Disentangling Stimulus Plausibility and Contextual Congruency: Electro-Physiological Evidence for Differential Cognitive Dynamics

Moreno I. Coco[a,e], Susana Araujo[b,e], Karl Magnus Petersson[b,c,d]

[a]*School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 3 Charles Street, Edinburgh, EH8 9AD, UK (tel. +441316517112, email: moreno.cocoi@gmail.com)*
[b]*Cognitive Neuroscience Research Group, Centre for Biomedical Research (CBMR), University of Algarve*
[c]*Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands*
[d]*Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands.*
[e]*Faculdade de Psicologia, Universidade de Lisboa*

Abstract

Expectancy mechanisms are routinely used by the cognitive system in stimulus processing and in anticipation of appropriate responses. Electrophysiology research has documented negative shifts of brain activity when expectancies are violated within a local stimulus context (e.g., reading an implausible word in a sentence) or more globally between consecutive stimuli (e.g., a narrative of images with an incongruent end). In this EEG study, we examine the interaction between expectancies operating at the level of stimulus *plausibility* and at more global level of contextual *congruency* to provide evidence for, or against, a disassociation of the underlying processing mechanisms. We asked participants to verify the congruency of pairs of cross-modal stimuli (a sentence and a scene), which varied in plausibility. ANOVAs on ERP amplitudes in selected windows of interest show that congruency violation has longer-lasting (from 100 to 500ms) and more widespread effects than plausibility violation (from 200 to 400ms). We also observed critical interactions between these factors, whereby incongruent and implausible pairs elicited stronger negative shifts than their congruent counterpart, both early on (100-200ms) and between 400-500ms. Our results suggest that the integration mechanisms are sensitive to both global and local effects of expectancy in a modality independent manner. Overall, we provide novel insights into the interdependence of expectancy during meaning integration of cross-modal stimuli in a verification task.

Keywords: Event plausibility, contextual congruency, cross-modal verification, stimulus processing, electro-physiology

1. Introduction

The cognitive system heavily relies on expectations of real-world events to optimize the processing of incoming information and forward appropriate responses (Rao and Ballard, 1999; Bar, 2007; Friston, 2010; Wacongne et al., 2012; Clark, 2013; Pickering and Clark, 2014). Behavioral and neural evidence suggests that expectancy mechanisms are found across a variety of tasks. During reading, for example, the predictability of a word directly mediates the amount of attention allocated and associated patterns of brain activity (e.g., Kutas and Hillyard 1980; Van Berkum et al. 1999; Halgren et al. 2002; DeLong et al. 2005, and Rayner 2009; Kutas and Federmeier 2011 for reviews on the topic). Similar findings are obtained in visual tasks, where the expected target location regulates eye-movement responses, memory recognition, and associated brain activity (e.g, Biederman et al. 1973; Loftus and Mackworth 1978; De Graef et al. 1990; Boyce and Pollatsek 1992; Henderson et al. 1999; Davenport and Potter 2004; Võ and Wolfe 2013; Coco et al. 2014).

Expectation [1] is an important concept in electro-physiology (EEG) research on the dynamics of stimulus processing, and the underlying mechanisms of semantic integration. A key observation in these studies is that negative shifts in the EEG activity may reflect processing costs due to expectation violations of linguistic and non-linguistic stimuli. With linguistic stimuli, for example, a seminal study by Kutas and Hillyard 1980 demonstrated that an unexpected word within a sentence (e.g., *the boy spreads butter with* ***socks***) generates negative EEG activity around 400 ms from stimulus onset (i.e., N400 ERP component), when compared to an expected word (e.g., ***knife***). Likewise with non-linguistic stimuli (e.g., a visual scene), a N390 is found when participants watch a visual scene (e.g., a soccer field with a player) and an unexpected (vs. expected) object is cued in it (e.g., a toilet-roll vs. a ball, Ganis and Kutas 2003). Moreover, earlier negative shifts are also observed (between 250-300 ms) when unexpected objects are embedded in the scene (Mudrik et al., 2010; Võ and Wolfe, 2013; Mudrik et al., 2014).

