2016; Nieuwland and Martin, 2017; Peña and Melloni, 2012; Rommers et al., 2013), but some of these effects are better explained in terms of predictability than in terms of prediction (Lewis et al., 2015; Lewis and Bastiaansen, 2015; Wang et al., 2012, 2018). Specifically, Lewis et al. (2015) argue that gamma-band activity reflects the checking of incoming information against representations pre-activated in working memory. On this account, increased gamma-band synchronization is reflective of a confirmed prediction (see also Herrmann et al., 2004). In our experiment, the increase in gamma-band power for anaphoric panels could reflect the match between the predicted and the actually presented panel. These predictions might not be as specific as those found in language studies, in which it has been shown that people make predictions at several levels of representation. However, even in visual sequences, people do have clear expectations about the content of upcoming panels. Not only do they have general expectations about referential and semantic continuity (Cohn, 2020b; Cohn et al., 2014), but they also make specific predictions about how visual narrative sequences are likely to be followed (Coderre et al., 2020). In visual narratives, there is both visual and semantic overlap between the anaphoric panel and its antecedent (i.e., the refiner is a zoomed-in version of the full panel), so it is not unlikely that participants had expectations for what the anaphoric panel would look like. The reduced N400 for panels in later sequence positions is in line with this possibility. It has been shown that predictability modulates the N400 amplitude in a similar way in visual narratives and in sentence processing (Coderre et al., 2020), so we consider it reasonable that the gamma-band increase for anaphoric compared to cataphoric panels reflects the predictability of anaphoric panels.

5. Conclusion

Visual narrative comprehension is thought to rely on neurocognitive mechanisms that are also recruited by the language system. In this EEG study, we studied reference and co-reference in visual narrative comprehension through electrophysiological measures in the time and frequency domain. While we mainly examined the consequences of co-reference in terms of the effects elicited by the second element in a co-referential dependency, an interesting question for future research is how real-time processing of visual narratives is affected by the first element, whose properties might cue the existence of a (co-referentially) related panel downstream. For instance, are the predictive processes initiated by antecedent panels, such as cataphoric refiners, similar to those involved in comprehending distance dependencies in language (e.g., wh-phrases, cataphoric pronouns)? A time-frequency approach to such questions will complement the ERP literature and thereby lead to a stronger embedding of the study of visual narratives in the cognitive neuroscience of language.

Credit author statement

Cas W. Coopmans: Conceptualization, Methodology, Formal analysis, Writing – original draft, Visualization; **Neil Cohn:** Conceptualization, Methodology, Investigation, Writing – review & editing

Declaration of competing interest

None.

References

- Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59 (4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>.
- Barkley, C., Kluender, R., Kutas, M., 2015. Referential processing in the human brain: an Event-Related Potential (ERP) study. *Brain Res.* 1629, 143–159. <https://doi.org/10.1016/j.brainres.2015.09.017>.
- Bastiaansen, M.C.M., Hagoort, P., 2003. Event-induced theta responses as a window on the dynamics of memory. *Cortex* 39 (4), 967–992. [https://doi.org/10.1016/S0010-9452\(08\)70873-6](https://doi.org/10.1016/S0010-9452(08)70873-6).
- Bastiaansen, M.C.M., van der Linden, M., ter Keurs, M., Dijkstra, T., Hagoort, P., 2005. Theta responses are involved in lexical-semantic retrieval during language processing. *J. Cognit. Neurosci.* 17 (3), 530–541. <https://doi.org/10.1162/0898929053279469>.
- Bastiaansen, M.C.M., Hagoort, P., 2015. Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *J. Cognit. Neurosci.* 27 (11), 2095–2107. https://doi.org/10.1162/jocn_a_00829.
- Bastiaansen, M.C.M., Mazaheri, A., Jensen, O., 2013. Beyond ERP's: oscillatory neuronal dynamics. In: Luck, S.J., Kappenman, E.S. (Eds.), *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press, Oxford, pp. 31–49.
