

Better together: integrating multivariate with univariate methods, and MEG with EEG to study language comprehension

Lin Wang & Gina R. Kuperberg

To cite this article: Lin Wang & Gina R. Kuperberg (2023): Better together: integrating multivariate with univariate methods, and MEG with EEG to study language comprehension, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2023.2223783](https://doi.org/10.1080/23273798.2023.2223783)

To link to this article: <https://doi.org/10.1080/23273798.2023.2223783>



Published online: 12 Jun 2023.



Submit your article to this journal [↗](#)



Article views: 161



View related articles [↗](#)



View Crossmark data [↗](#)



Better together: integrating multivariate with univariate methods, and MEG with EEG to study language comprehension

Lin Wang^{a,b} and Gina R. Kuperberg^{a,b}

^aDepartment of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; ^bDepartment of Psychology, Tufts University, Medford, MA, USA

ABSTRACT

We used MEG and EEG to examine the effects of Plausibility (*anomalous vs. plausible*) and Animacy (*animate vs. inanimate*) on activity to incoming words during language comprehension. We conducted univariate event-related and multivariate spatial similarity analyses on both datasets. The univariate and multivariate results converged in their time course and sensitivity to Plausibility. However, only the spatial similarity analyses detected effects of Animacy. The MEG and EEG findings largely converged between 300–500 ms, but diverged in their univariate and multivariate responses to anomalies between 600–1000 ms. We interpret the full set of results within a predictive coding framework. In addition to the theoretical significance, we discuss the methodological implications of the convergence and divergence between the univariate and multivariate results, as well as between the MEG and EEG results. We argue that a deeper understanding of language processing can be achieved by integrating different analysis approaches and techniques.

ARTICLE HISTORY

Received 10 November 2022
Accepted 5 June 2023

KEYWORDS


ERP; representational similarity analysis; Animacy; predictive coding; N400; P600; evidence accumulation


General introduction

To comprehend language, we must transform a continuous stream of sounds or letters into meaning. How do we accomplish this feat? How and when does the brain encode the semantic features of each incoming word during real-time comprehension? Does the process of retrieving these semantic features always produce an overall increase in brain activity? The exquisite temporal resolution of electroencephalography (EEG) and magnetoencephalography (MEG) make these neuroimaging techniques ideally suited for addressing these questions. As information encoded within each new linguistic input is passed up and down the cortical hierarchy, large numbers of spatially aligned pyramidal neurons produce electric dipoles, generating electric and magnetic fields (Buzsáki et al., 2012) that can be detected at the head surface using EEG and MEG. At each recording channel, the measured voltage or magnetic field reflects the sum of neural activity generated by multiple underlying dipole sources, scaled by a weighting factor that, in EEG, is influenced by the conductivities of head tissue, particularly the skull and scalp (Nunez, 1990), and, in MEG, is influenced by the folding pattern of the cortex (Hämäläinen et al., 1993).

In a typical EEG/MEG experiment, we measure activity at multiple recording channels at each time point following the onset of an “event” of interest, such as a critical word in a sentence. Our goal is to understand how our experimental variables influence this recorded activity so that we can make inferences about when and how the brain extracts information from the linguistic input as it unfolds in real time. To achieve this goal, we can take one of two different types of analytic approaches.

The first is to carry out a classic *event-related analysis*. This approach has a long history in psycholinguistic research, dating from the discovery of the N400 in 1980 (Kutas & Hillyard, 1980). The basic assumption of an event-related analysis is that an experimental variable of interest modulates the *amplitude* of activity measured at individual adjacent channels. This assumption is naturally incorporated in a *univariate* General Linear Model, which asks whether and when a particular variable explains variance in the amplitude of a dependent variable across multiple items (Baayen et al., 2008; Clark, 1973). If the experimental variable is categorical, then the intercept term in this type of model simply corresponds to the *average* amplitude of the response across all trials in the reference condition, and the

CONTACT Lin Wang  lwang48@mg.harvard.edu  Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02129, USA; Gina R. Kuperberg  gkuperberg@mg.harvard.edu  Department of Psychology, Tufts University, Medford, MA 02155, USA

 Supplemental data for this article can be accessed <http://doi.org/10.1080/23273798.2023.2223783>.

© 2023 Informa UK Limited, trading as Taylor & Francis Group

model's beta weight corresponds to the *mean difference* between experimental conditions (Smith & Kutas, 2015). Indeed, the classic way of carrying out an event-related analysis in EEG and MEG is simply to *average* the measured activity at each recording channel, across all trials within each experimental condition at each time point following the onset of the critical event (Luck, 2014a). This yields a time course of activity at each channel – a sequence of “waveforms” that in EEG are referred to as event-related potentials (ERPs), and in MEG are referred to as event-related fields (ERFs). These event-related responses can be described in terms of their amplitude, latency, duration, scalp topography, and, in the case of EEG, their polarity (whether the voltage is negative-going or positive-going). If we believe that a particular event-related response is reliably modulated by a particular type of psychological event, then we refer to it as an event-related *component* (Kappenman & Luck, 2012).

At a neurophysiological level, the interpretation of an event-related component is relatively straightforward: its *amplitude* (often averaged across a particular time window and adjacent channels) is taken to reflect the *strength* of time-locked (and phase-locked) neural activity produced by the underlying dipole source(s), and any *difference* in amplitude that can be explained by our experimental variable is taken to reflect differences in the strength of this underlying neural activity. At a cognitive level, this difference is usually interpreted as reflecting the effect of the experimental variable on a neurocognitive “process” (Kappenman & Luck, 2012).

This classic event-related analysis approach has yielded a large literature that has characterised a number of different event-related components, which have each been linked to different aspects of language processing (see Swaab et al., 2012 and Dikker et al., 2020 for reviews of language-related ERP and MEG components, respectively). For example, during sentence comprehension, N400 event-related component¹ is usually taken to reflect the ease of “retrieving” or “accessing” the semantic features of incoming words (Hagoort, 2013; Lau et al., 2008; Van Berkum, 2009).

The second approach that we can take to analyse EEG/MEG data has a much shorter history in neurophysiology and psycholinguistics. Instead of focusing on the *amplitude* of activity measured at individual adjacent recording channels, this approach aims to characterise whole *patterns* of activity, measured across *multiple* recording channels. This *multivariate* analysis approach was first developed in fMRI to describe patterns of activity observed across multiple voxels within neuroanatomical regions of interest (Haxby et al., 2001). It was argued that, in contrast to univariate methods,

which focused on the magnitude of activity at individual voxels, these spatial patterns better reflected the “representational information” encoded within these regions (Kriegeskorte & Bandettini, 2007).

Initially, it was assumed that multivariate methods were unsuited for analysing EEG/MEG data at the head surface. However, it has since become clear that unique spatial patterns produced within underlying neuroanatomical sources can give rise to unique spatial patterns of electrical and magnetic activity that can be detected either across the full set of recording channels (e.g. Cichy et al., 2014; Stokes et al., 2015) or across subsets of channels (e.g. Karimi-Rouzbahani et al., 2021) at the head surface.

In a multivariate analysis, the basic assumption is that the experimental variable of interest influences the geometric relationship amongst these spatial patterns. For example, one type of multivariate approach – *spatial similarity analysis*, otherwise known as representational similarity analysis asks whether and when the similarity/dissimilarity amongst items, with respect to the experimental variable, can explain the similarity/dissimilarity amongst the spatial patterns produced by these items (Kriegeskorte et al., 2008; Nili et al., 2014). In a typical representational similarity analysis stream, this is examined by constructing a “Model” dissimilarity matrix and correlating it with a “Neural” dissimilarity matrix at each time point following event onset, yielding a time series of *r* values (e.g. Cichy et al., 2014).

In contrast to event-related components, there is no *direct* neurophysiological interpretation of these multivariate time series. They simply tell us whether and when the spatial patterns of neural activity produced by the stimuli can be *discriminated* by the experimental variable of interest. From a cognitive perspective, this is usually taken to reflect whether and when “representational information” linked to this variable is neurally *decodable*.

The present study

To sum up, there is a long history of using EEG and MEG with event-related analyses to study the neural basis of online language comprehension. The combination of EEG/MEG with spatial similarity analysis and other multivariate methods is more recent. Because these newer multivariate approaches come from a different tradition and use different statistical methods, it is sometimes assumed that they detect neural activity that is functionally distinct and separable from that indexed by classic EEG/MEG components. However, this assumption is not always valid, and, despite close methodological links between the two approaches, few psycholinguistic

studies have systematically compared the results of these two types of analyses in the same datasets. In addition, there has been little discussion in the psycholinguistic literature about how to synthesise the results of event-related and spatial similarity analyses to inform theory. Finally, few studies have directly compared univariate or multivariate findings between MEG and EEG. In this investigation, we aimed to close these gaps. We addressed three sets of questions.

Question 1: when and how does the brain encode the Animacy of incoming words during online language comprehension?

Our first goal was to use multivariate and univariate methods with MEG/EEG to address a theoretical question: When and how does the brain encode the animacy of incoming words during online language comprehension, and how is this influenced by the prior context?

A large literature using event-related approaches has established that, between 300–500ms, incoming words whose semantic features match these that were pre-activated by the prior context produce a smaller N400 event-related component than words whose semantic features fail to match these predictions (ERP: Kuperberg et al., 2020; Kutas & Hillyard, 1980; Nieuwland et al., 2020; MEG: Halgren et al., 2002; Helenius et al., 1998; Ihara et al., 2007; Maess et al., 2006). For example, *plausible* nouns whose animacy-based semantic features match those that are pre-activated by a prior verb (Szewczyk & Schriefers, 2013; Wang et al., 2020) produce a smaller N400 than *anomalous* nouns whose features fail to match these verb-based predictions (Kuperberg et al., 2020; Paczynski & Kuperberg, 2011, 2012; Szewczyk & Schriefers, 2013).

In line with the idea that event-related components index “processing”, the smaller N400 response to expected (*versus* unexpected) words is usually taken to reflect the ease of semantic “retrieval” (Hagoort, 2013; Van Berkum, 2009) or lexico-semantic “access” (Lau et al., 2008). However, another way of understanding the N400 is as indexing the amount of *new information* that is encoded within the bottom-up input, i.e. the amount of information that was not already predicted by the prior context (Kuperberg, 2016). This is the premise of *predictive coding* – a general computational theory of brain function that posits that unpredicted information encoded within the bottom-up input is detected by local “error units”, which produce “prediction error” (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999), and an event-related response (Friston, 2005), see General Discussion for further elaboration. Critically, according to predictive coding, even if new

bottom-up information matches prior predictions, failing to activate error units and produce a large univariate response, it is still encoded within separate sets of “state units”. As such, it should still be possible to detect this expected information within the critical 300–500ms time window in which the bottom-up input first makes contact with semantic memory.

To test this theory, we measured MEG and EEG as participants read three-sentence discourse scenarios. We varied both the Plausibility and the Animacy of a critical word in the final sentence, such that it was either *animate* or *inanimate*, and either *plausible* (matching the animacy constraints of the prior verb) or *anomalous* (mismatching these animacy constraints). In addition to carrying out event-related univariate analyses, we carried out multivariate spatial similarity analyses on the same datasets. Previous fMRI studies have shown that multivariate methods can discriminate animate and inanimate visual objects (Devereux et al., 2013; Kriegeskorte et al., 2008; Proklova et al., 2016; Sha et al., 2015) and words (Devereux et al., 2013). In addition, previous EEG/MEG studies have used representational similarity analysis to show that animacy-based discriminations of visual objects occur rapidly (Carlson et al., 2013; Cichy et al., 2014; Cichy & Pantazis, 2017; Khaligh-Razavi et al., 2018). However, no previous study has used these methods to ask whether and when the brain distinguishes between *animate* and *inanimate* incoming words during language comprehension. Based on principles of predictive coding, we predicted that the spatial similarity analysis would detect an effect of Animacy within the same 300–500 ms time window in which the event-related analysis detected an effect of Plausibility on the N400, and that this effect would be detected regardless of whether the incoming words matched or mismatched the animacy-based constraint of the prior verb.

In addition to asking *when* the animacy-based features of incoming words were encoded, we were also interested in *how* these features were encoded. A long line of Cognitive Science research has established that the semantic features of *animate* items are more similar to one another than the semantic features of *inanimate* items (Garrard et al., 2001; McRae et al., 1997; Randall et al., 2004; Zannino et al., 2006). It has been proposed that these inherent differences in semantic similarity can account for the brain’s sensitivity to animacy-based categorical structure (Devlin et al., 1998; Gonnerman et al., 1997; Moss et al., 1998; Taylor et al., 2011; Tyler & Moss, 2001). On this account, the similarity between the spatial patterns produced by *animate* and *inanimate* nouns should mirror this internal similarity structure. We therefore hypothesised that,

between 300–500 ms, any spatial similarity effect of Animacy would be driven by a greater similarity amongst the spatial patterns produced by *animate* than *inanimate* nouns.

To test these hypotheses, we extracted the spatial patterns produced by each critical noun at each time point following its onset and calculated the spatial similarity between these patterns across all pairs of nouns. We then averaged these pairwise spatial similarity values (r values) within each of the four experimental conditions, giving rise to four spatial similarity time series. This allowed us to test our hypothesis regarding the directionality of the spatial similarity effect (*animate* > *inanimate*), as well as to directly compare its timing with the event-related effect of Plausibility.

Question 2: how do spatial similarity and event-related measures converge and diverge during language comprehension?

Our second goal was more methodological in nature. As discussed earlier, event-related and spatial similarity analyses each aim to characterise different aspects of the underlying neural signal – the strength of the dipoles *versus* the spatial patterns they produce. In principle, these two measures should be independent of each other. However, both analysis methods face the challenge of distinguishing very small signals from ongoing activity that is not time-locked to stimulus onset, and independent of the experimental variable(s) of interest, i.e. “noise”. Therefore, in practice, estimates of spatial similarity are likely to covary with the amplitude of evoked responses. This covariation between univariate and multivariate responses has been discussed in the fMRI literature (e.g. Haxby et al., 2001; Jimura & Poldrack, 2012; Walther et al., 2016), and to some degree, in the EEG/MEG literature (Guggenmos et al., 2018). However, the degree to which multivariate methods are sensitive to the classic event-related components and effects that are typically produced during language comprehension is unclear.

One reason why we know so little about how event-related and multivariate measures converge and diverge is that these methodologies use quite different analysis streams and methods of visualisation. As discussed above, in a typical event-related analysis, we average activity across items within each condition at each time point, with the goal of determining when one condition produces more or less activity than another (Kappenman & Luck, 2012). In contrast, in a typical representational similarity analysis, we examine correlations between a Model dissimilarity matrix and a Neural dissimilarity matrix at each time point, with the goal of determining when the conditions can be

discriminated (cf. Kriegeskorte et al., 2008; Nili et al., 2014). Therefore, directly comparing the results of the two methods is challenging.

The experimental design and the spatial similarity analysis methods that we employed in the current study allowed us to overcome this challenge. Our strategy of constructing separate spatial similarity time series for each condition allowed us to directly compare the shape and time course of these time series with the event-related waveforms produced by each condition. Moreover, by orthogonally manipulating Animacy and Plausibility in the same set of linguistic stimuli, and examining the effect of both variables on the same univariate and multivariate measures, we were able to determine when and how the univariate and multivariate results converged and diverged. We hypothesised that the event-related analyses would fail to detect any spatial similarity effects of Animacy that were detected by the spatial similarity analyses, but that the spatial similarity analyses would capture all the effects of Plausibility that were detected by the event-related analyses.

Question 3: how do the results of MEG and EEG converge and diverge during language comprehension?

Our third goal was to directly compare the results (both univariate and multivariate) of data that were collected using two different *techniques* – MEG and EEG.

