No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing

Dejan Draschkow[a],[*],[1],[2], Edvard Heikel[a], Melissa L.-H. Võ[a], Christian J. Fiebach[a],[b], Jona Sassenhagen[a]

[a] *Department of Psychology, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 6, Frankfurt am Main 60323, Germany*
[b] *Brain Imaging Center, Goethe University Frankfurt, Frankfurt am Main, Germany*

A B S T R A C T

Attributing meaning to diverse visual input is a core feature of human cognition. Violating environmental expectations (e.g., a toothbrush in the fridge) induces a late event-related negativity of the event-related potential/ERP. This *N400* ERP has not only been linked to the semantic processing of language, but also to objects and scenes. Inconsistent object-scene relationships are additionally associated with an earlier negative deflection of the EEG signal between 250 and 350 ms. This *N300* is hypothesized to reflect pre-semantic perceptual processes. To investigate whether these two components are truly separable or if the early object-scene integration activity (250–350 ms) shares certain levels of processing with the late neural correlates of meaning processing (350–500 ms), we used time-resolved multivariate pattern analysis (MVPA) where a classifier trained at one time point in a trial (e.g., during the N300 time window) is tested at every other time point (i.e., including the N400 time window). Forty participants were presented with semantic inconsistencies, in which an object was inconsistent with a scene's meaning. Replicating previous findings, our manipulation produced significant N300 and N400 deflections. MVPA revealed above chance decoding performance for classifiers trained during time points of the N300 component and tested during later time points of the N400, and vice versa. This provides no evidence for the activation of two separable neurocognitive processes following the violation of context-dependent predictions in visual scene perception. Our data supports the early appearance of high-level, context-sensitive processes in visual cognition.

1. Introduction

Learned regularities and previous experience with our visual environment regulate predictions about *which* objects should occur *where* in a scene, alleviating the computational load of perceptual processes (Bar, 2004, 2007, 2009; Biederman, 1981; Biederman et al., 1982). For example, objects that are not easily identifiable when presented without scene context can be easily identified if the scene background is provided (Brandman and Peelen, 2017). These predictions can be investigated by showing observers images containing violations of different forms. Thus, seeing a bathtub in a living room would violate what we have interpreted in previous work as *semantic* predictions about what object belongs in the scene, while finding a toilet brush next to the toothpaste would violate *spatial* predictions (Draschkow and Võ,

2017; Võ and Wolfe, 2013, 2015). These paradigms have revealed that violations of predictions can lead to slower and less accurate identification of objects (Bar, 2004; Biederman et al., 1982; Davenport and Potter, 2004), elicit longer and more frequent fixations on critical objects (Cornelissen and Võ, 2016; Henderson et al., 1999; Loftus and Mackworth, 1978), and impede visual search (Castelhano and Heaven, 2011; Võ and Henderson, 2010). Studies of brain correlates of meaning in language identified a late event-related negativity (N400) sensitive to violations of semantic expectations (Kutas and Federmeier, 2011; Kutas and Hillyard, 1980). Object-scene violations are accompanied by a similar, yet more anteriorly distributed negativity: a scene N400 (Ganis and Kutas, 2003; Kovalenko et al., 2012; McPherson and Holcomb, 1999; Mudrik et al., 2010; Võ and Wolfe, 2013; Lauer et al., 2018). It is hypothesized to accompany semantic processing of scenes.

Inconsistent object-scene relationships are additionally associated with an earlier negative component (250–350 ms) – often referred to as N300. McPherson and Holcombe (1999) first demonstrated that objects preceded by an unrelated prime elicit a more frontally distributed event-related negativity around 300 ms, supporting the existence of two separate components, an anterior, image-specific N300 and a later, central/parietal concept-level N400. Ever since this initial finding, many studies using visual objects and scenes have used this terminology and/or have separated their analysis according to these proposed components (e.g., Federmeier and Kutas, 2001; Hamm et al., 2002; Lauer et al., 2018 Meade et al., 2018; Mudrik et al., 2010; Mudrik et al., 2014; Sitnikova et al., 2008; Võ and Wolfe, 2013; Willems et al., 2008). The N300 is hypothesized to reflect pre-semantic perceptual processes (Mudrik et al., 2010; Schendan and Kutas, 2002; Schendan and Maher, 2009) and in a direct comparison, Hamm et al. (2002) argued that the N300 and N400 are generated by distinct underlying networks of cortical activity and reflect two distinct semantic effects in object identification – categorization and amodal semantic mismatches respectively.

There is, however, also a body of evidence suggesting that the separation of these components might be artificially imposed due to data preprocessing, task (e.g. quicker access to pictures compared to words due to a less arbitrary relationship between object/scene pairs) and/or participant specific variance in the temporal manifestation of a single component. Thus, it cannot be fully excluded on the basis of the available empirical literature that the topographic differences might only be superficial in nature. This notion receives support from several lines of evidence: Willems et al. (2008) failed to observe a separate N300 effect and accordingly argued against this component being specific to the processing of pictures. Previous studies have also either failed to find a distinguishable N300 effect (Demiral et al., 2012; Federmeier and Kutas, 2001; Ganis and Kutas, 2003; Nigam et al., 1992), or found a similar time-course between the N400 of words and pictures, but diverging topographies (Ganis et al., 1996). Further, it is possible that even with superficial differences in topography, the exact same neural structures are engaged during these time windows and an overlap of a late posterior positivity from the P3 family results in the apparent reduction of the N400 effect at parietal and posterior scalp sites (Nobre and McCarthy, 1994). Finally, it can be questioned whether the early scene-specific N300 effect is at all exclusive for pictorial stimuli, as its time window (i.e., starting around 250 ms and ending at 350 ms) overlaps with the well-established time course of the N400 often reported in language studies (250–500 ms; e.g. Kutas and Federmeier, 2011).

The potential separability of N300 and N400 may ultimately not be resolvable using classical ERP analyses, as their ability to identify differences in the underlying neural substrate of two ERP components with similar scalp distributions is inherently limited. As shown by Urbach and Kutas (2002), common statistical procedures intended to identify a dissimilarity of the underlying source configuration do not in fact do so. A promising novel tool for identifying dissimilarities vs. commonalities of EEG signals across time is time-generalized multivariate decoding (e.g., King and Dehaene, 2014), which has also been applied to investigating the N400 component in language (Heikel et al., 2018). Time-generalized decoding consists of training machine learning classifiers to distinguish between experimental conditions at each point throughout the trial, based on their specific patterns of EEG activity. The resulting fitted classifiers are then each evaluated at all time-points. Applying this method to the investigation of N300 vs. N400 responses to object-scene inconsistency, one can train classifiers to learn neural patterns separating congruent from incongruent conditions during the N300 time window, and then test how well these classifiers perform when applied to classify N400 time window activity (and vice versa). While still operating in the space of scalp-recorded topographical patterns (and not, e.g., in source space representing the generators of the underlying neural processes), this procedure allows one to positively quantify the degree of overlap and similarity between neural patterns at

different time points. This bottom-up, data-driven approach, accordingly, transcends a theory-motivated, or descriptive, segmentation of the event-related potential into arbitrary windows, providing a more objective look at sequences of processing stages.