- Bastiaansen, M.C.M., Oostenveld, R., Jensen, O., Hagoort, P., 2008. I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain Lang.* 106 (1), 15–28. <https://doi.org/10.1016/j.bandl.2007.10.006>.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting Linear Mixed-Effects Models Using Lme4 arXiv preprint arXiv:1406.5823.
- Bonnefond, M., Jensen, O., 2012. Alpha oscillations serve to protect working memory maintenance against anticipated distractors. *Curr. Biol.* 22 (20), 1969–1974. <https://doi.org/10.1016/j.cub.2012.08.029>.
- Boudewyn, M.A., Long, D.L., Traxler, M.J., Lesh, T.A., Dave, S., Mangun, G.R., Carter, C. S., Swaab, T.Y., 2015. Sensitivity to referential ambiguity in discourse: the role of attention, working memory, and verbal ability. *J. Cognit. Neurosci.* 27 (12), 2309–2323. https://doi.org/10.1162/jocn_a_00837.
- Brodbeck, C., Gwilliams, L., Pylkkänen, L., 2016. Language in context: MEG evidence for modality-general and -specific responses to reference resolution. *ENeuro* 3 (6). <https://doi.org/10.1523/ENEURO.0145-16.2016>.
- Brouwer, H., Hoeks, J.C.J., 2013. A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Front. Hum. Neurosci.* 7 <https://doi.org/10.3389/fnhum.2013.00758>.
- Burkhardt, P., 2006. Inferential bridging relations reveal distinct neural mechanisms: evidence from event-related brain potentials. *Brain Lang.* 98 (2), 159–168. <https://doi.org/10.1016/j.bandl.2006.04.005>.
- Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. *Science* 304 (5679), 1926–1929. <https://doi.org/10.1126/science.1099745>.
- Clarke, A., Taylor, K.I., Tyler, L.K., 2011. The evolution of meaning: spatio-temporal dynamics of visual object recognition. *J. Cognit. Neurosci.* 23 (8), 1887–1899. <https://doi.org/10.1162/jocn.2010.21544>.

- Coderre, E.L., O'Donnell, E., O'Rourke, E., Cohn, N., 2020. Predictability modulates neurocognitive semantic processing of non-verbal narratives. *Sci. Rep.* 10 (1), 10326. <https://doi.org/10.1038/s41598-020-66814-z>.
- Cohen, M.X., 2014. *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press, Cambridge, MA.
- Cohn, N., 2013. *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. Bloomsbury, London.
- Cohn, N., 2014. You're a good structure Charlie Brown: the distribution of narrative categories in comic strips. *Cognit. Sci.* 38 (7), 1317–1359. <https://doi.org/10.1111/cogs.12116>.
- Cohn, N., 2019. Structural complexity in visual narratives: theory, brains, and cross-cultural diversity. In: Grishakova, M., Poulaki, M. (Eds.), *Narrative Complexity: Cognition, Embodiment, Evolution*. University of Nebraska Press, Lincoln, NE, pp. 174–199.
- Cohn, N., 2020a. Your brain on comics: a cognitive model of visual narrative comprehension. *Topics Cognit. Sci.* 12 (1), 352–386. <https://doi.org/10.1111/tops.12421>.
- Cohn, N., 2020b. *Who Understands Comics? Questioning the Universality of Visual Language Comprehension*. Bloomsbury, London.
- Cohn, N., Foulsham, T., 2020. Zooming in on the cognitive neuroscience of visual narrative. *Brain Cognit.* 146, 105634. <https://doi.org/10.1016/j.bandc.2020.105634>.
- Cohn, N., Kutas, M., 2015. Getting a cue before getting a clue: event-related potentials to inference in visual narrative comprehension. *Neuropsychologia* 77, 267–278. <https://doi.org/10.1016/j.neuropsychologia.2015.08.026>.