Because MEG and EEG both index postsynaptic activity that is produced by pyramidal neurons, they are sensitive to many of the same event-related components, including the N400. There is also evidence that MEG and EEG are sensitive to some of the same multivariate effects (Cichy & Pantazis, 2017; Wang et al., 2020). However, there are also important differences between the two techniques. Unlike the magnetic field, the electric field is smeared at the head surface (Geisler & Gerstein, 1961; Grynszpan & Geselowitz, 1973). In addition, MEG and EEG differ in their sensitivities to underlying dipole sources with different orientations (Cuffin & Cohen, 1979; Hämäläinen et al., 1993). Therefore, the two types of techniques may not always detect the same neurocognitive mechanisms. For example, MEG is less sensitive than EEG to the domain-general P300 event-related component (Siedenberg et al., 1996).

In the present study, we were particularly interested in comparing the MEG and EEG responses to anomalous inputs in a later time window between 600–1000 ms. When a linguistic anomaly cannot initially be incorporated into a high-level interpretation, the disruption of comprehension is thought to trigger multiple processes,

including reprocessing the bottom-up input, reactivating prior top-down predictions, and tracking conflict between these different sources of information to verify that the disruption stemmed from an external (rather than an internal) error (Kuperberg et al., [under review](#)).

In EEG, the prolonged late error-based response manifests as a large posteriorly distributed late positivity ERP effect, known as the P600 (Brothers et al., [2020](#); Brothers et al., [2022](#); Kuperberg, [2007](#); Kuperberg et al., [2003](#); Kuperberg et al., [2020](#); Kuperberg et al., [under review](#); van de Meerendonk et al., [2009](#)). The P600 ERP component shares several functional properties with the P300 (Coulson et al., [1998](#); Osterhout et al., [2012](#); Sassenhagen et al., [2014](#); Sassenhagen & Fiebach, [2019](#)), and it may play a particularly important role in tracking the source of linguistic errors (Brothers et al., [2022](#); Kuperberg et al., [under review](#)). However, previous MEG studies have not reported an analogous effect on anomalies in this late time window. This raises the possibility that, due to their differential sensitivity to distinct neuroanatomical sources, MEG and EEG may also be differentially sensitive to distinct aspects of the prolonged error-based response between 600–1000 ms.

To date, few psycholinguistic studies have directly compared the results of EEG and MEG, using the same stimuli and tasks. We therefore carried out the same experiment using MEG (reported in Study 1) and EEG (reported in Study 2). This allowed us to qualitatively compare both the event-related and spatial similarity results across the two techniques. We hypothesised that the MEG and EEG results would largely converge within the 300–500 ms time window, but that they would show differences in their sensitivities to the activity produced by semantic anomalies in the later 600–1000 ms time window.

Study 1: MEG

Introduction

We recorded MEG while participants read multiple three-sentence scenarios. We orthogonally manipulated the *Plausibility* and the *Animacy* of critical nouns in the final sentence, such that they either matched or violated the animacy-based constraints of the preceding verb. This 2 × 2 design gave rise to four experimental conditions in which the critical nouns were: (1) *plausible* and *animate* (e.g. “... greeted the guests ...”), (2) *plausible* and *inanimate* (e.g. “... clamped the wires ...”), (3) *anomalous* and *animate* (e.g. “... clamped the *guests ...”), and (4) *anomalous* and *inanimate* (e.g. “... greeted the *wires ...”).

We began by analysing the data using a traditional event-related analysis stream, averaging across trials within each condition at each time point after critical word onset. To test our hypotheses, we contrasted the amplitude of the ERFs across the four conditions within two time windows of interest: 300–500 ms (to capture the N400) and 600–1000 ms (to capture post-N400 activity). Based on previous MEG studies, we expected that between 300–500 ms, the amplitude of the N400 over left temporal sensors would be larger to the semantically *anomalous* than the *plausible* nouns (Halgren et al., [2002](#); Helenius et al., [1998](#); Ihara et al., [2007](#); Maess et al., [2006](#)). Based on the prior ERP literature (Brothers et al., [2020](#); Kuperberg, [2007](#); Kuperberg et al., [2020](#); van de Meerendonk et al., [2009](#); Van Petten & Luka, [2012](#)), we also considered the possibility that we would additionally see an effect of Plausibility (*anomalous* > *plausible*) in the later 600–1000 ms time window. However, because most previous MEG studies have not reported findings past the N400 time window, the size and topography of any late effect of Plausibility that we might see on these late ERFs were unclear.

We then carried out a spatial similarity analysis on the same MEG dataset. We aimed to address two questions. The first was whether this analysis would pick up on the same underlying neural activity detected by the event-related analysis. If so, then the timing of the rise and fall of the spatial similarity *r* values should track that of both early and late event-related components. In addition, this analysis should reveal similar effects of Plausibility (*anomalous* > *plausible*) as the event-related analysis within both the N400 (300–500 ms) and post-N400 (600–1000 ms) time windows.

Our second question was whether and when the spatial similarity analysis would detect *additional* effects of the Animacy on the critical nouns that were *not* detected by the event-related analysis. As noted in the General Introduction, previous MEG studies have shown that multivariate methods can discriminate between *animate* and *inanimate* visual objects (Carlson et al., [2013](#); Cichy et al., [2014](#); Cichy & Pantazis, [2017](#); Khaligh-Razavi et al., [2018](#)). However, previous studies have not asked whether and when these methods are able to discriminate the animacy of linguistic inputs during sentence-level comprehension, or what drives the discrimination in the spatial patterns produced by the animate *versus* inanimate concepts. To address these questions, we averaged the pairwise spatial similarity values (*r* values) *within* each experimental condition at each time point following critical word onset, and constructed four spatial similarity time series. This enabled us to determine both the timing and

directionality of any spatial similarity effects, as well as to directly compare the spatial similarity time series with the event-related time series.

Methods

Participants

The MEG study was carried out at the Martinos Centre at Massachusetts General Hospital. We report data from 32 participants (16 females, mean age: 23.4; range: 18–35). Thirty-three participants originally took part in the experiment, but the data of one participant was excluded because of technical problems. Written consent was obtained from all participants following the guidelines of the Mass. General Brigham Institutional Review Board (IRB). All participants were right-handed and had normal or corrected-to-normal vision. All were native speakers of English, with no additional language exposure before the age of five. Participants were not taking psychoactive medication and were screened to exclude the presence of psychiatric and neurological disorders.

Experimental design

Each participant read 200 three-sentence discourse scenarios. The first two sentences set up a discourse context, and the third sentence had a fixed structure: an adjunct phrase (ranging between 1–4 words), a subject, verb, determiner, direct object noun (the “critical noun”), followed by three additional words. In half the scenarios, the verb in the third sentence constrained for an *animate* direct object noun (e.g. “... greeted the ...”), and in the other half, the verb constrained for an *inanimate* direct object noun (e.g. “... clamped the ...”). In half the scenarios, the critical noun matched the verb’s animacy constraints (its selectional restrictions), rendering the scenarios *plausible*, and in the other half, it violated these constraints, rendering the scenarios *anomalous*. The *animate* nouns were slightly longer (mean number of letters \pm SD: 7.51 ± 2.14) than the *inanimate* nouns (6.56 ± 2.13 , $t_{(698)} = 4.56$, $p < 0.001$, $d = 0.35$). In addition, the *animate* nouns were slightly less frequent (mean log frequency \pm SD: 0.74 ± 0.92 ; based on Brysbaert & New, 2009) than the *inanimate* nouns (0.94 ± 0.82 , $t_{(698)} = -3.07$, $p = 0.001$, $d = -0.23$). A detailed description of how the stimuli were constructed and normed can be found in Wang et al. (2020) and Kuperberg et al. (2020). As discussed there, we also varied the lexical constraint of the prior contexts such that, in each of the four conditions, half the critical nouns followed high constraint contexts and half followed low constraint contexts. However, for the purpose of the analyses reported here, we collapsed

across contextual constraint. The scenarios were counterbalanced across four lists, with approximately 50 sentences of each condition within each list.

Similarity structure of the critical words

Semantic similarity between pairs of animate and inanimate nouns. To confirm that the *animate* critical nouns were semantically more similar to each other than the *inanimate* nouns, we quantified the semantic similarity between all pairs of *animate* nouns and all pairs of *inanimate* nouns using word2vec (Mikolov et al., 2013). Word2vec is a neural network model that learns distributed semantic representations of words by training it on the Google News corpus of ~100 billion words. Each word is represented as a 300-dimensional vector, and words with similar meanings are located close to each other in this high-dimensional space. We calculated the semantic similarity between each pair of words by measuring the cosine distance between their corresponding 300-dimensional word vectors, using the Genism natural language processing library in Python.

To test for a statistical difference in semantic similarity between the *animate* and *inanimate* nouns pairs, we used a permutation approach. Specifically, we calculated the mean difference in similarity between the two groups of nouns and used this as our test statistic. We then randomly reassigned the similarity values to either the *animate* or *inanimate* group and recalculated the mean difference in similarity between the two groups. We repeated this process 1000 times to create a null distribution of mean differences. We considered our observed test statistic to be statistically significant if it fell within the top or bottom 2.5% of the null distribution. This analysis confirmed that the semantic similarity amongst pairs of *animate* nouns (mean \pm SD: 0.13 ± 0.11) was significantly greater than amongst pairs of *inanimate* nouns (mean \pm SD: 0.11 ± 0.12 , $p = 0.001$).

Similarities in word length and frequency between pairs of animate and inanimate nouns.

We also quantified the similarity in length (number of letters) and log frequency (Brysbaert & New, 2009) amongst the *animate* and *inanimate* nouns by computing the absolute difference in these values for each pair of items. For each of these variables, we took the difference between the mean similarity value (i.e. the absolute difference value) in each group of nouns as our test statistic and used a permutation test (1000 permutations) to determine if there were significant differences between the two groups. We found that the pairs of *animate* nouns were slightly *less* similar to one another (i.e. showed

greater differences) than the pairs of *inanimate* nouns, both in word length (*animate*: 2.41 ± 1.84 , *inanimate*: 2.34 ± 1.91 , $p = 0.001$) and frequency (*animate*: 1.05 ± 0.78 , *inanimate*: 0.91 ± 0.71 , $p = 0.001$).

Experimental procedure

The discourse scenarios were presented over eight blocks, with short breaks in between each block. Before the experimental session, a practice session was conducted to familiarise participants with the stimulus presentation and tasks.

In each scenario, the first two sentences were each presented as a whole for 3900 ms (100 ms interstimulus interval, ISI). Then, after a white fixation (“++++”) for 550 ms (100 ms ISI), the third sentence appeared word-by-word, with each word presented for 450 ms (100 ms ISI). Following the sentence-final word, a pink question-mark (“?”) appeared for 1400 ms (100 ms ISI), which cued participants to press one of two buttons with their left hand to indicate whether or not the scenario made sense. In addition, following a proportion of trials (24/200), which were semi-randomly distributed across blocks, participants responded to a yes/no comprehension question that appeared on the screen for 1900ms (100 ms ISI). After each trial, a blank screen appeared with a variable duration, ranging from 100 to 500 ms, followed by a green fixation (“++++”) for 900 ms (100 ms ISI) during which participants were encouraged to blink.

MEG data acquisition and preprocessing

MEG data were acquired with a Neuromag VectorView system (Elekta-Neuromag Oy, Finland) with 102 triplets of sensors. Each triplet was comprised of two orthogonal planar gradiometers and one magnetometer. EEG data were acquired at the same time as the MEG data (see Study 2). MEG signals were digitised at 1000 Hz, with an online bandpass filter of 0.03 - 300 Hz. Vertical and horizontal electrooculography (EOG) data, as well as electrocardiogram (ECG) data were collected with bipolar recordings. Impedances were kept at less than 30k Ω .

The data were preprocessed using the Minimum Norms Estimates (MNE) software package in Python (Gramfort et al., 2014). We first identified “bad” sensors in each participant in each block, resulting in the removal of seven (on average) out of the 306 MEG sensors. The data of these bad MEG sensors were later interpolated using spherical spline interpolation (Perrin et al., 1989). Eye-movement and blink artifacts were automatically removed (Gramfort et al., 2014), and Signal-Space Projection (SSP) correction (Uusitalo & Ilmoniemi, 1997) was used to correct for ECG artifacts.

After applying a bandpass filter at 0.1–30 Hz, the data were segmented into –100–1000 ms epochs relative to critical word onset. Epochs in which amplitudes exceeded pre-specified cutoff thresholds (4e-10 T/m for gradiometers and 4e-12 T for magnetometers) were removed, leaving 40 - 42 trials (on average) in each of the four conditions.

Event-related analysis

At each recording site, we combined the activity produced by the two gradiometer sensors by computing their root mean square value. We carried out an event-related analysis using Fieldtrip, an open-source Matlab toolbox (Oostenveld et al., 2011). In each participant, we calculated the evoked responses, relative to critical word onset, by averaging activity across trials within each of the four conditions at each sampling point following critical word onset, using a –100–0 ms baseline.

In MEG, the topographic distribution of the N400 effect at the head surface has been well characterised: Several previous studies have shown that it is produced primarily over left temporal sensors (e.g. Lau et al., 2013; Wang et al., 2018; Wang et al., 2023). Therefore, to compare the magnitude of the N400 across conditions, we averaged activity across nine left temporal sites (highlighted in the topographic plot in Figure 1A) between 300–500 ms. These mean values served as the dependent variable in a repeated-measures ANOVA, with two within-subject variables: Plausibility (*plausible*, *anomalous*) and Animacy (*animate*, *inanimate*).

We also carried out statistical analyses on the mean activity between 600–1000 ms. In this time window, previous studies using ERPs have reported late effects of Plausibility (*anomalous* > *plausible*) on a positive-going component known as the P600. However, most previous MEG studies have failed to examine activity within this later time window, and so the topography of any late MEG effect was unclear. We therefore took a mass univariate statistical approach to determine whether and where there were differences across conditions in evoked activity within this late time window, carrying out tests at all 102 combined gradiometer sensors. To test for the main effects of each variable, we carried out dependent-samples t-tests, collapsing across the two levels of the other variable. To test for an interaction between Plausibility and Animacy, we computed difference waveforms (*inanimate* minus *animate*) separately for the *anomalous* and *plausible* conditions, and contrasted these two difference waveforms. To correct for multiple comparisons, we used a cluster-based permutation test (Maris & Oostenveld, 2007): All spatially adjacent data samples that exceeded a preset uncorrected significance threshold of 5% were taken as a cluster,