In our specific case, we leverage time-generalized decoding to test whether or not the congruence effect during the N300 window differs from that during the N400 window. While not being conclusive proof for identity vs. non-identity, given the inherent limitations of scalp EEG and the nature of falsificationist hypothesis testing, the possible results clearly map onto different hypotheses. If N300 and N400 effects reflect just one continuous process, classifiers trained during the N400 time window should perform well during the N300 time window, and vice versa. If, however, the two ERP effects reflect different cognitive stages in a processing chain with different cortical substrates, then no generalization should be found between N300 and N400. Note that the results of this analysis can speak only to a *neurocognitive* theory of differences in the underlying neurocognitive events. It is, in principle, possible that N300 and N400 window reflect one and the same cortical source configuration (in which case substantial temporal generalization should be observed), but very different cognitive computations (performed by one and the same brain area). However, finding such N300/N400 cross-decoding would argue against strong interpretations of N300 and N400 time windows as reflecting two different cognitive processing stages.
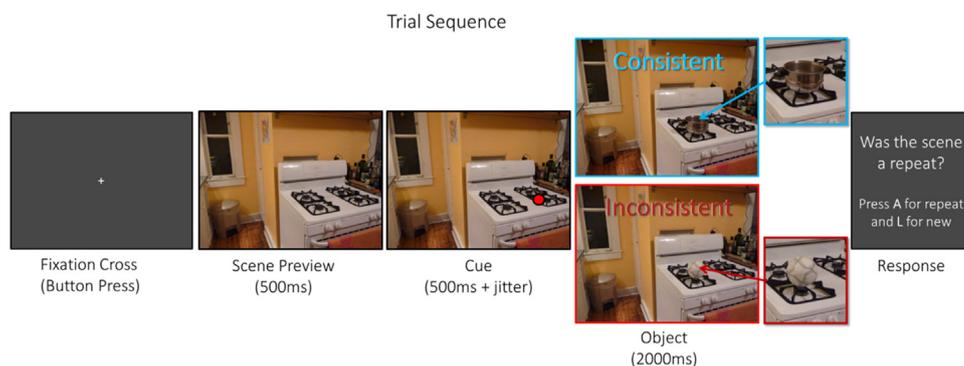
2. Methods

2.1. Participants

Participants were recruited at the Goethe University Frankfurt, until forty complete data sets were obtained (mean age = 21.8, range = 18–41, 33 female, 4 left-handed). As this is the first analysis of its kind, no sensible power analysis could be conducted; we instead simply chose a sample size that is large compared to similar studies (Mudrik et al., 2010; Võ and Wolfe, 2013), while still being feasible. All participants had normal or corrected-to-normal vision. All were volunteers receiving course credit or financial compensation and had given informed consent according to protocols approved by the local ethics committee. None reported a history of neurological or psychiatric disorders.

2.2. Stimuli and procedure

The stimulus material (318 saliency controlled images of real-world scenes) and procedure were nearly identical to the study of Võ and Wolfe (2013). The 318 images consisted of 152 unique scenes in either a semanitcally consistent or inconsistent version (i.e., the object was either from the same category as the scene or not), as well as 10 additional scenes used as targets for a repetition detection task and 4 practice images. Stimuli were presented in a dimly-lit room using OpenSesame (Mathôt et al., 2012) on a 24-in. monitor (resolution = 1920 × 1080, viewing distance approx. 65 cm, scenes subtending approx. 24° (horizontal) by 18° (vertical) of visual angle).

Participants were told that they would see a series of scenes, each containing one critical object marked by a cue. Each trial began with a blink phase. The participant could initiate the trial by pressing the spacebar, which was followed by the presentation of a scene without the critical object for 500 ms. Next, a red dot appeared at a location in the scene, indicating where to move the eyes and where to expect the critical object to appear. Five hundred milliseconds after onset of the cue (plus a random jitter between 0 and 300 ms, to prevent anticipatory effects), the critical object appeared in the scene and remained visible together with the scene for 2000 ms (Fig. 1). To keep participants engaged in viewing the scenes without explicitly probing the object-scene inconsistencies, we asked them to view each scene carefully and to press a button whenever they spotted an exact repetition (i.e., the same scene with the same object in the same location as seen on a previous

Trial Sequence

Fixation Cross
(Button Press)

Scene Preview
(500ms)

Cue
(500ms + jitter)

Consistent

Inconsistent

Object
(2000ms)

Was the scene
a repeat?

Press A for repeat
and L for new

Response

Fig. 1. Trial sequence. Each trial started with the presentation of a fixation cross that indicated blinking was encouraged. Once ready, subjects pressed a button, which triggered the presentation of a preview scene without the critical object (500 ms). Next, a cue appeared (500 ms plus randomly sampled jitter between 0 and 300 ms), and participants moved their eyes to the cued location. Then the object appeared at the cued location and remained visible on the screen together with the scene (2000 ms). The object could either be consistent or inconsistent with the scene. Finally, the participants indicated if they had seen the current object-scene combination before during the experiment.

trial).

The 169 experimental trials included 152 unique and 17 repetition trials. Each of the 152 unique scenes was used in either a semantically consistent or inconsistent version (Fig. 1), resulting in 76 trials per condition. Each participant saw each of the scenes only once during the event-related potential (ERP) experiment, except for the additional 17 trials containing repeated (i.e. target) scenes for a repetition detection task. All target scenes for the repetition detection task were excluded from subsequent analysis (i.e., their first and subsequent presentations), thus the analysis was conducted on data from the 152 experimental trials. Assignment of scenes to the two conditions was counterbalanced across participants and the order of scene presentation was random. Participants were acquainted with the procedure through 4 practice trials before the start of the experiment. The experiment lasted ~30 min.

2.3. Data acquisition and pre-processing

The complete pre-processing and analysis scripts can be found alongside the experimental data as html files and as reproducible scripts (jupyter notebooks; (Kluyver et al., 2016)) at https://github.com/DejanDraschkow/n3n4).