- Cohn, N., Kutas, M., 2017. What's your neural function, visual narrative conjunction? Grammar, meaning, and fluency in sequential image processing. *Cognit. Res.: Princip. Implicat.* 2 (1), 27. <https://doi.org/10.1186/s41235-017-0064-5>.
- Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P.J., Kuperberg, G.R., 2012. (Pea)nuts and bolts of visual narrative: structure and meaning in sequential image comprehension. *Cognit. Psychol.* 65 (1), 1–38. <https://doi.org/10.1016/j.cogpsych.2012.01.003>.
- Cohn, N., Jackendoff, R., Holcomb, P.J., Kuperberg, G.R., 2014. The grammar of visual narrative: neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia* 64, 63–70. <https://doi.org/10.1016/j.neuropsychologia.2014.09.018>.
- Cohn, N., van Middelaar, L., Foulsham, T., & Schilperoord, J. (in prep.). Zooming in at a Distance on Anaphoric Dependencies in Visual Narrative.
- Coopmans, C.W., Nieuwland, M.S., 2020. Dissociating activation and integration of discourse referents: evidence from ERPs and oscillations. *Cortex* 126, 83–106. <https://doi.org/10.1016/j.cortex.2019.12.028>.
- Deloigné, F., Brouwer, H., Crocker, M.W., 2019. Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain Cognit.* 135, 103569. <https://doi.org/10.1016/j.bandc.2019.05.007>.
- Draschkow, D., Heikel, E., Vö, M.L.-H., Fiebach, C.J., Sassenhagen, J., 2018. No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia* 120, 9–17. <https://doi.org/10.1016/j.neuropsychologia.2018.09.016>.
- Drijvers, L., Özürek, A., Jensen, O., 2018. Alpha and beta oscillations index semantic congruency between speech and gestures in clear and degraded speech. *J. Cognit. Neurosci.* 30 (8), 1086–1097. https://doi.org/10.1162/jocn_a_01301.
- Engel, A.K., Fries, P., Singer, W., 2001. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2 (10), 704–716. <https://doi.org/10.1038/35094565>.
- Federmeier, K.D., Kutas, M., 2001. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *J. Exp. Psychol. Learn. Mem. Cognit.* 27 (1), 202–224. <https://doi.org/10.1037/0278-7393.27.1.202>.
- Fedorenko, E., Scott, T.L., Brunner, P., Coon, W.G., Pritchett, B., Schalk, G., Kanwisher, N., 2016. Neural correlate of the construction of sentence meaning. *Proc. Natl. Acad. Sci. Unit. States Am.* 113 (41), E6256–E6262. <https://doi.org/10.1073/pnas.1612132113>.
- Filik, R., Sanford, A.J., 2008. When is cataphoric reference recognised? *Cognition* 107 (3), 1112–1121. <https://doi.org/10.1016/j.cognition.2007.11.001>.
- Foulsham, T., Cohn, N., 2021. Zooming in on visual narrative comprehension. *Mem. Cognit.* 49 (3), 451–466. <https://doi.org/10.3758/s13421-020-01101-w>.
- Foxe, J.J., Snyder, A.C., 2011. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Psychol.* 2 <https://doi.org/10.3389/fpsyg.2011.00154>.
- Ganis, G., Kutas, M., Sereno, M.I., 1996. The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *J. Cognit. Neurosci.* 8 (2), 89–106. <https://doi.org/10.1162/jocn.1996.8.2.89>.
- Giglio, A.C.A., Minati, L., Boggio, P.S., 2013. Throwing the banana away and keeping the peel: neuroelectric responses to unexpected but physically feasible action endings. *Brain Res.* 1532, 56–62. <https://doi.org/10.1016/j.brainres.2013.08.017>.
- Gordon, P.C., Hendrick, R., 1998. The representation and processing of co-reference in discourse. *Cognit. Sci.* 22 (4), 389–424. [https://doi.org/10.1016/S0364-0213\(99\)80045-7](https://doi.org/10.1016/S0364-0213(99)80045-7).