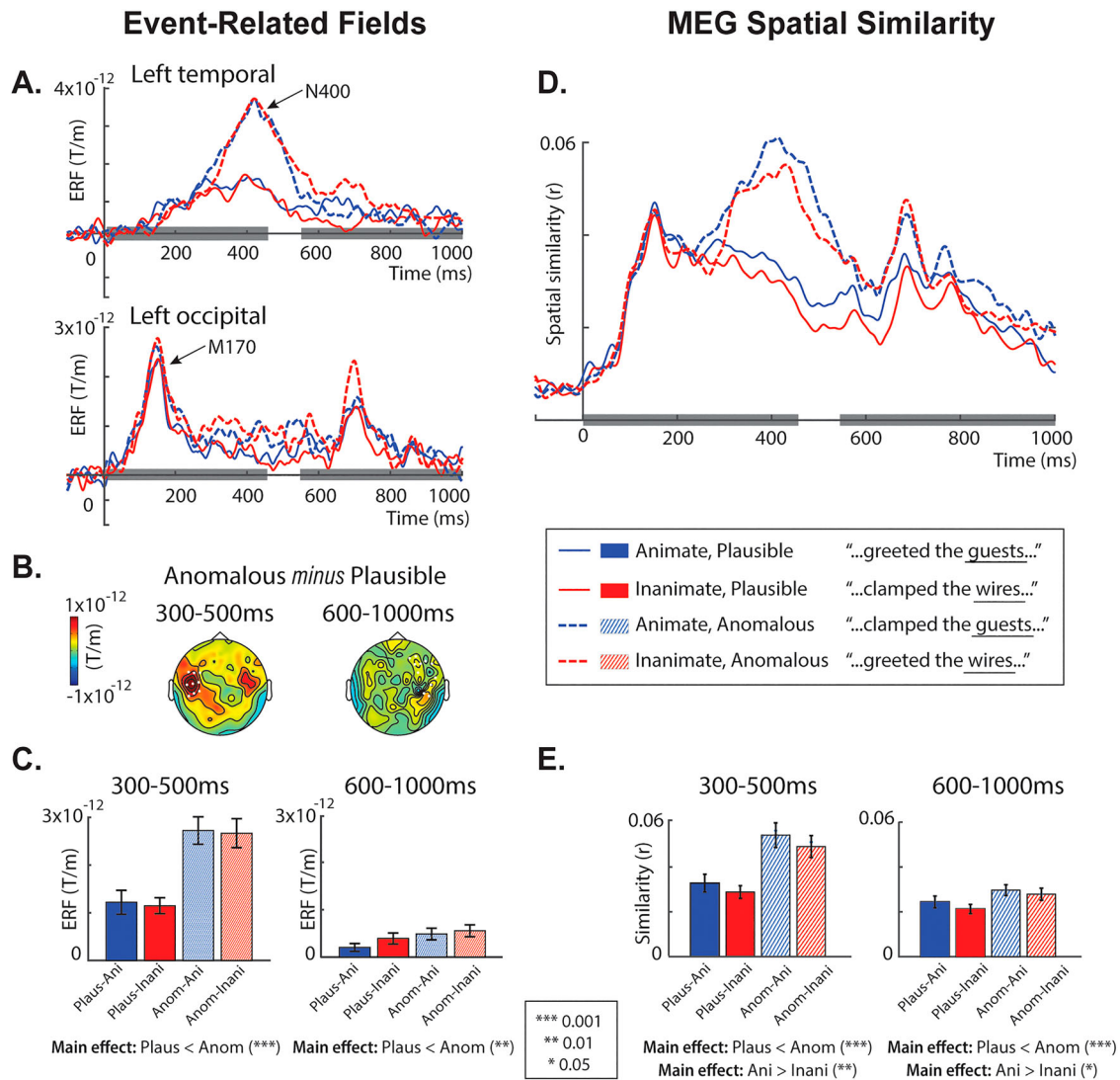


Figure 1. Results of the MEG event-related and spatial similarity analyses ($n = 32$). **A.** Grand-average ERF waveforms following the onset of critical nouns at a left temporal site (MEG1512 + 1513) and a left occipital sensor site (MEG1932 + 1933). The duration of the critical nouns (0 - 450 ms) and the subsequent word (550 - 1000 ms) are marked with gray bars on the x-axes. **B.** Topographic maps showing the distributions of the effects of Plausibility (collapsed across the *animate* and *inanimate* conditions) between 300–500 ms (left) and 600–1000 ms (right). MEG sensors that showed significant differences are indicated with white dots. **C.** Bar graphs showing the amplitude of the MEG event-related fields (ERFs) in each condition, averaged between 300–500 ms at a representative left temporal sensor (left), and between 600–1000 ms at a representative right frontal sensor (right). Error bars represent ± 1 standard error of the mean. Between 300–500 ms, the *anomalous* nouns produced a significantly larger N400 than the *plausible* nouns. The amplitude of the N400 did not differ significantly between the *animate* and *inanimate* nouns, either in the *plausible* or the *anomalous* condition. Between 600–1000 ms, the *anomalous* nouns produced slightly larger ERF than the *plausible* words at right frontal MEG sensors. Again, the magnitude of this ERF did not differ significantly between the *animate* and *inanimate* nouns in either the *plausible* or the *anomalous* condition. **D.** Grand-average spatial similarity time series following the onset of critical nouns in each of the four conditions. The duration of the critical nouns (0 - 450 ms) and the subsequent word (550 - 1000 ms) are marked with gray bars on the x-axis. **E.** Bar graphs showing the mean MEG spatial similarity values for each condition, averaged between 300–500 ms (left) and between 600–1000 ms (right). Error bars represent ± 1 standard error of the mean. Between 300–500 ms, the *anomalous* nouns produced larger spatial similarity values than the *plausible* nouns, mirroring the ERF findings. In addition, the *animate* nouns produced significantly larger spatial similarity values than the *inanimate* nouns in both *plausible* and *anomalous* conditions. The same pattern was observed between 600–1000 ms.

and individual t-statistics within each cluster were summed to yield cluster-level test statistics. We built a null distribution by randomly assigning condition labels within each participant 1000 times, calculating

cluster-level statistics for each randomisation, and entering the largest cluster-level statistic into the null distribution. We then compared the observed cluster-level test statistics against the null distribution and took any

clusters falling within the highest or lowest 2.5th percentile to be statistically significant.

Spatial similarity analysis

We carried out a spatial similarity analysis on the preprocessed MEG data using custom-written scripts. In each participant, on each trial, and at each time point following critical word onset, we extracted a vector that characterised the spatial pattern of neural activity produced across all sensors at all 102 recording sites (i.e. 306 sensors in total – 204 gradiometer and 102 magnetometer), and computed the similarity between the spatial patterns produced by each pair of critical nouns by correlating these spatial vectors (Pearson r). We then averaged these r values across all pairs of trials within each of the four conditions to create four spatial similarity time series.

For statistical analyses, we averaged these spatial similarity values across the same 300–500 ms and 600–1000 ms time windows that were used for the event-related analysis, and carried out repeated measures ANOVAs with Plausibility (*plausible*, *anomalous*) and Animacy (*animate*, *inanimate*) as within-subject variables. To visualise the data, we averaged the time series for each condition across all participants, yielding four grand-average spatial similarity time series.

Results

Event-related responses

Description of ERF time courses. As shown in Figure 1A, between 100–200 ms, the critical words in all four conditions produced an M170 ERF component (Tarkiainen et al., 1999) at bilateral occipital sensors, particularly over the left hemisphere. Between 300–500 ms, the critical words, particularly the *anomalous* words, produced an N400 that peaked at 400 ms, and was maximal over bilateral temporal sensors, particularly over the left hemisphere. Within the later 600–1000 ms time window, there was no clear evidence of a distinct waveform, although the M170 evoked by the following word was visible at occipital sensors between 650–750 ms.

300–500ms. As shown in Figure 1 (left), the N400 ERF was significantly larger to the *anomalous* than the *plausible* nouns (Main effect of Plausibility: $F_{(1,31)} = 44.33$, $p < 0.001$, $\eta^2 = 0.583$). There was no evoked effect of Animacy ($F_{(1,31)} = 2.57$, $p = 0.119$, $\eta^2 = 0.077$), or any interaction between Plausibility and Animacy ($F_{(1,31)} = 0.34$, $p = 0.854$, $\eta^2 = 0.001$).²

600–1000ms. Within this later time window, the cluster-based permutation tests showed that the *anomalous*

words produced a slightly larger ERF than the *plausible* words at right frontal MEG sensors (cluster-based permutation test: $p = 0.014$), although the magnitude of this effect was small, see Figure 1B (right). No clusters revealed a significant effect of Animacy or an interaction between Plausibility and Animacy within this late time window.

MEG spatial similarity results

Description of the MEG spatial similarity time series.

As shown in Figure 1D, the overall shape of the spatial similarity time series mirrored that of the ERF time courses: The rise and fall of the spatial similarity values between 100–200 ms corresponded to the M170 ERF component, and the rise and fall of the spatial similarity values between 300–500 ms, particularly to the *anomalous* nouns, corresponded to the N400. The rise and fall of spatial similarity values between 650–750 ms corresponded to the M170 evoked by the following word.

300–500ms. As shown in Figure 1D and E, the spatial similarity values, averaged between 300–500 ms, were significantly larger to the *anomalous* than to the *plausible* nouns (Main effect of Plausibility: $F_{(1,31)} = 50.06$, $p < 0.001$, $\eta^2 = 0.617$). In addition, both the *plausible* and *anomalous* critical words produced spatial similarity values that were slightly larger to the *animate* than to the *inanimate* nouns (Main effect of Animacy: $F_{(1,31)} = 7.86$, $p = 0.009$, $\eta^2 = 0.202$). There was no interaction between Plausibility and Animacy ($F_{(1,31)} = 0.053$, $p = 0.820$, $\eta^2 = 0.002$).

600–1000ms. Both the spatial similarity effects described above continued into the 600–1000 ms window, see Figure 1D and E. The spatial similarity values remained larger to the *anomalous* than *plausible* nouns (Main effect of Plausibility: $F_{(1,31)} = 16.384$, $p < 0.001$, $\eta^2 = 0.346$), and slightly larger to the *animate* than *inanimate* nouns, regardless of Plausibility (Main effect of Animacy: $F_{(1,31)} = 5.073$, $p = 0.032$, $\eta^2 = 0.141$). Again, there was no interaction between Plausibility and Animacy ($F_{(1,31)} = 0.38$, $p = 0.54$, $\eta^2 = 0.012$).

Follow-up exploratory analyses

To sum up, the spatial similarity analyses picked up on the same effects of Plausibility that were detected by the classic event-related analyses. In addition, the spatial similarity analyses detected effects of Animacy (*animate* > *inanimate*) that were not detected by the event-related analyses. To further explore the relationship between the univariate and multivariate findings, we carried out two additional analyses, focusing on the effect of Animacy in the plausible sentences

between 300–500 ms (which we replicated using EEG in Study 2). We addressed two questions.

First, was the failure to detect a univariate effect of Animacy within this time window simply because we constrained our univariate analysis to left temporal sensors? As discussed under Methods, the reason why we chose to average activity across these sensors is that previous MEG studies have shown that this is where the N400 effect is largest at the head surface (e.g. Lau et al., 2013; Wang et al., 2018; Wang et al., 2023). However, few MEG studies have tested for an effect of Animacy on univariate activity. It is therefore possible that by taking this analysis approach, we missed a univariate effect of Animacy over other sensors, which could have driven the multivariate effect of Animacy observed in the 300–500 ms time window. To determine whether this was the case, we tested for an effect of Animacy on the mean response between 300–500 ms in the plausible sentences at all 102 (combined) gradiometer sensors, and used cluster-based permutation tests to control for multiple comparisons (Maris & Oostenveld, 2007), i.e. the same approach that we took to carry out the univariate analysis in the 600–1000 ms time window. This mass univariate analysis failed to reveal any spatial clusters showing an effect of Animacy.

Having excluded the possibility that the spatial similarity effect of Animacy was driven by a univariate effect at any group of individual channels, we can conclude that it was driven by differences in similarity amongst the spatial patterns produced across multiple channels. This led to a second question: Would we be able to detect the same spatial similarity effect if we used only the subset of sensors that are known to contribute to the N400 evoked response? To address this question, we repeated the spatial similarity analysis, but instead of using all 306 sensors at all 120 recording sites, we used only the sensors at the nine left temporal sites that we used in the univariate analysis (i.e. 27 sensors in total – 18 gradiometer and 9 magnetometer). This subset analysis indeed revealed the same effect of Animacy on spatial similarity values: the spatial patterns produced by pairs of *animate plausible* nouns were again more similar than those produced by pairs of *inanimate plausible* nouns, $t_{(31)} = 2.29$, $p = 0.029$, $d = 0.787$.

Discussion

The event-related and spatial similarity analyses of the MEG data both revealed effects of Plausibility. The larger evoked N400 response to *anomalous* than *plausible* nouns replicates findings from previous studies (Halgren et al., 2002; Helenius et al., 1998; Ihara et al.,

2007; Maess et al., 2006; Wang et al., 2023). Here, we show that it was followed by a smaller evoked effect of Plausibility between 600–1000 ms at right frontal recording sites. We further show that these same effects of Plausibility were detected by the spatial similarity analyses.

The spatial similarity analysis additionally revealed an effect of Animacy. This effect was smaller than the effect of Plausibility, but it showed a similar time course, appearing between 300–500 ms. It was driven by a greater similarity amongst the spatial patterns for *animate* than *inanimate* nouns. These differences in spatial similarity cannot be explained by differences in similarity amongst the *animate* and *inanimate* nouns in their word length or frequency, which went in the opposite direction (see Study 1 Methods: *Similarities in word length and frequency between pairs of animate and inanimate nouns*). Instead, they mirror the difference amongst the *animate* and *inanimate* nouns in their semantic similarity structures (*animate* > *inanimate*) (Garrard et al., 2001; McRae et al., 1997; Randall et al., 2004; Zannino et al., 2006). These findings have important theoretical implications, which we consider in detail in the General Discussion.

In this Discussion section, we will focus on why the multivariate and univariate measures converged and diverged from a *methodological* perspective.

Convergence between the spatial similarity and event-related results

The results of the spatial similarity analyses converged with the results of the event-related analysis in two ways. First, the overall spatial similarity time series revealed a sequence of “waveforms” whose timing coincided with the event-related waveforms. Second, the spatial similarity analysis captured all the *effects* detected by the event-related analysis. For example, between 300–500 ms (the N400 time window), the spatial similarity values were larger for the *anomalous* than for the *plausible* nouns, mirroring the larger N400 event-related response to *anomalous* than *plausible* nouns.

From a mathematical perspective, the Pearson’s correlation between a pair of spatial vectors should be independent of each vector’s magnitude. It may therefore not be obvious why the similarity between pairs of spatial patterns, which describe neural activity across multiple channels, covaried with the amplitude of evoked activity at adjacent channels.

The reason for this is that evoked responses and spatial similarities are derived from the same very small neural signals that are embedded in noise. By “noise”, we mean any measurement that is of no

theoretical interest to the experimenter, i.e. that is neither time-locked to the experimental stimuli, nor modulated by the experimental variables of interest.³ To extract this signal from noise, *both* types of analyses rely on the assumption that the signal of interest is *consistently* produced across *multiple* trials of the same condition, while noise fluctuates randomly across trials.

This idea has been discussed extensively in the event-related literature where the goal is to determine whether and how an experimental variable influences the *strength* of the underlying neural signal produced by individual items within a particular condition. To isolate this signal, one can “filter out” noise by averaging activity across multiple items (Luck, 2014a), as we did in the current study. Alternatively, one can use a General Linear Model, which models noise within the error term (Baayen et al., 2008; Clark, 1973; Smith & Kutas, 2015). In a spatial similarity analysis, the same basic principle applies except that, instead of estimating the *strength* of the underlying neural signal produced by individual items, the goal is to estimate the *similarity* between the spatial patterns of signals produced by pairs of items. Once again, to isolate this signal, one can either filter out noise by averaging spatial similarity values across multiple pairs of items, as we did in the present study, or one can use a single trial model-based approach, such as correlating a Model dissimilarity matrix with a Neural dissimilarity matrix (e.g. Kriegeskorte et al., 2008; Nili et al., 2014).

The sensitivity of these two types of analyses to overlapping sources of signal and noise can explain both aspects of the convergence we observed between the event-related and spatial similarity results. First, it explains *why*, *regardless* of experimental condition, the rise and fall of the spatial similarity values tracked the rise and fall (relative to baseline) of the evoked M170 and N400 ERF components. At baseline, when no input was presented, there was no signal – only noise. Therefore, the expected value of both the event-related response and Pearson’s *r* was zero. However, as the information associated with each critical word became available to each level of the cortical hierarchy, it produced transient increases and decreases in the strength of the neural signal. Because each word consistently produced the same transient increases and decreases in activity within the same neural sources, the event-related analysis was able to detect a consistent rise and fall in the amplitude of the magnetic field across multiple items at particular sets of recording sensors. Because these increases and decreases in activity had consistent scalp topographies (i.e. they were consistently produced at some recording channels but not others), the spatial similarity analysis was able to

detect the same increases and decreases in signal strength as increases and decreases in the similarity in the spatial patterns produced across multiple pairs of words.