The electroencephalogram (EEG) was recorded with a sampling rate of 1000 Hz from 64 active Ag/AgCl electrodes (arranged in an extended 10–20 layout using either a brainAmp amplifier or an actiChamp amplifier (both: Brain Products GmbH, Gilching, Germany). EEG data analysis was conducted in MNE-Python (Gramfort et al., 2013; https://mne-tools.github.io/). First, data was referenced to linked mastoid electrodes. Then, it was down-sampled to 200 Hz, high-pass filtered at 0.1 Hz and low-pass filtered at 40 Hz. Eye movements and muscle artefacts were corrected via independent component analysis (ICA; Jung et al., 2000). ICA components were estimated on data which was high-pass filtered at 8 Hz using FastICA. Eye movement components were detected by (1) correlating the filtered data with the electro-oculography (EOG) signal plus (2) manually selecting a subset of typical component maps and identifying the best group match to them (Viola et al., 2009). Selected components were then removed from the 0.1 Hz filtered data and a 20 Hz low-pass filter was applied.[3] Then, EOG channels were dropped, leaving 60 channels in total. Subsequently, data was segmented into 1100-ms epochs time-locked to the onset of the cued object (i.e., − 200 to + 900 ms relative to target stimulus onset). Each epoch was baseline-adjusted by subtracting the mean amplitude in the prestimulus period (− 200 to 0 ms) from all the data points in the epoch. Finally, fully automated artifact rejection with

default values using peak-to-peak thresholding was used to interpolate artefactual channels and to drop contaminated epochs (Jas et al., 2017), leaving on average 149 trials (132–152) per subject, with a mean of 74.6 trials in the consistent and 74.4 in the inconsistent condition.

2.4. Data analysis

2.4.1. Univariate analysis

Event-related brain potentials (ERPs) were calculated by first averaging trials within subjects, and then averaging these waveforms across subjects, separately for consistent and inconsistent trials. For statistical analysis, the mean amplitudes were calculated for two consecutive time windows, i.e., 250–350 ms (N300) and 350–500 (N400), across the mid-central region (electrodes FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, and CP2) which was previously shown to display strong scene-related N300 and N400 effects (Ganis and Kutas, 2003; Mudrik et al., 2014; Võ and Wolfe, 2013). Paired sample *t*-tests were used for the critical comparisons between conditions.

2.4.2. Multivariate pattern analysis

To test to which degree similar neural patterns occur at different time points of object-scene integration, a multivariate pattern analysis (MVPA) was implemented on the epoched EEG data. Independently for each subject, a Logistic Regression implemented in scikit-learn (Pedregosa et al., 2011), with default parameters, was trained to classify trials as being consistent or inconsistent based on brain activity. Classifiers were trained separately on EEG activity at each time point. A 5-fold stratified cross-validation procedure with balanced classes (i.e., equal number of consistent vs. inconsistent trials per fold, within each subject) was used: Each participant's epochs were split into five equal-sized folds. For each time point in each epoch in a fold, trial type (inconsistent vs. consistent) was predicted by a classifier that had been fitted (i.e., 'trained') on the other four folds. To assess the quality of the predictions - i.e., correctly vs. incorrectly labelled trials – the Area under the Curve of the Receiver-Operating Characteristic was calculated, as a sensitive, yet robust scoring procedure: higher scores (on a scale from 0 to 1) indicate that brain activity more strongly differs between the two conditions (with 0.5 corresponding to guessing, and 1 to perfect accuracy).

To investigate if neurocognitive patterns are shared between early and late stages of the N300/N400 complex, MVPA was applied in a time-generalized manner (King and Dehaene, 2014). In this procedure, a classifier is not only tested at the time point it was trained on (e.g., during the N300 time window), but also used for predicting the condition of the trial at every other time point (i.e., including the N400 time window). This is schematically illustrated in Fig. 2. Calculating classification scores based on EEG activity at each time for classifiers trained at each time point results in a Generalization Across Time/GAT matrix (Fig. 5A) that shows training times on the y-axis against testing times on the x-axis. The diagonal of this GAT matrix represents training

---

[3] We also repeated the MVPA analyses on data filtered even more modestly – below 55 Hz – to exclude artificially induced temporal generalization. Due to the much lower signal to noise ratio, overall decoding accuracies were lower; nevertheless, the qualitative pattern of results did not change.
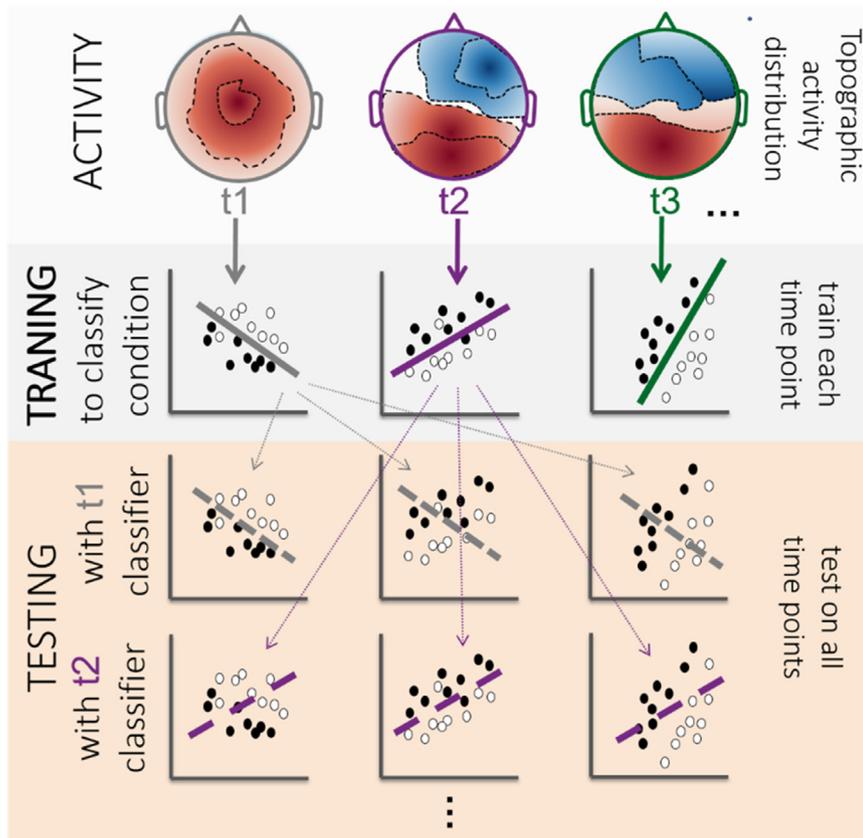
Fig. 2. Visualization of the generalization across time (GAT) procedure. First row: at time points in the trial where neurocognitive processes differ between two experimental conditions, distinct spatio-temporal responses are evoked which appear on individual trials (mixed with noise). MVPA methods use powerful pattern classifications algorithms to learn multivariate patterns that distinguish between the condition-specific EEG responses (second row). To test whether neurophysiological patterns generalize to other time points, a classifier trained at one time point (t) is also scored concerning its predictions based on neurophysiological patterns at all other time points, i.e. testing at time point t2, t3, etc. This procedure is repeated for all time points of a trial.

and testing at the same time point (e.g., trained at 350 ms and tested on 350 ms) – i.e., the strength of the neural pattern dissociating violation from control trials. Off-diagonal entries show pattern persistence or reoccurrence – that is, time points t + x where a classifier trained at time point t can still successfully classify trials, indicating that similar EEG patterns, and thus, by inference, similar cognitive processes, characterize both time points. If the N300 was functionally distinct from the N400, one would expect that classifiers trained during the N300 time windows would not generalize well to later time points of the N400 component. Demonstrating above chance classification accuracy of these classifiers (i.e., temporal generalization from N400 to N300, and the reverse) would, however, indicate that similar neural patterns are generating the two components.