- Guan, Y., Farrar, M.J., Keil, A., 2018. Oscillatory brain activity differentially reflects false belief understanding and complementation syntax processing. *Cognit. Affect Behav. Neurosci.* 18 (1), 189–201. <https://doi.org/10.3758/s13415-018-0565-9>.
- Hamm, J.P., Johnson, B.W., Kirk, I.J., 2002. Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clin. Neurophysiol.* 113 (8), 1339–1350. [https://doi.org/10.1016/S1388-2457\(02\)00161-X](https://doi.org/10.1016/S1388-2457(02)00161-X).
- Heine, A., Tamm, S., Hofmann, M., Bösel, R.M., Jacobs, A.M., 2006. Event-related theta activity reflects memory processes in pronoun repetition. *Neuroreport* 17 (18), 1835–1839. <https://doi.org/10.1097/WNR.0b013e328010a096>.
- Herrmann, C.S., Munk, M.H.J., Engel, A.K., 2004. Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cognit. Sci.* 8 (8), 347–355. <https://doi.org/10.1016/j.tics.2004.06.006>.
- Jensen, O., Mazaheri, A., 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4 <https://doi.org/10.3389/fnhum.2010.00186>.
- Karimi, H., Swaab, T.Y., Ferreira, F., 2018. Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *J. Mem. Lang.* 102, 68–82. <https://doi.org/10.1016/j.jml.2018.05.003>.
- Kennison, S.M., Fernandez, E.C., Bowers, J.M., 2009. Processing differences for anaphoric and cataphoric pronouns: implications for theories of discourse processing. *Discourse Process* 46 (1), 25–45. <https://doi.org/10.1080/01638530802359145>.
- Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cognit. Sci.* 16 (12), 606–617. <https://doi.org/10.1016/j.tics.2012.10.007>.
- 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 (1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>.
- Lewis, A.G., Bastiaansen, M., 2015. A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex* 68, 155–168. <https://doi.org/10.1016/j.cortex.2015.02.014>.
- Lewis, A.G., Wang, L., Bastiaansen, M., 2015. Fast oscillatory dynamics during language comprehension: unification versus maintenance and prediction? *Brain Lang.* 148, 51–63. <https://doi.org/10.1016/j.bandl.2015.01.003>.
- Loschky, L.C., Larson, A.M., Smith, T.J., Magliano, J.P., 2020. The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics Cognit. Sci.* 12 (1), 311–351. <https://doi.org/10.1111/tops.12455>.
- Luck, S.J., 2014. *An introduction to the event-related potential technique*. MIT Press, Cambridge, MA.
- Magliano, J.P., Larson, A.M., Higgs, K., Loschky, L.C., 2016. The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Mem. Cognit.* 44 (2), 207–219. <https://doi.org/10.3758/s13421-015-0558-7>.
- Manfredi, M., Cohn, N., Kutas, M., 2017. When a hit sounds like a kiss: an electrophysiological exploration of semantic processing in visual narrative. *Brain Lang.* 169, 28–38. <https://doi.org/10.1016/j.bandl.2017.02.001>.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>.
- Marslen-Wilson, W., Levy, E., Tyler, L.K., 1982. Producing interpretable discourse: the establishment and maintenance of reference. In: Jarvela, R.J., Klein, W. (Eds.), *Speech, Place, and Action*. Wiley, New York, NY, pp. 339–378.
- Martinovic, J., Gruber, T., Müller, M.M., 2007. Induced gamma band responses predict recognition delays during object identification. *J. Cognit. Neurosci.* 19 (6), 921–934. <https://doi.org/10.1162/jocn.2007.19.6.921>.
- McPherson, W.B., Holcomb, P.J., 1999. An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology* 36, 53–65.
- Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., Friederici, A.D., 2015. Frontal-posterior theta oscillations reflect memory retrieval during sentence comprehension. *Cortex* 71, 205–218. <https://doi.org/10.1016/j.cortex.2015.06.027>.
- Miller, E.K., Lundqvist, M., Bastos, A.M., 2018. Working memory 2.0. *Neuron* 100 (2), 463–475. <https://doi.org/10.1016/j.neuron.2018.09.023>.
- Mitra, P.P., Pesaran, B., 1999. Analysis of dynamic brain imaging data. *Biophys. J.* 76 (2), 691–708. [https://doi.org/10.1016/S0006-3495\(99\)77236-X](https://doi.org/10.1016/S0006-3495(99)77236-X).
- Nieuwland, M.S., 2014. "Who's he?" Event-related brain potentials and unbound pronouns. *J. Mem. Lang.* 76, 1–28. <https://doi.org/10.1016/j.jml.2014.06.002>.
- Nieuwland, M.S., Coopmans, C.W., Sommers, R.P., 2019. Distinguishing old from new referents during discourse comprehension: evidence from ERPs and oscillations. *Front. Hum. Neurosci.* 13 <https://doi.org/10.3389/fnhum.2019.00398>.
- Nieuwland, M.S., Martin, A.E., 2017. Neural oscillations and a nascent corticohippocampal theory of reference. *J. Cognit. Neurosci.* 29 (5), 896–910. https://doi.org/10.1162/jocn_a_01091.
- Nieuwland, M.S., Otten, M., van Berkum, J.J.A., 2007. Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *J. Cognit. Neurosci.* 19 (2), 228–236. <https://doi.org/10.1162/jocn.2007.19.2.228>.
- Nieuwland, M.S., van Berkum, J.J.A., 2006. Individual differences and contextual bias in pronoun resolution: evidence from ERPs. *Brain Res.* 1118 (1), 155–167. <https://doi.org/10.1016/j.brainres.2006.08.022>.
- Nieuwland, M.S., van Berkum, J.J.A., 2008. The neurocognition of referential ambiguity in language comprehension. *Lang. Linguist. Compass* 2 (4), 603–630. <https://doi.org/10.1111/j.1749-818X.2008.00070.x>.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 156869 <https://doi.org/10.1155/2011/156869>.
- Peña, M., Melloni, L., 2012. Brain oscillations during spoken sentence processing. *J. Cognit. Neurosci.* 24 (5), 1149–1164. https://doi.org/10.1162/jocn_a_00144.
- Piai, V., Anderson, K.L., Lin, J.J., Dewar, C., Parvizi, J., Dronkers, N.F., Knight, R.T., 2016. Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proc. Natl. Acad. Sci. Unit. States Am.* 113 (40), 11366–11371. <https://doi.org/10.1073/pnas.1603312113>.

- Prystauka, Y., Lewis, A.G., 2019. The power of neural oscillations to inform sentence comprehension: a linguistic perspective. *Lang. Linguist. Compass* 13 (9), e12347. <https://doi.org/10.1111/lnc3.12347>.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>.
- Reid, V.M., Striano, T., 2008. N400 involvement in the processing of action sequences. *Neurosci. Lett.* 433 (2), 93–97. <https://doi.org/10.1016/j.neulet.2007.12.066>.
- Rommers, J., Dijkstra, T., Bastiaansen, M., 2013. Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *J. Cognit. Neurosci.* 25 (5), 762–776. https://doi.org/10.1162/jocn_a.00337.
- Rommers, J., Federmeier, K.D., 2018. Predictability's aftermath: downstream consequences of word predictability as revealed by repetition effects. *Cortex* 101, 16–30. <https://doi.org/10.1016/j.cortex.2017.12.018>.
- Roux, F., Uhlhaas, P.J., 2014. Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cognit. Sci.* 18 (1), 16–25. <https://doi.org/10.1016/j.tics.2013.10.010>.