Second, the shared sensitivity of the event-related and spatial similarity analyses to common sources of signal and noise also explains why whenever the event-related analyses detected a larger response to the *anomalous* than the *plausible* words, pairs of *anomalous* nouns produced more similar patterns than pairs of *plausible* words. To understand why the evoked and spatial similarity *effects* (i.e. the *differences* between conditions) covaried, the key factor to consider is the signal-to-noise *ratio* in each condition. In the presence of signal and no noise, two identical spatial patterns would produce a Pearson’s *r* value of 1. For example, if two *anomalous* words each produced twice as much signal at the same three measurement channels as two *plausible* words, then, in the absence of noise, the spatial similarity between the two *anomalous* words would be the same as between the two *plausible* words (in both cases, Pearson’s *r* is 1). However, any noise will result in an underestimation of the true spatial similarity value ($r < 1$), such that the weaker the signal-to-noise ratio, the greater the reduction of the estimated value. Therefore, on the assumption that noise is held constant, the larger signal produced by the two *anomalous* words will result in a greater signal-to-noise ratio. Therefore, our estimation of the spatial similarity between these two *anomalous* words will be more accurate, with an *r* value that is larger (closer to the true Pearson’s *r* value of 1) than between the two *plausible* words.

These considerations raise the question of whether it is possible to carry out a multivariate analysis that is *not* sensitive to univariate activity. For example, with the current design, is there a type of multivariate analysis that would detect the effects of Animacy, but *not* the rise and fall of condition-non-specific evoked activity, or the effects of Plausibility on the evoked response?

One approach that some researchers have taken to reduce the influence of univariate activity is to remove condition-non-specific responses before carrying out a multivariate analysis. This so-called “cocktail blank removal” removes the “common” mean pattern of activity across all conditions by subtracting it from activity produced by each individual condition before computing Pearson’s correlations (Haxby et al., 2001; Op de Beeck et al., 2006). However, this method can introduce false dependencies when examining differences in spatial similarity between conditions (see Die-drichsen et al., 2011 and Garrido et al., 2013 for discussion).

A second approach, previously attempted in the fMRI literature, is to remove univariate *effects* (differences between conditions) before carrying out a multivariate analysis. Here, for each condition, within each spatial region, the mean response across all voxels and all trials is subtracted from the activity produced by each individual trial at each voxel (Jimura & Poldrack, 2012; LaRocque et al., 2013). However, this approach introduces other practical and conceptual difficulties (for discussion, see Hebart & Baker, 2018).

A third approach for dissociating multivariate and univariate activity is to use a different measure of spatial similarity – the Euclidean distance, instead of Pearson’s correlations (or Pearson’s distance, i.e. $1 - r$). In contrast to these Pearson’s measures, the Euclidean distance is invariant to condition-non-specific responses (Guggenmos et al., 2018; Walther et al., 2016). In addition, at least in fMRI, there is evidence that this measure can be insensitive to univariate *differences* between conditions (e.g. see Walther et al., 2016, Figure 8). However, to yield results that can be meaningfully interpreted, it is necessary to use a *cross-validated* Euclidean distance measure (Guggenmos et al., 2018; Walther et al., 2016).⁴ Cross-validation requires the dataset to be split into at least two partitions. While partitioning the data is feasible in studies of visual perception in which the same visual stimulus is usually presented multiple times, this would introduce significant challenges in studies of language processing. This is because we typically present multiple different linguistic items within each experimental condition, and each individual linguistic item is usually presented only once to avoid psychological confounds. Therefore, to compute the cross-validated Euclidean distance, we would need to match items across partitions on numerous potentially confounding variables. Moreover, cross-validation becomes impossible if we are interested in isolating specific neural patterns that are unique to individual items (e.g. Wang et al., 2018).

Divergence between the spatial similarity and event-related results

In contrast to the effect of Plausibility, the effect of Animacy was only detected by the multivariate spatial similarity analysis. We saw no evidence that this effect was detected by the univariate analysis, either when we collapsed across left temporal sites, or when we carried out a mass univariate analysis at all sensors. Moreover, the multivariate effect was observed, regardless of whether the absolute amplitude of the event-related responses (and spatial similarity values) was relatively small (on the *plausible* nouns) or large (on the *anomalous* nouns).

The reason why multivariate measures, such as spatial similarity analysis, are sensitive to effects that are invisible to univariate measures is that they can capture systematic differences between conditions in how *patterns* of activity, measured across multiple channels, vary across trials. In the present study, the spatial similarity analysis was able to exploit the fact that, across trials, the spatial patterns produced by pairs of *animate* nouns were consistently more similar than those produced by pairs of *inanimate* nouns.

Study 2: EEG

Introduction

In this second study, we presented participants with the same set of stimuli, but this time measured EEG rather than MEG. We had two goals. The first was to replicate the broad pattern of convergence and divergence between the evoked and spatial similarity effects using a complementary technique. The second was to determine whether, where and how the EEG and MEG univariate and multivariate measures converged and diverged.

Although MEG and EEG are both sensitive to tangential dipole sources, only EEG can detect radial dipoles (Ahlfors et al., 2010; Hämäläinen et al., 1993; Nunez, 1990). It is therefore able to detect activity that is invisible to MEG (Ahlfors et al., 2010; Siedenberg et al., 1996). In the present study, we were particularly interested in whether EEG would detect additional activity produced by *anomalous* nouns in the post-N400 time window, between 600–1000 ms. Although MEG did reveal a late effect of Plausibility over right frontal sensors in this late time window (*anomalous* > *plausible*), the size of this effect was small. In contrast, previous EEG studies have shown that between 600–1000 ms, *anomalous* words that disrupt the process of deep comprehension produce a robust posteriorly distributed positive-going ERP component, known as the “semantic P600” (Brothers et al., 2020; Kuperberg, 2007; Kuperberg et al., 2003; Kuperberg et al., 2020; van de Meerendonk et al., 2009). This ERP component is thought to be functionally related to the P300 ERP component, to which MEG is known to be relatively insensitive (Ahlfors et al., 2010; Siedenberg et al., 1996). We were therefore interested in directly comparing the magnitude of the ERP and ERF effects produced by the same set of *anomalous* (versus *plausible*) words in this late time window, as well as exploring any differences between EEG and MEG in how Animacy modulated spatial similarity in this same time window.

Finally, we were interested in whether EEG would detect an interaction between Plausibility and Animacy

in the earlier N400 time window. Although the MEG study showed no evidence of this interaction, a previous ERP study with a similar design, but using a different set of stimuli, showed that *animate* nouns produced a smaller effect of Plausibility than *inanimate* nouns (Paczynski & Kuperberg, 2011, Experiment 1). This interaction was largely driven by a less negative waveform to the *anomalous animate* than the *anomalous inanimate* nouns. In the present study, we were interested in whether this ERP finding would replicate.

Methods

We collected EEG data in two separate cohorts of participants. The first cohort was the same set of 32 participants who participated in Study 1 at Massachusetts General Hospital. In these participants, EEG data were acquired at the same time as the MEG data, using a 70-channel MEG-compatible scalp electrode system (BrainProducts, München), with an online reference at the left mastoid. Signals were digitised at 1000 Hz, with an online bandpass filter of 0.03 - 300 Hz.

The second EEG dataset was collected in a new cohort of 40 participants (19 females, mean age: 21.5 years; range: 18–32 years) at Tufts University, using a subset of the scenarios used in Study 1 (160 out of 200). Written consent was obtained from these participants, based on guidelines from the Tufts University Social, Behavioural, and Educational Research Institutional Review Board. These EEG data were acquired using Biosemi ActiveTwo acquisition system from 32 active electrodes using a modified 10/20 system montage. Signals were digitised at 512 Hz, with a bandpass of DC–104 Hz. The EEG data were referenced offline to the average of the left and right mastoid channels.

To maximise statistical power, we combined these two EEG datasets, yielding a dataset of 72 participants in total.⁵

EEG data preprocessing

The two EEG datasets were preprocessed using Fieldtrip (Oostenveld et al., 2011). Because they were acquired with different online filtering settings (0.03 - 300 Hz vs. DC - 104 Hz), we applied an offline low-pass filter of 30 Hz to the first EEG dataset and an offline band-pass filter of 0.1 - 30 Hz to the second EEG dataset. In addition, because the two datasets were acquired with different sampling rates (1000 Hz vs. 512 Hz), we down-sampled both datasets to 500 Hz. In the first dataset, we removed, on average, seven “bad” EEG channels (out of the 70 channels), which were subsequently interpolated using spherical spline interpolation (Perrin et al., 1989).

Each individual’s EEG data was segmented into –100–1000 ms epochs, relative to noun onset. We applied an Independent Component Analysis to correct for eye movement artifacts (Bell & Sejnowski, 1997; Jung et al., 2000). Remaining artifacts were removed based on visual inspection, leaving 39 - 40 artifact-free trials in each of the four conditions across the two EEG datasets.

Event-related analysis

The spatial and temporal characteristics of both the N400 and P600 ERP effects have been very well characterised in previous studies. For statistical analyses, we therefore selected spatiotemporal regions of interest based on those used in a recent study that examined both these ERP components (Kuperberg et al., 2020). We operationalised the N400 as the average activity between 300–500 ms across six centro-parietal electrode sites (C3, Cz, C4, CP1, CPz, CP2). We operationalised the P600 as the average activity between 600–1000 ms across six posterior electrode sites (P3, Pz, P4, O1, Oz, O2). For each component, we carried out a repeated-measures ANOVA with two within-subject variables: Plausibility (*plausible*, *anomalous*) and Animacy (*animate*, *inanimate*).

Spatial similarity analysis

As in the MEG study, in each participant, at each time point, we correlated the vectors that characterised the spatial patterns of activity between every pair of trials within each of the four conditions, and then averaged the resulting pairwise spatial similarity values to produce four spatial similarity time series. We then averaged these spatial similarity values between 300–500 ms and between 600–1000 ms, and tested for significant differences across conditions by carrying out 2×2 repeated measures ANOVAs as described above. To visualise the EEG spatial similarity results and compare them with the ERPs, we constructed four grand-average spatial similarity time series.

Results

ERP results

Description of ERP time courses. As shown in Figure 2A, the critical nouns evoked an N1/P1, which peaked at ~100 ms over fronto-central (N1) and posterior (P1) electrode sites, followed by a P2 component, peaking at ~200 ms over fronto-central sites. These early components were followed by the N400, which was larger to the *anomalous* than the *plausible* continuations, and was visible between 300–500 ms over centro-parietal sites. Between 600–1000 ms, the *anomalous* nouns produced a large posteriorly distributed positive-going slow

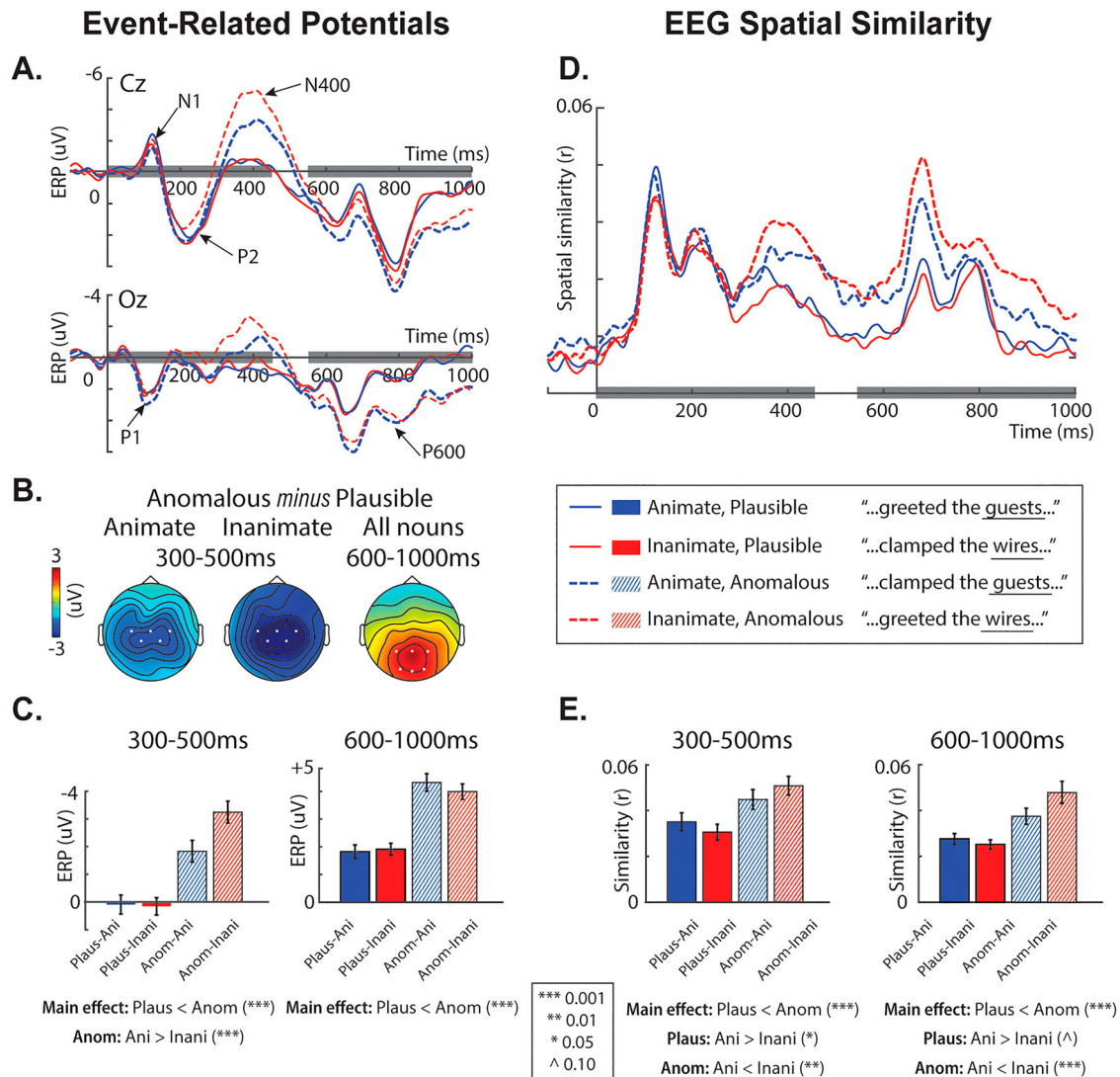


Figure 2. Results of the EEG event-related and spatial similarity analyses ($n = 72$). **A.** Grand-average ERP waveforms following the onset of critical nouns at a midline central electrode (Cz) and a midline posterior electrode (Oz). The duration of the critical nouns (0 - 450 ms) and the subsequent word (550 - 1000 ms) are marked with gray bars on the x-axes. **B.** Topographic maps show the distributions of the effects of Plausibility separately for the *animate* and *inanimate* conditions between 300–500 ms (left and middle) and for the collapsed *animate* and *inanimate* conditions between 600–1000 ms (right). The centro-parietal and posterior EEG electrodes that showed significant differences are marked with white dots. **C.** Bar graphs showing the amplitude of the event-related potentials (ERPs) in each condition, averaged between 300–500 ms across central electrode sites (left), and between 600–1000 ms across posterior electrode sites (right). Error bars represent ± 1 standard error of the mean. Between 300–500 ms, the *anomalous* nouns produced a significantly larger (more negative-going) N400 than the *plausible* nouns. There was no significant difference in the amplitude of the N400 evoked by the *animate* and *inanimate* nouns, in either the *plausible* or the *anomalous* condition. However, in the *anomalous* condition, the *animate* nouns produced a smaller negativity than the *inanimate* nouns. As discussed in the main manuscript, we suggest that this may reflect an earlier onset of the P600 produced by the *animate* (versus *inanimate*) anomalies. Between 600–1000 ms, the *anomalous* nouns produced a larger P600 than the *plausible* nouns. The amplitude of the P600 did not differ significantly between the *animate* and *inanimate* nouns in either the *plausible* or *anomalous* condition. **D.** Grand-average spatial similarity time series following the onset of critical nouns in each of the four conditions. The duration of the critical nouns (0 - 450 ms) and the subsequent word (550 - 1000 ms) are marked with gray bars on the x-axis. **E.** Bar graphs showing the mean EEG spatial similarity values in each condition, averaged between 300–500 ms (left) and between 600–1000 ms (right). Error bars represent ± 1 standard error of the mean. Between 300–500 ms, the *anomalous* nouns produced larger spatial similarity values than the *plausible* nouns. In the *plausible* condition, the *animate* nouns produced significantly larger spatial similarity values than the *inanimate* nouns. In the *anomalous* condition, however, the spatial similarity values were significantly larger in the *inanimate* than the *animate* condition. Between 600–1000 ms, the *anomalous* nouns produced significantly larger spatial similarity values than the *plausible* nouns. In the *plausible* condition, the *animate* nouns produced larger spatial similarity values than the *inanimate* nouns, although this effect only approached significance. In the *anomalous* condition, the spatial similarity values were significantly larger in the *inanimate* than the *animate* condition.

wave – the P600 effect. Within this late time window, the P1 evoked by the word following the critical noun was visible at occipital electrodes, peaking at ~650 ms.