Ample evidence has been gathered about the N400 component (e.g., Kutas et al. 2006; Hagoort and van Berkum 2007; Lau et al. 2008 for reviews); but its root causes are still debated (e.g., Kutas and Federmeier 2011). In fact, even though negative shifts are observed when unexpected stimuli are processed, a wide range of factors is directly implicated in the latency and distribution of such shifts. One of the most important factor is the contextual information that surrounds an unexpected stimulus.

In particular, two types (or levels) of context can be distinguished: (a) *local*, such as a short sentence enclosing an unexpected word (e.g., Marslen-Wilson and Tyler 1980; Kutas and Hillyard 1980; DeLong et al. 2005), or an image onto which an additional visual stimulus is superimposed (e.g., Ganis and Kutas 2003); and (b) *global*, such as a discourse preamble before reading the critical sentence (e.g., Kutas 1993; Camblin et al. 2007; Menenti et al. 2009; Lau et al. 2013) or a narrative of images with an incongruent ending (e.g., West and Holcomb 2002; Sitnikova et al. 2008; Cohn et al. 2012). The information conveyed by a global context bears direct consequences on the processing of a local context, as observed with both linguistic and non-linguistic information. Camblin et al. (2007), for example, showed that N400 effects elicited by unassociated word pairs (e.g, arms-nose, versus the associated arms-legs) in a local sentence context, can be reduced when preceded by a supportive global discourse statement. West and Holcomb (2002) similarly found a large negativity (at ≈ 300 and $\approx 500ms$ after scene onset) when presenting a global narrative of images and an incongruous ending image (local context) than a congruous one. Furthermore, larger negativities are observed for scenes containing ambiguous objects, especially when the context of the scene is neutral with respect to the semantics of the object (Dyck and Brodeur, 2015). Moreover, a global context (e.g., a narrative of images depicting a man cutting a loaf of bread) could generate expectations

[1] In this study, we mainly discuss the notion of *expectation*, rather than *predictability*, and refer to processing mechanisms that are mediated by the likelihood of expectancy at a local (within a stimulus) or a more global (across stimuli) scale. We avoid predictability, because it generally refers to incremental processing in psycholinguistic research, such as expecting a particular word, given its prior context, before it is actually presented. This is not the type of experimental manipulation implemented in the current study, so we opted against entering into this debate. We refer the interested reader to Van Petten and Luka 2012 for an insightful discussion about predictability and expectation.

that might or might not be consistent with a local context (e.g., a final image where the man is ironing rather than cutting the bread). Sitnikova et al. (2008) investigated this particular case showing earlier, and longer-lasting, negative shifts when the congruency between global and local context was violated as compared to when the local context was congruent with the global context.

Verification tasks also provide additional insights about the role of congruency on processing costs. Dikker and Pylkkanen (2011), for example, used a word-picture matching task, and demonstrated that when the content of a word does not completely match the content of a subsequently presented picture, a negative shift of brain activity is observed as early as 100 ms after picture onset (cf., Brunellière et al. 2013 for corroborating evidences in spoken word recognition). Similar results are obtained with other cross-modal verification tasks when the congruency is manipulated between: (a) the source of an audio signal and its location in the visual context (i.e., left and right) (Teder-Sälejärvi et al., 2005), or (b) the emotional valency of speech and an associated face expression (Pourtois et al., 2000).

To sum up, negative shifts of EEG brain activity result from expectation violations. Expectancy mechanisms seem to operate at two levels: (1) the local plausibility of a specific stimulus, and (2) the congruency between a global and a local context. In the current study, we precisely examine the processing costs arising when both types of expectancy are simultaneously violated. Our main goal is to provide evidences for, or against, a disassociation of expectancy mechanisms driven by stimulus plausibility and message congruency. We do so by designing a cross-modal (sentence-scene) verification paradigm, which naturally affords a fully-crossed 2x2 design of plausibility and congruency [2] (refer Clark and Chase 1972; Carpenter and Just 1975 for seminal psycholinguistic work on this task).