- Schneider, T.R., Debener, S., Oostenveld, R., Engel, A.K., 2008. Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage* 42 (3), 1244–1254. <https://doi.org/10.1016/j.neuroimage.2008.05.033>.
- Singer, W., 2011. Dynamic formation of functional networks by synchronization. *Neuron* 69 (2), 191–193. <https://doi.org/10.1016/j.neuron.2011.01.008>.
- Sitnikova, T., Holcomb, P.J., Kuperberg, G.R., 2008. Neurocognitive mechanisms of human comprehension. In: Shipley, T.F., Zacks, J.M. (Eds.), *Understanding Events: from Perception to Action*. Oxford University Press, Oxford, pp. 639–683.
- Tallon-Baudry, C., 2003. Oscillatory synchrony and human visual cognition. *J. Physiol. Paris* 97 (2), 355–363. <https://doi.org/10.1016/j.jphysparis.2003.09.009>.
- Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representation. *Trends Cognit. Sci.* 3 (4), 151–162. [https://doi.org/10.1016/S1364-6613\(99\)01299-1](https://doi.org/10.1016/S1364-6613(99)01299-1).
- van Berkum, J.J.A., 2009. The neuropragmatics of “simple” utterance comprehension: an ERP review. In: Sauerland, U., Yatshiro, K. (Eds.), *Semantics and Pragmatics: from Experiment to Theory*. Palgrave Macmillan, pp. 276–316.
- van Berkum, J.J.A., Brown, C.M., Hagoort, P., 1999. Early referential context effects in sentence processing: evidence from event-related brain potentials. *J. Mem. Lang.* 41 (2), 147–182. <https://doi.org/10.1006/jmla.1999.2641>.
- van Berkum, J.J.A., Brown, C.M., Hagoort, P., Zwitserlood, P., 2003. Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology* 40 (2), 235–248. <https://doi.org/10.1111/1469-8986.00025>.
- van Berkum, J.J.A., Koornneef, A.W., Otten, M., Nieuwland, M.S., 2007. Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171. <https://doi.org/10.1016/j.brainres.2006.06.091>.
- van Dijk, T., Kintsch, W., 1983. *Strategies of Discourse Comprehension*. Academic Press, New York, NY.
- van Petten, C., Kutas, M., 1990. Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cognit.* 18 (4), 380–393. <https://doi.org/10.3758/BF03197127>.
- Varela, F., Lachaux, J.-P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2 (4), 229–239. <https://doi.org/10.1038/35067550>.
- Vonk, W., Hustinx, L.G.M.M., Simons, W.H.G., 1992. The use of referential expressions in structuring discourse. *Lang. Cognit. Process.* 7 (3–4), 301–333. <https://doi.org/10.1080/01690969208409389>.
- Wang, L., Hagoort, P., Jensen, O., 2018. Gamma oscillatory activity related to language prediction. *J. Cognit. Neurosci.* 30 (8), 1075–1085. https://doi.org/10.1162/jocn_a_01275.
- Wang, L., Zhu, Z., Bastiaansen, M., 2012. Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Front. Psychol.* 3 <https://doi.org/10.3389/fpsyg.2012.00187>.
- West, W.C., Holcomb, P.J., 2002. Event-related potentials during discourse-level semantic integration of complex pictures. *Cognit. Brain Res.* 13 (3), 363–375. [https://doi.org/10.1016/S0926-6410\(01\)00129-X](https://doi.org/10.1016/S0926-6410(01)00129-X).
- Willems, R.M., Oostenveld, R., Hagoort, P., 2008. Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension. *Brain Res.* 1219, 78–90. <https://doi.org/10.1016/j.brainres.2008.04.065>.
- Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., Kuperberg, G.R., 2014. The difference between “giving a rose” and “giving a kiss”: sustained neural activity to the light verb construction. *J. Mem. Lang.* 73, 31–42. <https://doi.org/10.1016/j.jml.2014.02.002>.