2.4.3. Statistical analysis

Time-generalized decoding scores were statistically evaluated in the two time windows introduced above, i.e., N300 and N400, thereby tracking over the time course of the entire epoch the classification accuracy scores of classifiers trained on data from these two time windows. Specifically, the performance of N300 vs. N400 classifiers over time was evaluated by (1) separately averaging the accuracy scores of all classifiers that were trained on the data points from the two non-overlapping time windows (i.e., 250–350 ms and 350–500 ms, respectively), and then (2) inferential testing their classification accuracy against chance (0.5) or against each other (N300 vs. N400). Diagonal decoding scores were analyzed in the same way. GAT matrices were subjected to the same procedure.

For time series and GAT matrices, statistical results were subjected to Threshold-Free Cluster Enhancement/TFCE (Smith and Nichols, 2009), implemented in MNE Python. TFCE is a nonparametric method for identifying statistically significant contrasts that derives statistical power from the cluster structures in the data (Maris and Oostenveld,

2007). It does not require selecting parameter values. TFCE controls for multiple comparisons, while retaining high sensitivity.

To investigate if classifiers trained in one time window outperformed those trained in the other time, mean scores of all decoders trained on one time window were averaged first within the training window, and then within the other window (i.e., N300 classifiers - > N300 time window, N300 classifiers - > N400 time window, etc.). If the N300 classifiers outperformed N400 classifiers during the N300 window, this would indicate that there were distinct patterns occurring only during the N300 time window, and which could therefore only be learned by N300-trained classifiers. To quantify if there was any effect specific to the N300 window that could not be explained by N400 window decoders, the 95% bootstrapped confidence interval for the difference between N300- > N300 and N400- > N300 cross-decoding was calculated, as well as for N400- > N400 vs. N300- > N400.

Finally, we conducted two analyses on aggregate activity in the N300 and N400 time windows.[4] First, we repeated the above analysis, but averaged across time points in a preceding step. Next, it might be argued that the method we employ here could only ever produce evidence in favor of two distinct patterns in the form of a negative finding: an inability to cross-decode. We implemented an analysis capable of providing positive evidence in the following form: separately for each dataset and each trial, we averaged activity in the N300 and the N400 time window. Then, we trained a classifier (Logistic Regression and 5-fold cross-validation; i.e., as before) to predict, based on these values, if a pattern was extracted from the N300 or the N400 time window. To ensure that temporal autocorrelation did not bias the classifier, we split the trials in half, i.e., putting N300 time window activity for all even trials and N400 activity for all odd trials in one run, and the remaining

---

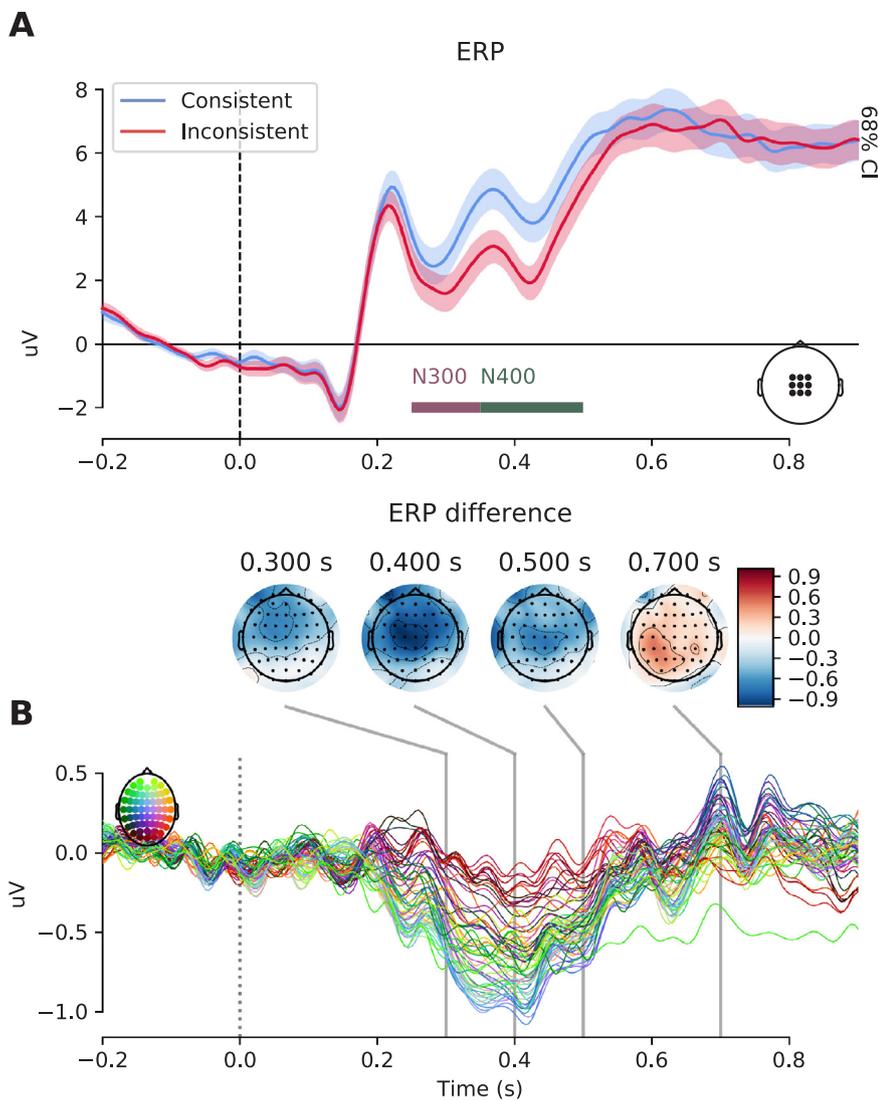[4] We thank an anonymous reviewer for suggesting these two analyses.

A

B

Fig. 3. Event-related potentials, consistent vs. inconsistent. A: ERP time-locked to scene plus object onset, for consistent and inconsistent conditions, at central electrodes. Shaded region indicates a 68% confidence interval. Purple and dark green horizontal bars indicate N300 and N400 time windows. B: Joint Butterfly + topographical map plot of inconsistent minus consistent difference. Each colored line represents one channel; see colored inset legend (left) for locations on the EEG cap. Adjacent channels receive adjacent colors. Additionally, topographical maps are shown for representative time points. Topomaps indicate the similarity of patterns at 300, 400 and 500 ms; colored line plots show that these patterns are representative for the entire time course of the negative-polarity ERP effect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

trials in another, and averaged the results. A positive outcome would indicate that the classifier could pick up on neural patterns indicating if the trial was from the early or the late time window. We calculated via bootstrapping the 95% confidence interval across datasets for the resulting decoding scores.