300–500ms. As shown in Figure 2 (left), between 300–500 ms, the N400 was larger to the *anomalous* than the *plausible* nouns (Main effect of Plausibility: $F_{(1,71)} = 132.65$, $p < 0.001$, $\eta^2 = 0.651$). There was also a significant interaction between Plausibility and Animacy ($F_{(1,71)} = 15.91$, $p < 0.001$, $\eta^2 = 0.183$). This interaction was driven by a smaller negativity to the *anomalous animate* than the *anomalous inanimate* nouns ($t_{(71)} = 5.41$, $p < 0.001$, $d = 0.64$), but no difference between the *plausible animate* and *plausible inanimate* nouns ($t_{(71)} = -0.27$, $p = 0.786$, $d = -0.03$).

600–1000ms. Between 600–1000 ms, the *anomalous* nouns produced a larger late posteriorly distributed P600 than the *plausible* nouns (Main effect of Plausibility: $F_{(1,71)} = 108.48$, $p < 0.001$, $\eta^2 = 0.604$). There was no main effect of Animacy ($F_{(1,71)} = 0.51$, $p = 0.479$, $\eta^2 = 0.007$), and no interaction between Plausibility and Animacy ($F_{(1,71)} = 1.87$, $p = 0.176$, $\eta^2 = 0.026$) in this time window.

EEG spatial similarity results

Description of the EEG spatial similarity time series. As shown in Figure 2D, the shape of the EEG spatial similarity time series mirrored that of the ERP waveforms, with peaks at ~100 ms (corresponding to the N1/P1), at ~200 ms (corresponding to the P200), and at 400 ms (corresponding to the N400). We also observed a slow rise and fall of spatial similarity values between 600–1000 ms, which corresponded to the slow rise and fall of the P600 waveform. We note, however, that whereas the polarity of the P600 ERP component is opposite to that of the earlier N400 component, spatial similarity values have no polarity, and so these values are positive in both the 300–500 ms and 600–1000 ms time windows. Superimposed on the slow rise and fall of spatial similarity values between 600–1000 ms, one can see a rise and fall of spatial similarity values that mirrors the early ERP components produced by the following word.

300–500ms. Between 300–500 ms the spatial similarity values were larger for the *anomalous* than the *plausible* nouns (Main effect of Plausibility: $F_{(1,71)} = 40.33$, $p < 0.001$, $\eta^2 = 0.362$). In addition, mirroring the ERP N400 results, Plausibility interacted with Animacy ($F_{(1,71)} = 8.838$, $p = 0.004$, $\eta^2 = 0.11$) due to smaller spatial similarity values for the *anomalous animate* than the *anomalous inanimate* nouns ($t_{(71)} = -2.084$, p

$= 0.041$, $d = 0.25$). On the *plausible* nouns, the spatial similarity analysis detected an effect of Animacy that was not detected by the ERPs, with greater spatial similarity values for the *plausible animate* than the *plausible inanimate* nouns ($t_{(71)} = 2.15$, $p = 0.035$, $d = 0.25$).

600–1000ms. Mirroring the late ERP effect, the spatial similarity values were larger for the *anomalous* than the *plausible* nouns (Main effect of Plausibility: $F_{(1,71)} = 27.75$, $p < 0.001$, $\eta^2 = 0.28$). In addition, the spatial similarity analysis again detected effects of Animacy that were not detected by the event-related analysis. However, within this later 600–1000 ms time window, the effect of Animacy went in a different direction, depending on whether the nouns were *plausible* or *anomalous*, reflected by a significant interaction between Plausibility and Animacy ($F_{(1,71)} = 14.10$, $p < 0.001$, $\eta^2 = 0.17$). On the *plausible* nouns, the spatial similarity values were larger for the *animate* than for *inanimate* nouns (although this effect only approached significance: $t_{(71)} = 1.63$, $p = 0.099$, $d = 0.20$). For the *anomalies*, however, *inanimate* nouns produced larger spatial similarity values than *animate* nouns ($t_{(71)} = 3.99$, $p < 0.001$, $d = 0.47$).

Follow-up exploratory analyses

Just as for MEG, we carried out two additional analyses to further explore the relationship between the univariate and multivariate effects of Animacy between 300–500 ms after the onset of the plausible nouns.

First, to ensure that we did not miss any evoked effect of Animacy over electrode sites that were not included in our *a priori* selected regions of interest, we tested for an effect of Animacy on the mean evoked activity between 300–500 ms at *all* electrode sites, and used cluster-based permutation tests to control for multiple comparisons (Maris & Oostenveld, 2007). This analysis failed to reveal any spatial clusters showing a significant event-related effect of Animacy.

Second, we tested for a spatial similarity effect of Animacy between 300–500 ms using only the subset of centro-parietal electrode sites that we used in the univariate analysis to examine the N400. In contrast to MEG, this analysis did not reveal a significant spatial similarity effect ($t_{(71)} = 0.76$, $p = 0.45$, $d = 0.18$).

Discussion

This EEG study replicated the basic pattern of convergence and divergence between the event-related and spatial similarity results that we observed in the MEG study (Study 1). First, the spatial similarity time series again picked up on the rise-and-fall morphology of

both early and late event-related components (here, the N1/P1, P2, N400 and P600), and it again revealed effects of Plausibility (anomalous > plausible) in both the 300–500 ms and 600–1000 ms time windows. Second, the spatial similarity analysis again revealed additional effects of Animacy that were not detected by the event-related analysis: Between the 300–500 ms, the spatial patterns produced by the *plausible animate* nouns were again more similar to each other than those produced by the *plausible inanimate* nouns, mirroring their semantic similarity structures, although, as discussed further below, the effect of Animacy on spatial similarity values went in the opposite direction on the *anomalous* nouns.

In the Discussion of Study 1, we focused on why the univariate and multivariate findings converged and diverged from a methodological perspective. In this Discussion section, we will focus on why the present EEG findings converged and diverged from the previous MEG findings, also from a methodological perspective. We will discuss the theoretical significance of the full set of results in the General Discussion.

Convergence between the MEG and EEG results

Both MEG and EEG techniques are sensitive to tangentially-oriented dipoles located within cortical sulci. Therefore, the broad similarities between the MEG and EEG results are not surprising. Our finding that both MEG and EEG detect the well-known effect of Plausibility on the N400 replicates many previous studies (MEG: Halgren et al., 2002; Helenius et al., 1998; Ihara et al., 2007; Maess et al., 2006; Wang et al., 2023; ERP: Kuperberg et al., 2020; Kutas & Hillyard, 1980; Nieuwland et al., 2020). Our finding that EEG was able to detect the same multivariate effect of Animacy as MEG is consistent with previous studies reporting a similar convergence between MEG and EEG multivariate effects (Cichy & Pantazis, 2017; Wang et al., 2020). It is particularly encouraging that EEG was able to detect a spatial similarity effect of Animacy despite the fact that there were fewer EEG electrodes (70 electrodes in 32 participants, and 32 electrodes in the remaining 40 participants) than MEG sensors (306 sensors). We note, however, that in contrast to MEG, EEG failed to detect the multivariate effect of Animacy between 300–500 ms when we used only a subset of recording sites, probably because the EEG signal is more smeared than MEG due to the high resistance of the skull relative to the scalp (Geisler & Gerstein, 1961; Grynszpan & Geselowitz, 1973).

Divergence between the EEG and MEG results

In addition to the similarities between the EEG and MEG results, we saw some differences between the two

techniques in the effects that they detected on the *anomalous* nouns.

MEG was relatively insensitive to the P600 ERP effect.

Perhaps the most striking difference between EEG and MEG was in the magnitude of the late evoked effect produced by the *anomalous* (versus *plausible*) nouns between 600–1000 ms. Consistent with many previous studies (Brothers et al., 2020; Kuperberg, 2007; Kuperberg et al., 2003; Kuperberg et al., 2020; van de Meerendonk et al., 2009), EEG detected a robust positive-going P600 effect within this time window, with an effect size of $d = 1.23$ at posterior electrode sites. In contrast, the MEG evoked effect within this time window was much smaller, with an effect size of $d = 0.20$ (estimated from the right frontal sensors that showed a significant cluster in the mass univariate analysis). We also note that because MEG gradiometer sensors do not carry information about the polarity of the magnetic field (i.e. whether the magnetic flux is leaving or entering the head), the late MEG effect was of the same polarity as the earlier N400 effect.

We suggest that the reason why EEG is so much more sensitive than MEG to the P600 at the scalp surface is that it is able to detect widespread sources within frontal and parietal cortices that may contribute to this component. In principle, activity at the scalp surface can potentially stem from two types of dipole sources within these fronto-parietal regions: Radially oriented dipoles, which are produced within the crowns of gyri and the bottom of sulci, and tangentially orientated dipoles, which are produced within the walls of different sulci.⁶ MEG, however, is relatively blind to these sources. Radially orientated dipole sources typically do not result in a detectable magnetic field (Cuffin & Cohen, 1979; Hämäläinen et al., 1993), and if two tangential dipoles face in the same direction, which is common in frontal and parietal cortices, they will each produce magnetic fields that cancel each other out (Ahlfors et al., 2010; Ahlfors et al., 2010). In contrast, EEG is highly sensitive to both types of sources: in both cases, the electrical dipoles will summate to produce large electric fields (Nunez, 1990).

A similar argument has been made for why scalp-recorded EEG is so much more sensitive than MEG to the well-known P300 evoked response (the P3b; Ahlfors et al., 2010; Siedenberg et al., 1996), which is also thought to reflect widespread activity within frontal and parietal regions (Soltani & Knight, 2000). Indeed, several researchers have noted that the P600 ERP component is functionally related to the P300 (Coulson et al., 1998; Kuperberg et al., [under review](#);

Osterhout et al., 2012; Sassenhagen et al., 2014; Sassenhagen & Fiebach, 2019).

These differences between EEG and MEG have potential functional implications. When a linguistic anomaly cannot be integrated into a high-level interpretation, it is thought to trigger multiple error-based processes within the 600–1000 ms time window (see Kuperberg et al., *under review*). Thus, the differential sensitivity of MEG and EEG to distinct neuroanatomical sources between 600–1000 ms means that the two techniques may also be differentially sensitive to distinct aspects of the prolonged error-based response within this time window. We elaborate further on this idea in the General Discussion.

Between 300–500 ms, EEG but not MEG produced a smaller evoked response to anomalous animate than anomalous inanimate nouns. We also saw differences between the EEG and MEG results in the earlier 300–500 ms time window. Similar to the event-related MEG analysis, the ERP analysis showed that the *anomalous* continuations produced a larger N400 than the *plausible* continuations. However, in contrast to MEG, but replicating a previous ERP study using a similar design (Paczynski & Kuperberg, 2011, Experiment 1), we observed an *interaction* between Plausibility and Animacy because the *anomalous animate* nouns (e.g. “... clamped the *guests ...”) produced a smaller negativity than the *anomalous inanimate* nouns (e.g. “... greeted the *wires ...”).

We suggest that the smaller negativity produced by the *animate anomalous* nouns within the 300–500 ms time window reflected an early-starting positive-going P600, which was triggered because, in addition to violating the selectional constraints of the prior verb, the *animate* anomalies *also* violated the comprehender’s canonical expectations that *inanimate* direct object noun-phrases should follow *animate* subject noun-phrases – expectations based on the so-called “animacy hierarchy” (Paczynski & Kuperberg, 2011; Silverstein, 1976). In EEG, this early-starting positivity-going P600 would have artificially reduced the amplitude of the N400 produced by the *anomalous animate* nouns at the scalp surface as a result of component overlap; that is, because the N400 and late posterior positivity/P600 ERP components both have posterior scalp distributions, if they overlap in time, their opposite polarities will cancel each other out (Brouwer & Crocker, 2017; Kuperberg et al., 2007).⁷ This would explain why we saw no such interaction between Plausibility and Animacy in MEG where spatiotemporal component overlap is less of an issue (a) because the magnetic field does not have polarity, and so any ERF effect

produced between 600–1000 ms had the same polarity as the earlier N400, and (b) because MEG is relatively insensitive to the P600 effect, as discussed above (for further discussion of component overlap in EEG *versus* MEG, see Wang et al., 2023, Supplementary Materials, Discussion Section 2).

On anomalous nouns, MEG and EEG produced spatial similarity effects of Animacy that went in opposite directions.

When the critical nouns were *plausible* (i.e. when they matched the animacy-based constraints of the preceding verb), both MEG and EEG showed that pairs of the *animate* nouns produced more similar spatial patterns than pairs of the *inanimate* nouns, mirroring their semantic similarity structures (*animate* > *inanimate*). However, on the *anomalous* nouns, the MEG and EEG spatial similarity effects of Animacy went in the opposite direction: In MEG, the effect on the anomalies went in the same direction as on the *plausible* nouns (*animate* > *inanimate*). However, in EEG, the spatial patterns were more similar for the *inanimate anomalous* than the *animate anomalous* nouns; that is, instead of mirroring the similarity structure of animacy features that were encoded within the anomalous bottom-up input itself (*animate* > *inanimate*), the effect mirrored the semantic similarity structure of the animacy features that were *predicted*, based on the animacy-based constraints of the prior verb (*inanimate* > *animate*). The fact that EEG and MEG detected different spatial similarity effects of Animacy on the *anomalous* continuations is particularly notable in the late 600–100 ms window where there was no difference in the amplitude of the evoked response produced by the *animate* and *inanimate anomalous* nouns.⁸ Once again, these findings suggest that the two techniques may be differentially sensitive to distinct underlying neural sources activated by the anomalies within this late time window. We discuss the theoretical significance of this finding in the General Discussion.

General discussion

We measured MEG and EEG as participants read discourse scenarios with *animate* or *inanimate* critical words that were either *plausible* (matching the animacy-based constraints of the prior verb) or *anomalous* (mismatching these constraints). We examined the effects of both Plausibility and Animacy by carrying out both event-related and spatial similarity analyses on both datasets.

The univariate event-related analyses revealed effects of Plausibility (*anomalous* > *plausible*) both between 300–500 ms (the N400) and in the later 600–1000 ms

time window. The multivariate spatial similarity analyses picked up on the same event-related waveforms and the same effects of Plausibility. In addition, it detected effects of Animacy that were not detected by the event-related analyses.

We also saw convergence and divergence between MEG and EEG findings. Between 300–500 ms, the MEG and EEG univariate and multivariate findings largely converged. However, between 600–1000 ms, we saw differences between the effects detected by MEG and EEG in the *anomalous* condition.