3. Results

3.1. Behavioral analysis

The overall error rate on the repetition detection task averaged 5.5% ($SD = 3.7\%$). The false alarm rate, i.e., participants erroneously reporting a unique scene as a repeat, was 2.6% ($SD = 4.1\%$).

3.2. ERP results

Replicating previous findings, in the N300 time window (250–350 ms), semantic violations elicited a significantly more negative response than the consistent control condition, $t(39) = 4.52$, $p < .001$ (see Fig. 3A). The same was true for the N400 time window (350–500 ms), $t(39) = 5.14$, $p < .001$. Visualizing the spatiotemporal structure of contrast effects for N300 and N400 time windows (see Fig. 3B) indicated highly similar patterns.

3.3. MVPA results

In agreement with raw ERP results, examining the distribution of patterns across time reinforces the interpretation that essentially the same neural pattern is expressed throughout the 200–550 ms range. Note that all multivariate analyses are not based on channel pre-selection, instead taking into account the full topographic distribution.

Temporally resolved decoding indicated above-chance decoding beginning around 200 ms after stimulus onset (Fig. 4A), i.e., it was possible to classify trials as consistent vs. inconsistent based on brain activity. The corresponding classifier patterns (i.e., the neural patterns dissociating brain activity in inconsistent vs. consistent trials) are shown in Fig. 4B. Much like scalp topographies of raw ERP differences, classifier patterns throughout the N300 and N400 windows (extracted by the method presented by Haufe et al. (2014)) were visually highly similar both throughout the time windows, and compared to the ERP results.

Classifiers trained during any of these windows showed statistically significant above-chance decoding throughout the entire time window, as evidenced by the Generalization Across Time matrix displayed in Fig. 5A; e.g., a classifier trained on EEG data at 400 ms after stimulus presentation (y-axis, 400 ms) could successfully classify trials based on EEG activity at 300, 400 and 500 ms (x-axis, 300, 400, 500 m).

To quantify the separability of neural processes reflected in the N300 vs. N400 time windows, we averaged the time-courses of
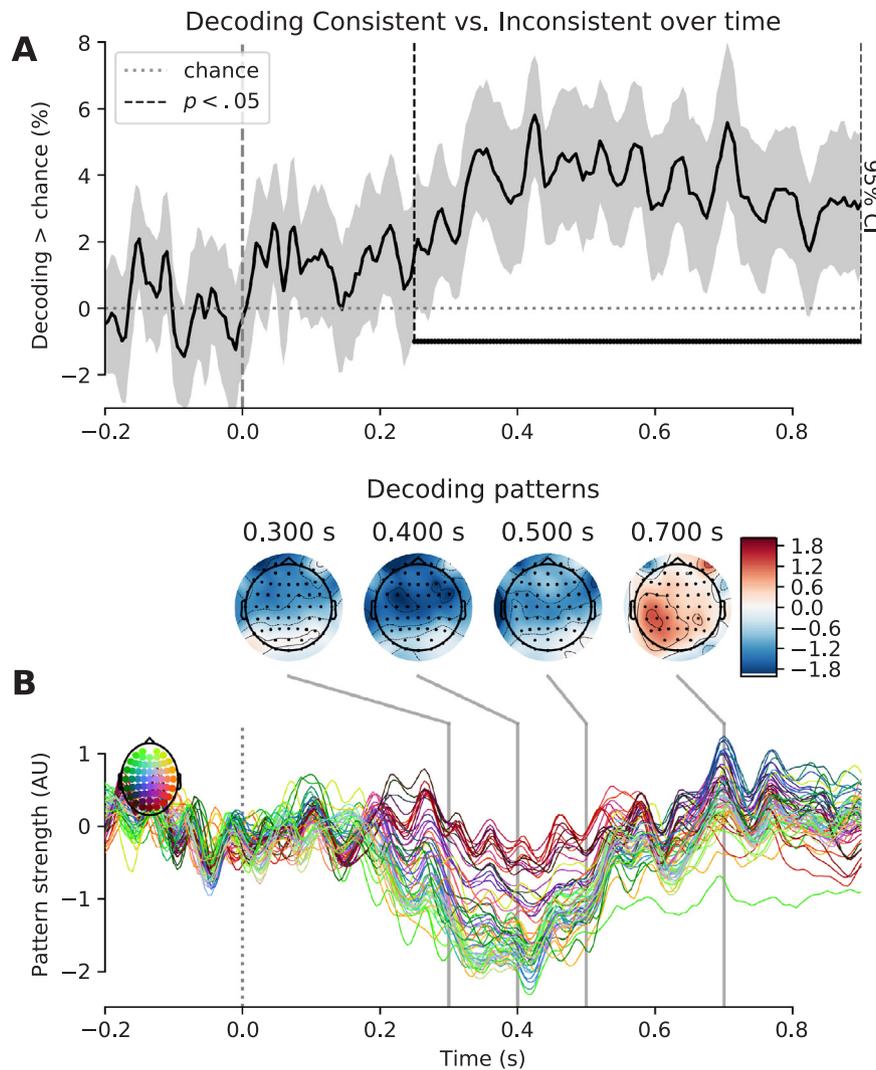
Fig. 4. Decoding across time. A: Temporally resolved decoding of inconsistent vs. consistent trials, for each time point. Thick horizontal black line indicates statistically significant ($p < .05$, TFCE) time points. Shaded region indicates 95% confidence interval. B: corresponding spatial patterns learned by the decoder, separately at each time point, visualized as the ERP contrast in 3B (joint Butterfly + topographical map plot). Patterns were extracted from decoder coefficients via the technique discussed by Haufe et al. (2014).

performance scores of all classifiers trained between 250 and 350 ms ('*N300 classifier*') and of all classifiers trained between 350 and 500 ms ('*N400 classifier*'). Fig. 5B and C show the degree to which N300 and N400 classifiers can decode trial type at different time points in the trial. The temporal evolution of the decoding performance of both classifiers across the entire trial epoch is shown in Fig. 5B. Both classifiers demonstrated significant decoding performance throughout the entire duration of the N300/N400 complex, starting as early as ~200 ms post object onset ($p < .05$).

The early (i.e., N300) classifier reliably predicted activity throughout the entire window of the N300/N400 complex, albeit not up to the full extent of the epoch. Moreover, the late (i.e., N400) classifier generalized beyond its window of training and was able to predict earlier EEG activity (as well as later activity extending up to 800 ms). Fig. 5C (left panel) shows that in the N300 time window, N300 classifiers do not perform better than N400 classifiers. During the N300 time window, decoding performance was higher or equivalent for N400 vs. N300 classifiers, but not to a statistically significant degree (all $p > .05$) and with a confidence interval very narrowly centered around zero: N300 $= > $ N300 vs. N400-$= > $ N300 mean decoding scores were -0.03 ($-.66$ to $.72$), i.e., a substantial advantage of N300 classifiers (over N400 classifiers) when classifying N300 time window

activity can be confidently excluded. That means N400 time window classifiers were not any worse at classifying N300 time window patterns than N300 classifiers were. For comparison, the N400 time window is also shown (Fig. 5C, left panel). The N400 $= > $ N400 result was in fact greater than the N300 $= > $ N400 decoding (mean: 1.4; 95% CI: . 0.672.21).