In Study 1, we offered a detailed discussion of why the multivariate and the event-related measures converged and diverged from a methodological perspective. In Study 2, we offered a detailed discussion of the convergence and divergence between the MEG and EEG findings, also from a methodological perspective. In this General Discussion, we focus on the *theoretical interpretation* of our findings. We argue that the full set of results can be understood within a predictive coding framework in which animacy-based semantic features of incoming words are encoded between 300–500 ms regardless of their predictability, but this only produces prediction error and a larger univariate response when these features fail to match prior predictions. We suggest that, between 600–1000 ms, a failure of the predictive coding algorithm to settle on an *anomalous* interpretation triggers both reprocessing of the bottom-up input, as well as a re-activation of prior predictions.

We conclude by considering the *broader implications* of our findings, offering some general recommendations for how best to integrate these different analysis approaches (univariate and multivariate) and techniques (MEG and EEG) to gain deeper insights into the neural basis of language comprehension.

Theoretical interpretation

The univariate N400 effect of Plausibility

The univariate effect of Plausibility on the N400 (*anomalous* > *plausible*) replicates many previous findings (e.g. MEG: Halgren et al., 2002; Helenius et al., 1998; Ihara et al., 2007; Maess et al., 2006; Wang et al., 2023; ERP: Kuperberg et al., 2020; Kutas & Hillyard, 1980; Nieuwland et al., 2020). The N400 is thought to reflect the ease of accessing/retrieving the semantic features associated with new lexical inputs (Kuperberg, 2016; Kutas & Federmeier, 2011). If these semantic features have already been pre-activated by the prior context, then they will be easier to retrieve and so the amplitude of the N400 will be reduced. In the present study, before observing the critical noun, the brain was able to use the

animacy-based selectional constraints of the preceding verb to pre-activate upcoming semantic features (Szewczyk & Schriefers, 2013; Wang et al., 2020). For example, the context “... he *greeted* the...” led to the pre-activation of semantic features associated with <animate> entities (e.g. <can breathe>, <can move>). It was therefore relatively easy to retrieve/access the semantic features of *plausible* nouns that matched these prior predictions, resulting in a relatively small N400. However, when an *anomalous* noun was encountered, all these semantic features had to be retrieved *de novo*, resulting in a larger N400 (Kuperberg et al., 2020; Paczynski & Kuperberg, 2011, 2012; Szewczyk & Schriefers, 2013).

The multivariate effect of Animacy between 300–500ms

The effect of Animacy revealed by the multivariate similarity analysis in this time window is more novel. Although previous MEG/EEG studies have shown that representational similarity analysis is able to distinguish the animacy of visual objects (Carlson et al., 2013; Cichy et al., 2014; Cichy & Pantazis, 2017; Khaligh-Razavi et al., 2018), this is the first study to show that these methods can discriminate between *animate* and *inanimate* incoming words during real-time language comprehension. Crucially, our results also offer new insights into the underlying mechanisms and temporal dynamics of this process.

First, we show that this multivariate distinction followed the basic time course of the N400 event-related component. This provides additional evidence that 300–500 ms provides a critical time window in which the perceptual features of incoming words first make contact with distributed knowledge stored within semantic memory (Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011).

Second, we show that the spatial patterns produced by pairs of *animate* nouns were more similar than those produced by pairs of *inanimate* nouns. This is important because it supports a longstanding proposal in Cognitive Science that categorical distinctions might emerge *implicitly* from differences in patterns of covariation amongst the “microfeatures” that comprise different concepts (Hinton et al., 1986; Kemp & Tenenbaum, 2008; Rogers & McClelland, 2008), including animacy (Devlin et al., 1998; Gonnerman et al., 1997; Moss et al., 1998; Taylor et al., 2011). Specifically, a large body of previous work has established that pairs of *animate* words are more similar to one another and have more semantic features in common than pairs of *inanimate* words (Garrard et al., 2001; McRae et al., 1997; Randall et al., 2004; Zannino et al., 2006). For

example, the two *animate* words, “guests” and “sailors”, share many semantic features (e.g., <sentient>, <can move>, <can breathe>), while the two *inanimate* words, “wire” and “fruit”, are each associated with distinct semantic features (e.g. <straight> for “wire”, <tasty> for “fruit”). Using Word2Vec (Mikolov et al., 2013), we confirmed that this was the case for the current stimulus set (see Study 1 Methods: *Semantic similarity between pairs of animate and inanimate nouns*). On the assumption that the unique set of micro-features associated with each individual word is represented by a unique set of neuroanatomical sources that produce a unique spatial pattern of electric and magnetic fields at the head surface (Wang et al., 2018), the spatial patterns produced by pairs of *animate* words will be more similar than those produced by pairs of *inanimate* words.

A spatial similarity analysis at the scalp surface cannot tell us which underlying sources produced the effect. However, we know that within the N400 time window, lexical representations, which are thought to be represented in multiple regions of the left temporal cortex (Hirshorn et al., 2016; Lau et al., 2008; Woolnough et al., 2021), play a crucial role in mapping orthographic and phonological forms onto semantic features (Wang et al., 2023). These semantic features are thought to be represented within multimodal regions that are widely distributed across the cortex (Huth et al., 2016; Martin & Chao, 2001). They may also be “bound together” as conceptual representations within the anterior and medial temporal lobe (Lambon-Ralph et al., 2017), and there is some evidence that semantic features may be topographically coded within these temporal regions (e.g. Chao et al., 1999; Haxby et al., 2001). We therefore speculate that the scalp-recorded spatial patterns stemmed from *both* sources within the left temporal lobe as well as more widespread cortical regions.

The idea that these spatial patterns stemmed from left temporal regions would be consistent with our finding in the MEG study that the same spatial similarity effect was produced when we repeated the analysis using only the subset of left temporal MEG sensors that reflect lexico-semantic activity between 300–500 ms within the left temporal cortex (Wang et al., 2023). It would also be consistent with evidence that patients with localised temporal lesions can show animacy-based category specific deficits (e.g. Warrington & McCarthy, 1987; Warrington & Shallice, 1984), and that the animacy of objects and words can be decoded based on spatial patterns produced within the left temporal lobe (Devereux et al., 2013; Fairhall & Caramazza, 2013; Liuzzi et al., 2020).

The idea that these spatial patterns stemmed from widely distributed cortical regions would be consistent with evidence that animacy-based categorical structure can emerge from patterns of covariation across semantic features within a widely distributed semantic system (Taylor et al., 2011; Tyler & Moss, 2001), including evidence for animacy-based categorisation deficits in patients with non-focal neuropathologies such as mid-stage Alzheimer’s disease (Gonnerman et al., 1997; see also Devlin et al., 1998 for discussion).

Better together: synthesising the univariate and multivariate results between 300–500ms

Taken together, therefore, the univariate and multivariate results complement each other. Both methods tell us that during sentence comprehension, the brain encodes or accesses the semantic features of incoming words between 300–500 ms after their onset. The spatial similarity results tell us that this occurs *regardless* of their predictability. The event-related results tell us that only the process of encoding *unpredicted* implausible semantic features produces changes in the overall strength of neural activity (*anomalous* > *plausible*). We now need a computational theory that can explain both sets of findings.

One such theory is *predictive coding* – a computational account of how the brain approximates Bayesian inference (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999). According to predictive coding, as new bottom-up information reaches each level of the cortical hierarchy, it activates two types of connectionist units – “state units” and “error units”. State units that encode contextual information at higher levels of the hierarchy continually generate predictions that attempt to “explain” expected information at the level below. Any new bottom-up information that matches these top-down predictions is suppressed within lower-level error units, while any residual unpredicted information that cannot be explained activates these error units, producing “prediction error” (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999; Spratling, 2016). This lower-level prediction error is then used to update the information encoded within the higher-level state units, which, in turn, produce more accurate top-down predictions. By minimising prediction error over multiple iterations of the predictive coding algorithm, the brain converges on the set of representations that best explain the bottom-up input.

Several researchers have suggested that the N400 may reflect prediction error produced during predictive coding (Bornkessel-Schlesewsky & Schlesewsky, 2019; Kuperberg et al., 2020; Rabovsky & McRae, 2014; Xiang & Kuperberg, 2015). In recent work, we have used a

predictive coding computational model to show that the effect of predictability on the N400, as well as its rise-and-fall morphology, can be simulated by the magnitude of prediction error produced by error units at lexical and semantic levels of representation, i.e. lexico-semantic prediction error (Nour Eddine, 2021; Nour Eddine et al., 2022).⁹

To understand how this framework can explain both the univariate and multivariate findings between 300–500 ms in the present study, consider the state of the language system before the incoming critical word was encountered. At this point, the brain will have pre-activated expected upcoming animacy-based semantic features, based on the animacy constraints of the prior verb. Consistent with this idea, we have previously shown that when a verb constrains for upcoming *animate* nouns (e.g. “... he greeted the ...”), the brain produces spatial patterns that are more similar than when a verb constrains for upcoming *inanimate* nouns (e.g. “... he clamped the ...”) (Wang et al., 2020).

According to predictive coding, upon encountering a *plausible* noun that confirms these prior predictions, these animacy-based features are reinstated between 300–500 ms within state units. However, any activation of error units (i.e. prediction error) is immediately suppressed by the prior predictions. This explains why we found a spatial similarity effect of Animacy (*animate* > *inanimate*) between 300–500 ms on the *plausible* critical words (reflecting state activity), even though these words produced a relatively small event-related N400 response (reflecting a relatively small prediction error within error units). In contrast, upon encountering an *anomalous* noun whose animacy features cannot be explained by prior predictions, at the same time as state units are accessing the animacy features associated with the bottom-up input, error units produce prediction error because this newly encoded semantic information is not suppressed by prior predictions. This can explain why the *anomalous* inputs produced a larger N400 evoked response than the *plausible* inputs.

Responses to the anomalous nouns between 600–1000ms: divergence between MEG and EEG findings

In the present study, the implausible nouns were not just unpredictable; they were *anomalous*, and therefore could not be incorporated into the comprehender’s high-level interpretation between 300–500 ms. When a linguistic error disrupts the process of comprehension, this can trigger a number of prolonged processes between 600–1000 ms. We found that MEG and EEG diverged in both their univariate and multivariate responses to the anomalies in this later time window,

raising the possibility that these two techniques were differentially sensitive to different aspects of this late error-based activity.

Between 600–1000 ms, MEG detected only a small evoked response to the anomalies. In previous work, we source localised this late evoked response to lower-level cortical regions, providing evidence that it reflects bottom-up *reprocessing* of the anomalous input (Wang et al., 2023). Consistent with this idea, the MEG multivariate analysis revealed more similar spatial patterns to the *anomalous animate* than *inanimate* nouns within this 600–1000 ms time window. As discussed above, this mirrors the internal semantic similarity structure of *anomalous* input. These findings therefore suggest that the animacy-based semantic features associated with the anomalous *bottom-up inputs* were re-activated within this later time window, and that MEG was sensitive to this late bottom-up activity.

In contrast, between 600–1000 ms, the EEG detected a much larger univariate ERP effect on the anomalies – the posteriorly distributed semantic P600 ERP component (Brothers et al., 2020; Brothers et al., 2022; Kuperberg, 2007; Kuperberg et al., 2003; Kuperberg et al., 2020; Kuperberg et al., [under review](#); van de Meerendonk et al., 2009). Moreover, within this later time window, the EEG multivariate effect of Animacy on the *anomalous* nouns went in the opposite direction, with more similar patterns observed in the *anomalous inanimate* than the *anomalous animate* condition; that is, within this later time window, instead of mirroring the internal semantic similarity structure of observed *anomalous animate* and *inanimate* nouns, these patterns mirrored the semantic similarity structure of the patterns that were *expected*, based on the constraints of the prior verb (e.g. “... he greeted the ... <animate>”, “... he clamped the ... <inanimate>”). These findings suggest that the incorrect *prior predictions* were *reactivated* during this time window and that EEG was relatively more sensitive than MEG to this late feedback activity.

The P600 ERP component itself may function to track these conflicting sources of bottom-up and top-down information in this late time window. Specifically, we have recently argued that the P600 reflects a decision variable that accumulates evidence that the original disruption in comprehension stemmed from an external error in the bottom-up input, as opposed to an internal error in processing (Kuperberg et al., [under review](#)). This type of confidence tracking may play an important role in ensuring that the brain responds optimally to linguistic errors over multiple time scales. Similar evidence accumulation decision-making mechanisms have been linked to the well-known P300 ERP component (the P3b or pre-decisional centro-parietal positivity; see

Twomey et al., 2015), which, as noted earlier, shares several functional properties with the P600 (Coulson et al., 1998; Osterhout et al., 2012; Sassenhagen et al., 2014; Sassenhagen & Fiebach, 2019).¹⁰ It is therefore particularly noteworthy that, like the P600, the P300 event-related response is larger in EEG than in MEG (Siedenberg et al., 1996). This may be because these decision-making mechanisms are implemented within widespread fronto-parietal cortices (Brosnan et al., 2020), and MEG is relatively insensitive to these sources (Ahlfors et al., 2010; see Study 2 for discussion).

Better together: synthesising the univariate and multivariate results of the MEG and EEG datasets between 600–1000ms

Once again, therefore, the results of these different approaches complement each other. We again suggest that this full set of findings can be understood by appealing to the principles of predictive coding. In the case of *anomalous* inputs, however, the predictive coding algorithm would *fail* to converge between 300–500 ms, leading to two downstream consequences within the later 600–1000 ms time window.

The first consequence is that higher levels of representation will fail to produce accurate top-down predictions that switch off lower-level prediction error. This can explain why the *anomalous* bottom-up input was reprocessed in this later 600–1000 ms time window (see Wang et al., 2023 for discussion), as detected by the univariate and multivariate MEG findings.

The second consequence is that the prior context will continue to generate inaccurate top-down predictions based on the preceding verb. For example, in the *anomalous* sentence, “... he greeted the *wires...”, the brain would continue to generate predictions of <animate> features, while in the *anomalous* sentence, “... he clamped the *guests...”, the brain would continue to generative predictions of <inanimate> features. This can explain why the EEG spatial similarity effect reflected the internal semantic similarity structure of these incorrect predictions, rather than that of the observed *anomalous* bottom-up *anomalous* inputs. As discussed above, the simultaneous activation of conflicting top-down and bottom-up information within this time window would inform the evidence accumulation process reflected by the P600 ERP component.

General implications

Our findings have several broad implications that go beyond the specific set of results reported in this study. In this final section, we summarise these

implications. We offer three take-home messages, and provide recommendations for how to effectively use multivariate together with univariate methods, as well as MEG together with EEG, to deepen our understanding of language processing.

Interpret multivariate results in the light of the large existing univariate ERP and MEG literatures

Our first take-home message is that it is critical to interpret any spatial similarity MEG/EEG results in the light of the large existing univariate ERP and MEG literatures. This is because, as we show clearly in the present study, any multivariate measures estimated using Pearson’s r will covary with the magnitude of well-known event-related components like the N400 and P600.

Perhaps because of the belief that newer multivariate measures detect “representational information” that is functionally distinct from the “cognitive process” detected by event-related responses (see further below), some multivariate MEG/EEG language studies make only minimal reference to the ERP/ERF literature. However, a failure to fully consider this literature not only misses the opportunity to capitalise on a large body of existing knowledge, but can also lead to false inferences about the functional significance of multivariate effects. For example, we know from the ERP literature that the N400 event-related component is highly sensitive to numerous lexical and semantic variables (e.g. word frequency: Rugg, 1990; Van Petten & Kutas, 1990; word concreteness or semantic richness: Amsel, 2011; Holcomb et al., 1999; Kounios & Holcomb, 1994; Rabovsky et al., 2012). In addition, during sentence comprehension, the N400 evoked by each incoming word is highly sensitive to its relationship with the prior context, including its lexical predictability (DeLong et al., 2005; Federmeier et al., 2007; Kutas & Hillyard, 1984), its plausibility (Kuperberg et al., 2020; Nieuwland et al., 2020), its semantic overlap with a incorrectly predicted word (e.g. Federmeier & Kutas, 1999), and its semantic association with prior words in the local context (e.g. Van Petten et al., 1997). All these event-related effects are likely to be detected by multivariate measures that use Pearson’s r . Therefore, whenever one sees a multivariate effect between 300–500 ms following stimulus onset, it is important to consider the possibility that it is driven by the effects of these variables on the well-known N400 component.

As discussed in detail in the Discussion section of Study 1, there is no simple way of capturing multivariate effects that are not driven by univariate effects. We therefore recommend that, when using Pearson correlations or Pearson distance measures as a similarity metric, the best way to determine which multivariate

effects are and are not driven by univariate effects is to carry out *both* a multivariate and univariate analysis on the same dataset (see Haxby et al., 2001 and Aly & Turk-Browne, 2015, for similar recommendations using fMRI), ideally using analogous methods (as in the present study), and qualitatively comparing the spatial similarity and evoked response time series.

Abandon the assumption of a one-to-one mapping between analytic approach and functional interpretation

Our second take-home message is that we should abandon any assumption of a simple one-to-one mapping between analytic approach (univariate *versus* multivariate) and functional interpretation (“cognitive process” *versus* “representational content”). This distinction was first proposed in the fMRI literature (Kriegeskorte, 2011) where it was assumed that “processing” equated to the engagement of metabolically demanding cognitive resources that operated upon “representational information”, which remained static over time.

While this assumption may sometimes be valid for interpreting the sluggish hemodynamic response (Jimura & Poldrack, 2012), any clear-cut distinction between “process” and “representational information” falls apart when we come to interpret MEG/EEG activity produced during real-time language comprehension in which each new piece of information is rapidly encoded at multiple levels of linguistic representation, and so, at each time point following stimulus onset, the activation (encoding) of this information corresponds to its processing. For example, as should be clear from our discussion of the N400 component, within a predictive coding framework, the difficulty of semantic “processing” (or “retrieval” or “access”) between 300–500 ms corresponds directly to the amount of unpredicted semantic representational *information* that is being encoded (see Kuperberg, 2016; Nour Eddine, 2021; Nour Eddine et al., 2023).

We therefore suggest that when interpreting EEG/MEG data, we should begin with the assumption that *both* univariate and multivariate measures can tell us whether and when the brain encodes information within new bottom-up input, and that the critical question that we should be asking is whether encoding (or, equivalently accessing or retrieving) this information is, or is not, accompanied by an overall increase in neural signal. As we show in the present study, the answer to this question can only be provided by carrying out *both* types of analyses. Our spatial similarity results alone were unable to tell us which of the effects of Plausibility or Animacy were accompanied by changes in response magnitude. It was only through the event-

related analysis that we learned that Plausibility but not Animacy modulated the amplitude of neural activity produced between 300–500 ms. This, in turn, provided important constraints on our theoretical interpretation.

Do not assume that MEG and EEG always reflect precisely the same underlying neural sources and mechanisms

Our third take-home message is that we cannot assume that MEG and EEG results will always converge. In the present study, this lack of convergence was particularly clear between 600–1000 ms (but, because of component overlap in EEG but not MEG, it can also impact the earlier 300–500 ms time window, see Discussion in Study 2; and see also Wang et al., 2023, Supplementary Materials, Discussion Section 2).

Between 600–1000 ms, MEG and EEG differed in both the univariate and multivariate effects in the *anomalous* condition, providing evidence that the two techniques were differentially sensitive to distinct neural sources that were activated in this late time window. This in turn, provided important theoretical constraints on our interpretation of the prolonged response produced by linguistic errors. We suggested that MEG and EEG may be differentially sensitive to the reprocessing of the bottom-up input and the re-activation of prior predictions respectively. If future studies confirm that EEG and MEG are indeed differentially sensitive to feedback *versus* feedforward activity within this later time window, then this would provide essential information that constrains the interpretation of future investigations using both techniques to study language comprehension.

Conclusion

In conclusion, our findings suggest that a comprehensive understanding of the neural and computational basis of language processing can only be achieved through the combined use of univariate and multivariate methods, along with different neuroimaging techniques. Our results illustrate the importance of interpreting the results of these different tools *together* to provide deeper insights into the complex mechanisms underlying language comprehension.

Notes

1. In the MEG literature, the N400 is sometimes referred to as the M400 or N400m.
2. As noted under Methods, there were small differences between the *animate* and *inanimate* nouns in their length and frequency. Effects of word length can

influence earlier ERP components (Hauk & Pulvermüller, 2004; Osterhout et al., 1997; Wydell et al., 2003), and differences in frequency can modulate the N400 (Dambacher et al., 2006; Payne et al., 2015; Van Petten & Kutas, 1990). These differences, however, are unlikely to explain the *absence* of a univariate effect of Animacy between 300-500ms.

3. This may include (a) spontaneous ongoing neural activity, (b) activity stemming from other physiological sources, and (c) noise from the environment or from our measurement instruments (Luck, 2014b). We also note that while spatial similarity and event-related analyses are sensitive to many of the same sources of noise, there are some differences: In an event-related analysis, spontaneous neural activity that is not phase-locked to an event of interest is *always* considered “noise”. In a spatial similarity analysis, however, non-phase-locked activity can, under certain circumstances, contribute to increases in spatial similarity. For example, if the phase value of a waveform measured at channel 1 varies across trials, then the event-related response at this channel will always be small. However, if the relative differences in amplitude across the three channels (e.g. $1 < 2 < 3$) are consistent across trials, then this will give rise to an increase in spatial similarity.
4. This is because without cross validation, estimates of the true Euclidean distance become increasingly more inaccurate (larger) with increases in noise. As discussed above, this is also true of Pearson’s distance estimates. However, the Pearson’s distance between two entirely random patterns has a maximal value of 1 (or, equivalently, a Pearson’s r value of 0), regardless of noise level. In contrast, the Euclidean distance is non-negative and unbounded. Therefore, without cross-validation, the noise inflation of Euclidean distance is greater than that of the Pearson’s distance. Therefore, without cross-validation, Euclidean distance values are far less interpretable than Pearson’s distance values (see Guggenmos et al., 2018; Walther et al., 2016 for discussion).
5. We also carried out supplementary analyses to check that the event-related and spatial similarity effects were similar between the two EEG datasets. For both types of analyses, the pattern of findings across the two datasets was indeed qualitatively similar.
6. Activity is also produced by tangentially orientated dipoles within the walls of a *single* sulcus. However, these dipoles frequently face in opposite directions, canceling each other out, and so they are not detected at the head surface by *either* MEG or EEG.
7. We note that this interpretation is slightly different from the one that we offered in our previous ERP study which used a similar design but a different set of stimuli (Paczynski & Kuperberg, 2011, Experiment 1). In that study, we also reported a smaller negativity to *animate* than *inanimate* anomalies in the N400 time window, and we also attributed this difference to the fact that the *animate* but not the *inanimate* anomalies violated expectations based on the animacy hierarchy. In that study, we suggested that this led to reduced *semantic* processing of the *anomalous animate* (versus *anomalous inanimate*) nouns on the N400 component itself.

However, because in the present investigation, Study 1 showed that the amplitude of the MEG N400 was just as large to the *anomalous animate* as to the *anomalous inanimate* nouns, we think that a component overlap explanation for the attenuation of the ERP N400 to the *animate anomalous* nouns is more likely.

8. In the earlier 300-500ms time window, the smaller EEG spatial similarity values for the *anomalous animate* (versus *anomalous inanimate*) nouns can be explained by the smaller evoked response to these *animate* continuations, as a result of an early start of the P600 component, as discussed above.
9. In predictive coding, “prediction error” is computed as an inherent component of the inference algorithm – the process of inferring semantic features from a word’s linguistic form. This differs from how prediction error has been simulated in other computational models of the N400, where the error is computed outside the model’s architecture (Fitz & Chang, 2019; Rabovsky & McRae, 2014). It also differs from computational models that have simulated the N400 as the total activity (Cheyette & Plaut, 2017; Laszlo & Plaut, 2012) or the total change in activity produced by a single set of units (Brouwer et al., 2017; Rabovsky et al., 2018; see Nour Eddine et al., 2022 for discussion and review). We also note that, although in the present study “prediction error” was produced by a linguistic “error” (an *anomalous* input), a linguistic error is not necessary to produce prediction error (or an N400) within a predictive coding framework. According to this framework, prediction error is produced at the lexico-semantic level *whenever* incoming lexico-semantic information cannot be explained by prior lexico-semantic predictions. This will result in a larger N400 to unexpected but plausible words, compared to expected words. We have recently argued that, in the case of anomalous inputs, prediction error is additionally generated at a higher event level of representation because the newly inferred anomalous event cannot be explained by still higher-level predictions based on longer-term real-world knowledge (Wang et al., 2023).
10. Note, however, that unlike the P300, which is thought to index the accumulation of raw evidence to make an initial decision, we have argued that the P600 reflects the accumulation of evidence that is gathered after the initial disruption in comprehension. In this sense, this component may bear more resemblance to later positivities within the P300 family of ERP components (Boldt & Yeung, 2015; Desender et al., 2019; Desender et al., 2021; Murphy et al., 2015; Steinhäuser & Yeung, 2010), which are also thought to track the brain’s confidence in relation to a previous choice (Boldt & Yeung, 2015; Desender et al., 2019; Murphy et al., 2015; see Desender et al., 2021 for a review).

Acknowledgments

We thank Trevor Brothers and Victoria Sharpe for helpful discussion. We also thank Jeff Stibel for his support of Drs. Kuperberg and Wang.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Institutes of Health: [Grant Number R01HD082527].

References

- Ahlfors, S. P., Han, J., Belliveau, J. W., & Hämaläinen, M. S. (2010). Sensitivity of MEG and EEG to source orientation. *Brain Topography*, 23(3), 227–232. <https://doi.org/10.1007/s10548-010-0154-x>
- Ahlfors, S. P., Han, J., Lin, F. H., Witzel, T., Belliveau, J. W., Hämaläinen, M. S., & Halgren, E. (2010). Cancellation of EEG and MEG signals generated by extended and distributed sources. *Human Brain Mapping*, 31(1), 140–149.
- Aly, M., & Turk-Browne, N. B. (2015). Attention stabilizes representations in the human hippocampus. *Cerebral Cortex*, 26(2), 783–796. <https://doi.org/10.1093/cercor/bhv041>
- Amsel, B. D. (2011). Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. *Neuropsychologia*, 49(5), 970–983. <https://doi.org/10.1016/j.neuropsychologia.2011.01.003>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bell, A. J., & Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vision Research*, 37(23), 3327–3338. [https://doi.org/10.1016/S0042-6989\(97\)00121-1](https://doi.org/10.1016/S0042-6989(97)00121-1)
- Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *The Journal of Neuroscience*, 35(8), 3478–3484. <https://doi.org/10.1523/JNEUROSCI.0797-14.2015>
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, 10, 298. <https://doi.org/10.3389/fpsyg.2019.00298>
- Brosnan, M. B., Sabarodien, K., Silk, T., Genc, S., Newman, D. P., Loughnane, G. M., Fornito, A., O’Connell, R. G., & Bellgrove, M. A. (2020). Evidence accumulation during perceptual decisions in humans varies as a function of dorsal frontoparietal organization. *Nature Human Behaviour*, 4(8), 844–855. <https://doi.org/10.1038/s41562-020-0863-4>
- Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. *Neurobiology of Language*, 1(1), 135–160. https://doi.org/10.1162/nol_a_00006
- Brothers, T., Zeitlin, M., Choi Perrachione, A., Choi, C., & Kuperberg, G. (2022). Domain-general conflict monitoring predicts neural and behavioral indices of linguistic error processing during reading comprehension. *Journal of Experimental Psychology: General*, 151(7), 1502–1519. <https://doi.org/10.1037/xge0001130>
- Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology*, 8, 1327. <https://doi.org/10.3389/fpsyg.2017.01327>
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41, 1318–1352. <https://doi.org/10.1111/cogs.12461>
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>
- Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6), 407–420. <https://doi.org/10.1038/nrn3241>
- Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, 13, <https://doi.org/10.1167/13.10.1>
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2(10), 913–919. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10491613 http://www.nature.com/neuro/journal/v2/n10/pdf/nn1099_913.pdf. <https://doi.org/10.1038/13217>
- Cheyette, S. J., & Plaut, D. C. (2017). Modeling the N400 ERP component as transient semantic over-activation within a neural network model of word comprehension. *Cognition*, 162, 153–166. <https://doi.org/10.1016/j.cognition.2016.10.016>
- Cichy, R. M., & Pantazis, D. (2017). Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *Neuroimage*, 158, 441–454. <https://doi.org/10.1016/j.neuroimage.2017.07.023>
- Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, 17(3), 455–462. <https://doi.org/10.1038/nn.3635>
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–359. [https://doi.org/10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3)
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13(1), 21–58. <https://doi.org/10.1080/016909698386582>
- Cuffin, B. N., & Cohen, D. (1979). Comparison of the magnetoencephalogram and electroencephalogram. *Electroencephalography and Clinical Neurophysiology*, 47(2), 132–146. [https://doi.org/10.1016/0013-4694\(79\)90215-3](https://doi.org/10.1016/0013-4694(79)90215-3)
- Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084(1), 89–103. <https://doi.org/10.1016/j.brainres.2006.02.010>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. <https://doi.org/10.1038/nn1504>
- Desender, K., Murphy, P., Boldt, A., Verguts, T., & Yeung, N. (2019). A post-decisional neural marker of confidence predicts information-seeking in decision-making. *The Journal of Neuroscience*, 39(17), 3309–3319. <https://doi.org/10.1523/JNEUROSCI.2620-18.2019>