Finally, we found that classifiers trying to predict if a data slice stemmed from the N300 or the N400 time window were at chance performance (both $p > $ [T 0.2see Fig. 5C, right). Furthermore, we could not observe any 'home advantage' of N300-trained classifiers when classifying trials based on average activity in that window, compared to N400-trained classifiers, although there was a slight advantage of N400 over N300 classifiers when predicting N400 time window activity (Fig. 5C, center).

4. Discussion

Disentangling the role of contextual scene information in object identification is crucial to understanding the efficiency of perceptual processes. There are different ways in which scene and object processing could interact: scene and object information might be processed in parallel and integrated only post-perceptually (Hollingworth and
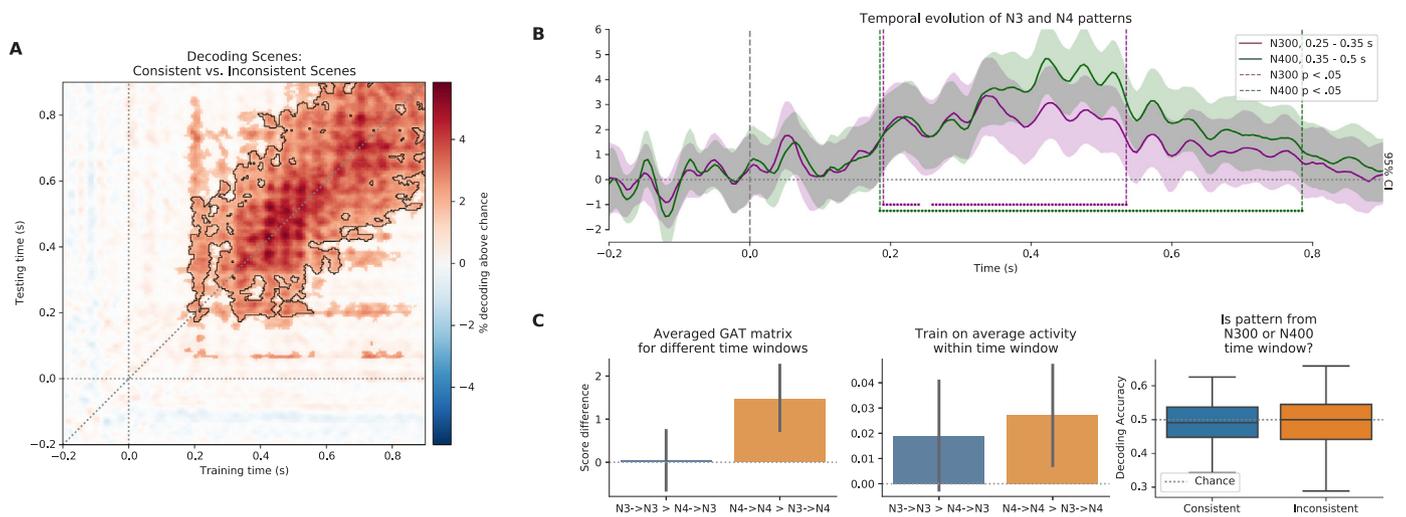
Fig. 5. Time-resolved decoding. A: GAT matrix. Decoding accuracy above chance (warmer colors indicate better performance) is plotted as a function of time points used for training the classifier (Y-axis) and time points used to test the classifier (X-axis). Diagonal corresponds to 4 A. Areas not significant at $p < .05$ (TFCE) are set to transparent; areas significant at $p < .01$ are surrounded by black contours. B: Decoding performance for selected time windows of interest representing the early (250–350 ms; purple) and late (350–500 ms; green) time windows of the N300/N400 complex. Performance time-courses represented here reflect the average of performance time-courses of all classifiers trained in the respective time window, i.e., an average of all rows of the GAT matrix shown in 5 A within that time window. Significant above chance ($> 50\%$) decoding ($p < .05$, TFCE) is depicted as horizontal solid lines with vertical dashed lines representing the beginning and the end of the significant periods. Shaded region indicates 95% confidence interval. C: Left: Average performance of N300 and N400 time window classifiers in N300 and N400 time windows. Center: performance of decoders specifically trained on averaged N300 or N400 time window activity when generalized to N300 and N400 time window. Right: performance of decoders trained to predict if a data point comes from N300 or N400 time windows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Henderson, 1999). Or scene information might facilitate the processing of objects already at a perceptual stage, with contextual information reducing the subset of possible object interpretations by activating candidate object representations (Bar, 2004; Brandman and Peelen, 2017; Trapp and Bar, 2015). An established method for investigating the role of scene context on object identification is to violate predicted object-scene pairings. Violations in expected object-scene relationships are associated with two negative ERP deflections: a later (350–500 ms) component similar to the N400 in language paradigms (Ganis and Kutas, 2003; Kutas and Federmeier, 2011) and an earlier (250–350 ms) component – referred to as N300 and hypothesized to reflect pre-semantic perceptual processes (Hamm et al., 2002; McPherson and Holcomb, 1999). The N300 has been taken as evidence for contextual information already biasing perceptual processes (Mudrik et al., 2010, 2014; Võ and Wolfe, 2013). This early interaction has been corroborated by high-resolution neuroimaging (Brandman and Peelen, 2017). It remained unresolved, however, if the N300 and N400 components reflect two distinct semantic processes in object identification – categorization and amodal semantic integration respectively – or if they share underlying networks of cortical activity.

In this study, we first of all replicate previous findings of congruency differences in both the N300 and N400 time windows (Mudrik et al., 2010, 2014; Võ and Wolfe, 2013). Beyond that, by applying MVPA to our EEG data, we found shared neural patterns across the observed N300 and N400 components - and therefore no evidence for distinct processes between the early and late object-scene integration stages. This supports the notion that similar neural patterns are active during both time windows. It argues against the interpretation of these components as reflecting separate perceptual vs. semantic processes. It also suggests the term N300 should be used with caution, or employed purely descriptively, when referring to the early part of the N400, so as to not suggest neurocognitive evidence for two distinct processing stages during this time window.

More generally, our results speak against an interpretation of the ERP as a fixed sequence of time windows, perhaps motivated by directly reading off peaks in the raw waveforms. Many ERP components vary systematically in their latency depending on various experimental or

internal contingencies (e.g., Sassenhagen et al., 2014; Verleger, 1997). Here, we show that essentially the same neurocognitive pattern can extend across what in other research has been understood as boundaries between such windows. That is, semantically inconsistent scene contexts trigger processes that encompass both early and later phases of object processing. In that vein, a recent study by Truman and Mudrik (2018) manipulated both object identifiability and semantic congruence of objects displayed in scenes to test the influence of context on object identification. This study demonstrates that experimental manipulations can distinguish between two functionally different events occurring in the same N300 time window – e.g., object identification based on visual features and semantic incongruence processing. Importantly, the waveforms for semantically incongruent objects diverged from visually unidentifiable ones later than semantically congruent objects, indicating these were only identified as objects later in time; this was taken as evidence for scene contexts affecting object identification. However, these different processes were also reflected in very distinct topographies (with a fronto-central pattern – likely the same pattern we are associating with the N400 here – for inconsistency, and an occipital pattern for object identifiability), indicating that two very different neurocognitive events play out in this time window.

While we do not find evidence for two separate neurocognitive processes underlying both early and later effects of scene contexts on object processing, our findings should also not be overinterpreted. First, while they indicate that some neural sources are activated throughout the combined N300/N400 window, it is also possible that there are other, i.e., distinct neural sources active in the N300 and/or N400 windows which can, however, not be detected with the methods applied here – e.g., non-phase-locked effects, or those with source configurations unlikely to be detected via EEG. Second, in our study the scene and object were sequentially presented. There is evidence from studies providing simultaneous presentation of object and scene (Mudrik et al., 2010, 2014), which indicate a possibly more pronounced N300 response – it might be that such a paradigm is more sensitive to an independent early process. Moreover, while our results indicate that the same neural substrates are active throughout both time windows, this cannot automatically be taken as evidence that only one cognitive

process unfolds; rather, the same substrate might in principle be involved in two entirely different processes from one time-point to the next.

As a specific example for how our results do not prove the complete identity between N300 and N400 time window activity, consider that N400 classifiers also perform above chance at later time points than N300 classifiers do. That is, later GAT decoding – in the top right of the GAT matrix – indicate that while N300 activity can be decoded by N400 classifiers as well as by N300 classifiers, N400 classifiers detect their patterns throughout a slightly longer window. In this late window, N300 classifiers no longer pick up activity. Moreover, there was a N400 classifier "home advantage" (see Fig. 5b, right), indicating that N300 time window patterns are a subset of the patterns found in the N400 time window. This could indicate that a further process – e.g., the late positive complex (e.g., Schendan and Kutas, 2002) – begins already towards the edge of the N400 time window, and is partially learned by N400, but not N300 classifiers. That is, both N400 and N300 classifiers are able to learn the pattern occurring in the N400 time window – which is observed throughout both the N300 and N400 time windows; but the N400 classifier additionally picks up on later patterns, perhaps reflecting an already initiated higher-level categorical process (see also Heikel et al., 2018). Evidence for the emergence of such a late positivity can be seen in the topographical maps of Fig. 3 and the spatial decoding patterns of Fig. 4. This finding, combined with the main finding of the similarity between N300 and N400, also establishes the potential of time-generalized decoding for illustrating the neurocognitive architecture of scene processing. It complements previous methods, and allows a new range of research questions to be addressed.

In sum, our results suggest that scene context already plays a role in early phases of object processing (Bar, 2004; Brandman and Peelen, 2017; Truman and Mudrik, 2018), without necessitating a two-component explanation of such effects. A more precise measurement of the onset of generalization depends on more specialized analyses of time-generalized cross-decoding, operating on higher-powered samples. However, *if* the observation of only one pattern sustained throughout the N300 and N400 time windows indeed indicates just one underlying neural event, then the neural substrates underlying comparatively high-level stages in scene processing/object-scene integration are already active very early, perhaps as early as 200 ms post stimulus presentation.

Acknowledgments

References

Bar, M., 2004. Visual objects in context. Nat. Rev. Neurosci. 5 (8), 617–629. https://doi.org/10.1038/nrn1476.

Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. Trends Cogn. Sci. 11 (7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005.

Bar, M., 2009. The proactive brain: memory for predictions. Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci. 364 (1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310.

Biederman, 1981. On the semantics of a glance at a scene. In: Kubovy, M., Pomerantz, J.R. (Eds.), Perceptual Organization. Lawrence Erlbaum, Hillsdale, New Jersey, pp. 213–263. Retrieved from: ⟨http://www.citeulike.org/user/ChristinaPavlo/article/6541577⟩.

Biederman, I., Mezzanotte, R.J., Rabinowitz, J.C., 1982. Scene perception: detecting and judging objects undergoing relational violations. Cogn. Psychol. 14 (2), 143–177.

Brandman, T., Peelen, M.V., 2017. Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. J. Neurosci. 37 (32) Retrieved from: ⟨http://www.jneurosci.org/content/37/32/7700⟩.

Castelhano, M.S., Heaven, C., 2011. Scene context influences without scene gist: eye movements guided by spatial associations in visual search (article). Psychon. Bull. Rev. 18 (5), 890–896. https://doi.org/10.3758/s13423-011-0107-8.

Cornelissen, T.H.W., Võ, M.L.-H., 2016. Stuck on semantics: processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. Atten. Percept. Psychophys. https://doi.org/10.3758/s13414-016-1203-7.

Davenport, J.L., Potter, M.C., 2004. Scene consistency in object and background perception. Psychol. Sci. 15 (8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x.

Demiral, Ş.B., Malcolm, G.L., Henderson, J.M., 2012. ERP correlates of spatially incongruent object identification during scene viewing: contextual expectancy versus simultaneous processing. Neuropsychologia 50, 1271–1285.

Draschkow, D., Võ, M.L.-H., 2017. Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. Sci. Rep. 7 (1), 16471. https://doi.org/10.1038/s41598-017-16739-x.

Federmeier, K.D., Kutas, M., 2001. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. J. Exp. Psychol.: Learn., Mem., Cogn. 27 (1), 202–224. https://doi.org/10.1037/0278-7393.27.1.202.

Ganis, G., Kutas, M., 2003. An electrophysiological study of scene effects on object identification. Cogn. Brain Res. 16 (2), 123–144. https://doi.org/10.1016/S0926-6410(02)00244-6.

Ganis, G., Kutas, M., Sereno, M.I., 1996. The search for "Common Sense": an electrophysiological study of the comprehension of words and pictures in reading. J. Cogn. Neurosci. 8 (2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7, 267. https://doi.org/10.3389/fnins.2013.00267.

Hamm, J.P., Johnson, B.W., Kirk, I.J., 2002. Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. Clin. Neurophysiol.: Off. J. Int. Fed. Clin. Neurophysiol. 113 (8), 1339–1350. Retrieved from: ⟨http://www.ncbi.nlm.nih.gov/pubmed/12140015⟩.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage 87, 96–110. https://doi.org/10.1016/J.NEUROIMAGE.2013.10.067.