- Desender, K., Ridderinkhof, K. R., & Murphy, P. (2021). Understanding neural signals of post-decisional performance monitoring: An integrative review. *Elife*, *10*, e67556. <https://doi.org/10.7554/eLife.67556>
- Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *The Journal of Neuroscience*, *33*(48), 18906–18916. <https://doi.org/10.1523/JNEUROSCI.3809-13.2013>
- Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience*, *10*(1), 77–94. <https://www.ncbi.nlm.nih.gov/pubmed/9526084>. <https://doi.org/10.1162/089892998563798>
- Diedrichsen, J., Ridgway, G. R., Friston, K. J., & Wiestler, T. (2011). Comparing the similarity and spatial structure of neural representations: A pattern-component model. *Neuroimage*, *55*(4), 1665–1678. <https://doi.org/10.1016/j.neuroimage.2011.01.044>
- Dikker, S., Assaneo, M. F., Gwilliams, L., Wang, L., & Kösem, A. (2020). Magnetoencephalography and language. *Neuroimaging Clinics of North America*, *30*(2), 229–238. <https://doi.org/10.1016/j.nic.2020.01.004>
- Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *Journal of Neuroscience*, *33*(25), 10552–10558. <https://doi.org/10.1523/JNEUROSCI.0051-13.2013>
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495. <https://doi.org/10.1006/jmla.1999.2660>
- Federmeier, K. D., & Laszlo, S. (2009). Psychology of learning and motivation. *Psychology of Learning and Motivation*, *51*, 1–44. [https://doi.org/10.1016/S0079-7421\(09\)51001-8](https://doi.org/10.1016/S0079-7421(09)51001-8)
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. <https://doi.org/10.1016/j.brainres.2006.06.101>
- Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52. <https://doi.org/10.1016/j.cogpsych.2019.03.002>
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360* (1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Garrard, P., Lambon-Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174. <https://doi.org/10.1080/02643290125857>
- Garrido, L., Vaziri-Pashkam, M., Nakayama, K., & Wilmer, J. (2013). The consequences of subtracting the mean pattern in fMRI multivariate correlation analyses. *Frontiers in Neuroscience*, *7*, 174. <https://doi.org/10.3389/fnins.2013.00174>
- Geisler, C. D., & Gerstein, G. L. (1961). The surface EEG in relation to its sources. *Electroencephalography and Clinical Neurophysiology*, *13*(6), 927–934. [https://doi.org/10.1016/0013-4694\(61\)90199-7](https://doi.org/10.1016/0013-4694(61)90199-7)
- Gonnerman, L. M., Andersen, E. S., Devlin, J. T., Kempler, D., & Seidenberg, M. S. (1997). Double dissociation of semantic categories in Alzheimer's disease. *Brain and Language*, *57* (2), 254–279. <https://doi.org/10.1006/brln.1997.1752>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460. <https://doi.org/10.1016/j.neuroimage.2013.10.027>
- Grynspan, F., & Geselowitz, D. B. (1973). Model studies of the magnetocardiogram. *Biophysical Journal*, *13*(9), 911–925. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1484372/pdf/biophysj00707-0056.pdf>. [https://doi.org/10.1016/S0006-3495\(73\)86034-5](https://doi.org/10.1016/S0006-3495(73)86034-5)
- Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *Neuroimage*, *173*, 434–447. <https://doi.org/10.1016/j.neuroimage.2018.02.044>
- Hagoort, P. (2013). Muc (memory, unification, control) and beyond. *Frontiers in Psychology*, *4*, 416. <https://doi.org/10.3389/fpsyg.2013.00416>
- Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*, *17*(3), 1101–1116. <https://doi.org/10.1006/nimg.2002.1268>
- Hämäläinen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J. E. T., & Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*(2), 413–497. <https://doi.org/10.1103/RevModPhys.65.413>
- Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090–1103. <https://doi.org/10.1016/j.clinph.2003.12.020>
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. <https://doi.org/10.1126/science.1063736>
- Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *Neuroimage*, *180*, 4–18. <https://doi.org/10.1016/j.neuroimage.2017.08.005>
- Helenius, P., Salmelin, R., Service, E., & Connolly, J. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, *121*(6), 1133–1142. <https://doi.org/10.1093/brain/121.6.1133>
- Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations* (pp. 77–109). MIT Press.
- Hirshorn, E. A., Li, Y., Ward, M. J., Richardson, R. M., Fiez, J. A., & Ghuman, A. S. (2016). Decoding and disrupting left midfusiform gyrus activity during word reading. *Proceedings of the National Academy of Sciences*, *113*(29), 8162–8167. <https://doi.org/10.1073/pnas.1604126113>
- Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning*,

- Memory, and Cognition*, 25(3), 721–742. <https://doi.org/10.1037/0278-7393.25.3.721>
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453–458. <https://doi.org/10.1038/nature17637>
- Ihara, A., Hayakawa, T., Wei, Q., Munetsuna, S., & Fujimaki, N. (2007). Lexical access and selection of contextually appropriate meaning for ambiguous words. *Neuroimage*, 38(3), 576–588. <https://doi.org/10.1016/j.neuroimage.2007.07.047>
- Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, 50(4), 544–552. <https://doi.org/10.1016/j.neuropsychologia.2011.11.007>
- Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2), 163–178. <https://doi.org/10.1111/1469-8986.3720163>
- Kappenman, E. S., & Luck, S. J. (2012). ERP components: The ups and downs of brainwave recordings. In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 3–30). Oxford University Press.
- Karimi-Rouzbahani, H., Ramezani, F., Woolgar, A., Rich, A., & Ghodrati, M. (2021). Perceptual difficulty modulates the direction of information flow in familiar face recognition. *Neuroimage*, 233, 117896. <https://doi.org/10.1016/j.neuroimage.2021.117896>
- Kemp, C., & Tenenbaum, J. B. (2008). Structured models of semantic cognition. *Behavioral and Brain Sciences*, 31(6), 717–718. <https://doi.org/10.1017/S0140525X08005931>
- Khaligh-Razavi, S. M., Cichy, R. M., Pantazis, D., & Oliva, A. (2018). Tracking the spatiotemporal neural dynamics of real-world object size and Animacy in the human brain. *Journal of Cognitive Neuroscience*, 30(11), 1559–1576. https://doi.org/10.1162/jocn_a_01290
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 804–823. <https://doi.org/10.1037/0278-7393.20.4.804>
- Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *Neuroimage*, 56(2), 411–421. <https://doi.org/10.1016/j.neuroimage.2011.01.061>
- Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage*, 38(4), 649–662. <https://doi.org/10.1016/j.neuroimage.2007.02.022>
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). How does nature program neuron types? *Frontiers in Neuroscience*, 2(1), 4. <https://doi.org/10.3389/neuro.01.016.2008>
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., & Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6), 1126–1141. <https://doi.org/10.1016/j.neuron.2008.10.043>
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>
- 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. <https://doi.org/10.1080/23273798.2015.1130233>
- Kuperberg, G. R., Alexander, E., & Brothers, T. (under review). The posterior P600 does not reflect error correction: An information seeking account of linguistic error processing.
- Kuperberg, G. R., Brothers, T., & Wlotko, E. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 32(1), 12–35. https://doi.org/10.1162/jocn_a_01465
- Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of Animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, 100(3), 223–237. <https://doi.org/10.1016/j.bandl.2005.12.006>
- 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. [https://doi.org/10.1016/S0926-6410\(03\)00086-7](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). *Annual Review of Psychology*, 62(1), 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205. <https://doi.org/10.1126/science.7350657>
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161–163. <https://doi.org/10.1038/307161a0>
- Lambon-Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18(1), 42–55. <https://doi.org/10.1038/nrn.2016.150>
- LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *The Journal of Neuroscience*, 33(13), 5466–5474. <https://doi.org/10.1523/JNEUROSCI.4293-12.2013>
- Laszlo, S., & Plaut, D. C. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain and Language*, 120(3), 271–281. <https://doi.org/10.1016/j.bandl.2011.09.001>
- Lau, E. F., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2013). Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *The Journal of Neuroscience*, 33(43), 17174–17181. <https://doi.org/10.1523/JNEUROSCI.1018-13.2013>
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933. <https://doi.org/10.1038/nrn2532>
- Liu, A. G., Aglinskis, A., & Fairhall, S. L. (2020). General and feature-based semantic representations in the semantic

- network. *Scientific Reports*, 10(1), 8931. <https://doi.org/10.1038/s41598-020-65906-0>
- Luck, S. J. (2014a). Chapter 8: Baseline correction, averaging, and time-frequency analysis. In S.J. Luck (Ed.), *An introduction to the event-related potential technique* (2 ed., pp. 249–282). MIT Press.
- Luck, S. J. (2014b). Chapter 6: Artifact rejection and correction. In S.J. Luck (Ed.), *An introduction to the event-related potential technique* (2nd ed, pp. 185–218). MIT Press.
- Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., & Friederici, A. D. (2006). Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. *Brain Research*, 1096(1), 163–172. <https://doi.org/10.1016/j.brainres.2006.04.037>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, 11(2), 194–201. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11301239. [https://doi.org/10.1016/S0959-4388\(00\)00196-3](https://doi.org/10.1016/S0959-4388(00)00196-3)
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126(2), 99–130. <https://doi.org/10.1037/0096-3445.126.2.99>
- Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. 1st International Conference on Learning Representations (ICLR), workshop track proceedings, Scottsdale, Arizona.
- Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., & Bunn, E. M. (1998). 'Two eyes of a see-through': Impaired and intact semantic knowledge in a case of selective deficit for living things. *Neurocase*, 4(4-5), 291–310. <https://doi.org/10.1080/13554799808410629>
- Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, 66(3), 241–251. <https://doi.org/10.1007/BF00198477>
- Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *Elife*, 4, e11946. <https://doi.org/10.7554/eLife.11946>
- Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., & Matthew Husband, E. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1791), 20180522. <https://doi.org/10.1098/rstb.2018.0522>
- Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, 10(4), e1003553. <https://doi.org/10.1371/journal.pcbi.1003553>
- Nour Eddine, S. (2021). *Divide and Concur: A predictive coding account of the N400 ERP component* [Doctoral dissertation]. Tufts University.
- Nour Eddine, S., Brothers, T., & Kuperberg, G. R. (2022). The N400 in silico: A review of computational models. In K. Federmeier (Ed.), *Psychology of learning and motivation* (Vol. 76, pp. 123–206). Academic Press.
- Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2023). A predictive coding model of the N400. *bioRxiv*, 04.
- Nunez, P. L. (1990). Localization of brain activity with electroencephalography. In S. Sato (Ed.), *Magnetoencephalography (advances in neurology, Vol. 54)* (Vol. 54, pp. 39–65). Raven Press.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 201, 1. <https://doi.org/10.1155/2011/156869>
- Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *The Journal of Neuroscience*, 26(50), 13025–13036. <https://doi.org/10.1523/JNEUROSCI.2481-06.2006>
- Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological Psychology*, 46(2), 143–168. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9288411. [https://doi.org/10.1016/S0301-0511\(97\)05250-2](https://doi.org/10.1016/S0301-0511(97)05250-2)
- Osterhout, L., Kim, A., & Kuperberg, G. R. (2012). The neurobiology of sentence comprehension. In M. Spivey, M. Joannisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 365–389). Cambridge University Press.
- Paczynski, M., & Kuperberg, G. R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb-argument processing. *Language and Cognitive Processes*, 26(9), 1402–1456. <https://doi.org/10.1080/01690965.2011.580143>
- Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, 67(4), 426–448. <https://doi.org/10.1016/j.jml.2012.07.003>
- Payne, B. R., Lee, C. L., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology*, 52(11), 1456–1469. <https://doi.org/10.1111/psyp.12515>
- Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72(2), 184–187. [https://doi.org/10.1016/0013-4694\(89\)90180-6](https://doi.org/10.1016/0013-4694(89)90180-6)
- Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, 28(5), 680–692. https://doi.org/10.1162/jocn_a_00924
- Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, 2(9), 693–705. <https://doi.org/10.1038/s41562-018-0406-4>
- Rabovsky, M., & McRae, K. (2014). Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning. *Cognition*, 132(1), 68–89. <https://doi.org/10.1016/j.cognition.2014.03.010>

- Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). The time course of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, 6, 11. <https://doi.org/10.3389/fnhum.2012.00011>
- Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: Behavioral and computational studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 393–406. <https://doi.org/10.1037/0278-7393.30.2.393>
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. <https://doi.org/10.1038/4580>
- Rogers, T. T., & McClelland, J. L. (2008). Précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, 31(6), 689–714. <https://doi.org/10.1017/S0140525X0800589X>
- Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, 18(4), 367–379. <https://doi.org/10.3758/BF03197126>
- Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *Neuroimage*, 200, 425–436. <https://doi.org/10.1016/j.neuroimage.2019.06.048>
- Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, 137, 29–39. <https://doi.org/10.1016/j.bandl.2014.07.010>
- Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*, 27(4), 665–678. https://doi.org/10.1162/jocn_a_00733
- Siedenberg, R., Goodin, D. S., Aminoff, M. J., Rowley, H. A., & Roberts, T. P. L. (1996). Comparison of late components in simultaneously recorded event-related electrical potentials and event-related magnetic fields. *Electroencephalography and Clinical Neurophysiology*, 99(2), 191–197. [https://doi.org/10.1016/0013-4694\(96\)95215-3](https://doi.org/10.1016/0013-4694(96)95215-3)
- Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Australian Institute of Aboriginal Studies.
- Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, 52(2), 157–168. <https://doi.org/10.1111/psyp.12317>
- Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Reviews™ in Neurobiology*, 14(3-4), 26. <https://doi.org/10.1615/CritRevNeurobiol.v14.i3-4.20>
- Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing*, 17(3), 279–305. <https://doi.org/10.1007/s10339-016-0765-6>
- Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *The Journal of Neuroscience*, 30(46), 15643–15653. <https://doi.org/10.1523/JNEUROSCI.1899-10.2010>
- Stokes, M. G., Wolff, M. J., & Spaak, E. (2015). Decoding rich spatial information with high temporal resolution. *Trends in Cognitive Sciences*, 19(11), 636–638. <https://doi.org/10.1016/j.tics.2015.08.016>
- Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 397–439). Oxford University Press.
- Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297–314. <https://doi.org/10.1016/j.jml.2012.12.002>
- Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122(11), 2119–2132. <https://doi.org/10.1093/brain/122.11.2119>
- Taylor, K. I., Devereux, B. J., & Tyler, L. K. (2011). Conceptual structure: Towards an integrated neuro-cognitive account. *Language and Cognitive Processes*, 26(9), 1368–1401. <https://doi.org/10.1080/01690965.2011.568227>
- Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, 42(1), 1636–1643. <https://doi.org/10.1111/ejn.12936>
- Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, 5(6), 244–252. <http://www.ncbi.nlm.nih.gov/pubmed/11390295>. [https://doi.org/10.1016/S1364-6613\(00\)01651-X](https://doi.org/10.1016/S1364-6613(00)01651-X)
- Uusitalo, M. A., & Ilmoniemi, R. J. (1997). Signal-space projection method for separating MEG or EEG into components. *Medical & Biological Engineering & Computing*, 35(2), 135–140. <https://doi.org/10.1007/BF02534144>
- Van Berkum, J. J. A. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Palgrave Macmillan.
- van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 3(5), 1211–1224. <https://doi.org/10.1111/j.1749-818X.2009.00163.x>
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18(4), 380–393. <https://doi.org/10.3758/BF03197127>
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190. <https://doi.org/10.1016/j.ijpsycho.2011.09.015>
- Van Petten, C., Weckerly, J., Mclsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, 8(3), 238–242. <https://doi.org/10.1111/j.1467-9280.1997.tb00418.x>
- Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, 137, 188–200. <https://doi.org/10.1016/j.neuroimage.2015.12.012>
- Wang, L., Hagoort, P., & Jensen, O. (2018). Language prediction is reflected by coupling between frontal gamma and posterior alpha oscillations. *Journal of Cognitive Neuroscience*, 30(3), 432–447. https://doi.org/10.1162/jocn_a_01190
- Wang, L., Kuperberg, G., & Jensen, O. (2018). Specific lexico-semantic predictions are associated with unique spatial

- and temporal patterns of neural activity. *Elife*, 7, e39061. <https://doi.org/10.7554/eLife.39061>
- Wang, L., Schoot, L., Brothers, T., Alexander, E., Warnke, L., Kim, M., Khan, S., Hämäläinen, M., & Kuperberg, G. R. (2023). Predictive coding across the left fronto-temporal hierarchy during language comprehension. *Cerebral Cortex*, 4478–4497. <https://doi.org/10.1093/cercor/bhac356>
- Wang, L., Wlotko, E., Alexander, E. J., Schoot, L., Kim, M., Warnke, L., & Kuperberg, G. R. (2020). Neural evidence for the prediction of Animacy features during language comprehension: Evidence from MEG and EEG Representational Similarity Analysis. *The Journal of Neuroscience*, 40(16), 3278–3291. <https://doi.org/10.1523/JNEUROSCI.1733-19.2020>
- Warrington, E. K., & McCarthy, R. (1987). Categories of knowledge. *Brain*, 110(5), 1273–1296. <https://doi.org/10.1093/brain/110.5.1273>
- Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107(3), 829–853. <https://doi.org/10.1093/brain/107.3.829>
- Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2021). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour*, 5(3), 389–398. <https://doi.org/10.1038/s41562-020-00982-w>
- Wydell, T. N., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural correlates of letter-string length and lexicality during reading in a regular orthography. *Journal of Cognitive Neuroscience*, 15(7), 1052–1062. <https://doi.org/10.1162/089892903770007434>
- Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30(6), 648–672. <https://doi.org/10.1080/23273798.2014.995679>
- Zannino, G. D., Perri, R., Pasqualetti, P., Caltagirone, C., & Carlesimo, G. A. (2006). Analysis of the semantic representations of living and nonliving concepts: A normative study. *Cognitive Neuropsychology*, 23(4), 515–540. <https://doi.org/10.1080/02643290542000067>