Heikel, E., Sassenhagen, J., Fiebach, C.J., 2018. Time-generalized multivariate analysis of EEG responses reveals a cascading architecture of semantic mismatch processing. Brain Lang. 184, 43–53. https://doi.org/10.1016/J.BANDL.2018.06.007.

Henderson, J.M., Weeks, P.A., Hollingworth, A., 1999. The effects of semantic consistency on eye movements during complex scene viewing. J. Exp. Psychol. Human. Percept. Perform. 25 (1), 210–228. Retrieved from: ⟨http://psycnet.apa.orgjournals/xhp/25/1/210⟩.

Hollingworth, A., Henderson, J.M., 1999. Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. Acta Psychol. 102 (2–3), 319–343. https://doi.org/10.1016/S0001-6918(98)00053-5.

Jas, M., Engemann, D.A., Bekhti, Y., Raimondo, F., Gramfort, A., 2017. Autoreject: automated artifact rejection for MEG and EEG data. NeuroImage 159, 417–429. https://doi.org/10.1016/J.NEUROIMAGE.2017.06.030.

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.

King, J.-R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. 18 (4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Willing, C., 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides, F., Schmidt, B. (Eds.), Positioning and Power in Academic Publishing. Players, Agents and Agendas, pp. 87–90.

Kovalenko, L.Y., Chaumon, M., Busch, N.A., 2012. A pool of pairs of related objects (POPORO) for investigating visual semantic integration: behavioral and electrophysiological validation. Brain Topogr. 25 (3), 272–284. https://doi.org/10.1007/s10548-011-0216-8.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647. 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, 203–205.

Loftus, G.R., Mackworth, N.H., 1978. Cognitive determinants of fixation location during picture viewing. J. Exp. Psychol. Human. Percept. Perform. 4 (4), 565–572.

Lauer, T., Cornelissen, T.H.W., Draschkow, D., Willenbockel, V., Võ, M.L.-H., 2018. The role of scene summary statistics in object recognition. Sci. Rep. https://doi.org/10.1038/s41598-018-32991-1.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190. https://doi.org/10.1016/J.JNEUMETH.2007.03.024.

Mathôt, S., Schreij, D., Theeuwes, J., 2012. OpenSesame: an open-source, graphical experiment builder for the social sciences. Behav. Res. Methods 44 (2), 314–324. https://doi.org/10.3758/s13428-011-0168-7.

McPherson, B.W., Holcomb, P.J., 1999. An electrophysiological investigation of semantic priming with pictures of real objects. Psychophysiology 36 (1), 53–65. Retrieved from: ⟨http://www.ncbi.nlm.nih.gov/pubmed/10098380⟩.

Meade, G., Lee, B., Midgley, K.J., Holcomb, P.J., Emmorey, K., 2018. Phonological and semantic priming in American Sign Language: N300 and N400 effects. Lang. Cogn. Neurosci. 1–15. https://doi.org/10.1080/23273798.2018.1446543.

Mudrik, L., Lamy, D., Deouell, L.Y., 2010. ERP evidence for context congruity effects during simultaneous object-scene processing. Neuropsychologia 48 (2), 507–517. https://doi.org/10.1016/j.neuropsychologia.2009.10.011.

Mudrik, L., Shalgi, S., Lamy, D., Deouell, L.Y., 2014. Synchronous contextual irregularities affect early scene processing: replication and extension. Neuropsychologia 56 (1), 447–458. https://doi.org/10.1016/j.neuropsychologia.2014.02.020.

Nigam, A., Hoffman, J.E., Simons, R.F., 1992. N400 to semantically anomalous pictures and words. J. Cogn. Neurosci. 4 (1), 15–22. https://doi.org/10.1162/jocn.1992.4.1.15.

Nobre, A.C., McCarthy, G., 1994. Language-related ERPs: scalp distributions and modulation by word type and semantic priming. J. Cogn. Neurosci. 6 (3), 233–255. https://doi.org/10.1162/jocn.1994.6.3.233.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12 (Oct), 2825–2830. Retrieved from: ⟨http://www.jmlr.org/papers/v12/pedregosa11a.html⟩.

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 Lang. 137, 29–39. https://doi.org/10.1016/J.BANDL.2014.07.010.

Schendan, H., Kutas, M., 2002. Neurophysiological evidence for two processing times for visual object identification. Neuropsychologia 40 (7), 931–945. https://doi.org/10.1016/S0028-3932(01)00176-2.

Schendan, H., Maher, S., 2009. Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. NeuroImage 44 (4), 1423–1438. https://doi.org/10.1016/j.neuroimage.2008.09.061.

Sitnikova, T., Holcomb, P.J., Kiyonaga, K.A., Kuperberg, G.R., 2008. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. J. Cogn. Neurosci. 20 (11), 2037–2057. https://doi.org/10.1162/jocn.2008.20143.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44 (1), 83–98. https://doi.org/10.1016/J.NEUROIMAGE.2008.03.061.

Trapp, S., Bar, M., 2015. Prediction, context, and competition in visual recognition. Ann. N.Y. Acad. Sci. 1339 (1), 190–198. https://doi.org/10.1111/nyas.12680.

Truman, A., Mudrik, L., 2018. Are incongruent objects harder to identify? The functional significance of the N300 component. Neuropsychologia 117, 222–232. https://doi.org/10.1016/j.neuropsychologia.2018.06.004.

Urbach, T.P., Kutas, M., 2002. The intractability of scaling scalp distributions to infer neuroelectric sources. Psychophysiology 39 (6), 791–808. https://doi.org/10.1017/S0048577202010648.

Verleger, R., 1997. On the utility of P3 latency as an index of mental chronometry. Psychophysiology 34 (2), 131–156. https://doi.org/10.1111/j.1469-8986.1997.tb02125.x.

Viola, F.C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., Debener, S., 2009. Semi-automatic identification of independent components representing EEG artifact. Clin. Neurophysiol. 120 (5), 868–877. https://doi.org/10.1016/J.CLINPH.2009.01.015.

Võ, M.L.-H., Henderson, J.M., 2010. The time course of initial scene processing for eye movement guidance in natural scene search (14.1-13). J. Vision. 10 (3). https://doi.org/10.1167/10.3.14.

Võ, M.L.-H., Wolfe, J.M., 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. Psychol. Sci. 24 (9), 1816–1823. https://doi.org/10.1177/0956797613476955.

Võ, M.L.-H., Wolfe, J.M., 2015. The role of memory for visual search in scenes. Ann. N.Y. Acad. Sci. 1339, 72–81. https://doi.org/10.1111/nyas.12667.

Willems, R.M., Özyürek, A., Hagoort, P., 2008. Seeing and hearing meaning: erp and fMRI evidence of word versus picture integration into a sentence context. J. Cogn. Neurosci. 20 (7), 1235–1249. https://doi.org/10.1162/jocn.2008.20085.