Participants first read a sentence (plausible or not, e.g., *the boy is eating a* ***brick***), building a global context, and then are exposed to a visual scene (local context), which matches it, or not, in content (e.g., a picture depicting a boy eating a brick, refer to Figure 1 for an example of the material used in this study). By examining EEG responses at the onset of the scene, we can capture how expectations from the global context interact with the plausibility in the downstream local context. This allows us to disentangle the mechanisms of congruency from those driven by plausibility under the same experimental design [3].

If different processing mechanisms are involved when the congruency between contexts is assessed and the plausibility of the stimuli is evaluated, then we should observe different ERP latencies and distributions when either, or both, are violated. Moreover, if such factors jointly contribute to the processing cost, we should observe an interaction between the two, i.e., the more the violations, the higher the processing cost.

[2] Note, our design departs from Sitnikova et al. (2008) by having a cross-modal verification paradigm where plausibility of stimuli can directly interact and compete with expectation processes of congruency.

[3] Differently from Knoeferle et al. 2011, we present the sentence as a global context for the scene, rather than vice-versa; and focus on the electro-physiological response during the processing of visual information.

First, we expect to replicate previous literature with respect to the main effects of congruency and plausibility. In line with cross-modal verification studies (e.g., Dikker and Pylkkanen 2011), we predict an early effect of congruency driven by the congruency/incongruency between the stimuli (i.e., sentence and scene), whereby a larger negative shift is expected with incongruous as compared to congruous trials, between 100-200 ms). Incongruent trials are also expected to display a larger negative shift between 300-400 ms and 400-500 ms (e.g., West and Holcomb 2002; Sitnikova et al. 2008). Plausibility, instead, is expected to kick in between 200-300 ms and 400-500 ms with implausible scenes triggering a larger negativity than plausible scenes (see Ganis and Kutas 2003; Mudrik et al. 2010; Sun et al. 2011; Võ and Wolfe 2013; Mudrik et al. 2014).

Second, and perhaps most importantly, our study makes it possible to establish whether these two sources of expectancy jointly contribute to processing costs. We expect a larger negativity for incongruent stimuli conveying implausible content where both plausibility and congruency are simultaneously violated. We predict this specific interaction to occur as soon as the content of both sentence and scene becomes available for the verification response (i.e., between 100-200 ms, Dikker and Pylkkanen 2011) and later on when such content needs to be integrated (i.e., 400-500 ms, akin to N400 effects). In fact, if the N400 reflects processes of semantic integration, and it is sensitive to global and local expectancies, then in this temporal window we should observe two costs on processing: one to integrate implausible information to the semantic network, and the other to resolve the incongruency between stimuli.

2. The present study

2.1. Method

The experimental design crossed *Plausibility* (Plausible, Implausible) of the information depicted in the sentence and scene with their *Congruency* (Congruent, Incongruent) as a pair, within participants. In previous literature, the term congruency was adopted to indicate both an inconsistency within a local context (e.g., Mudrik et al. 2010), as well as, the mismatch between a global and a local context (e.g., West and Holcomb 2002; Sitnikova et al. 2008). Here, *Congruency* indicates whether the content of the sentence-scene pair matched or not in content; and *Plausibility* whether the stimuli had expected or unexpected content (see Figure 1 for a visualization of the design and examples of the stimuli used).

Participants. Nineteen students (11 males and 8 females; mean age = 24.73 ± 5.02 years) at the University of Algarve, all native speakers of Portuguese, volunteered to participate in the study. One participant, from an initial pool of twenty, had to be discarded because the EEG signal was severely contaminated by electric noise. The experiment was granted by the Ethics Committee of the Department of Psychology, in accordance with the University's Ethics Code of Practice.

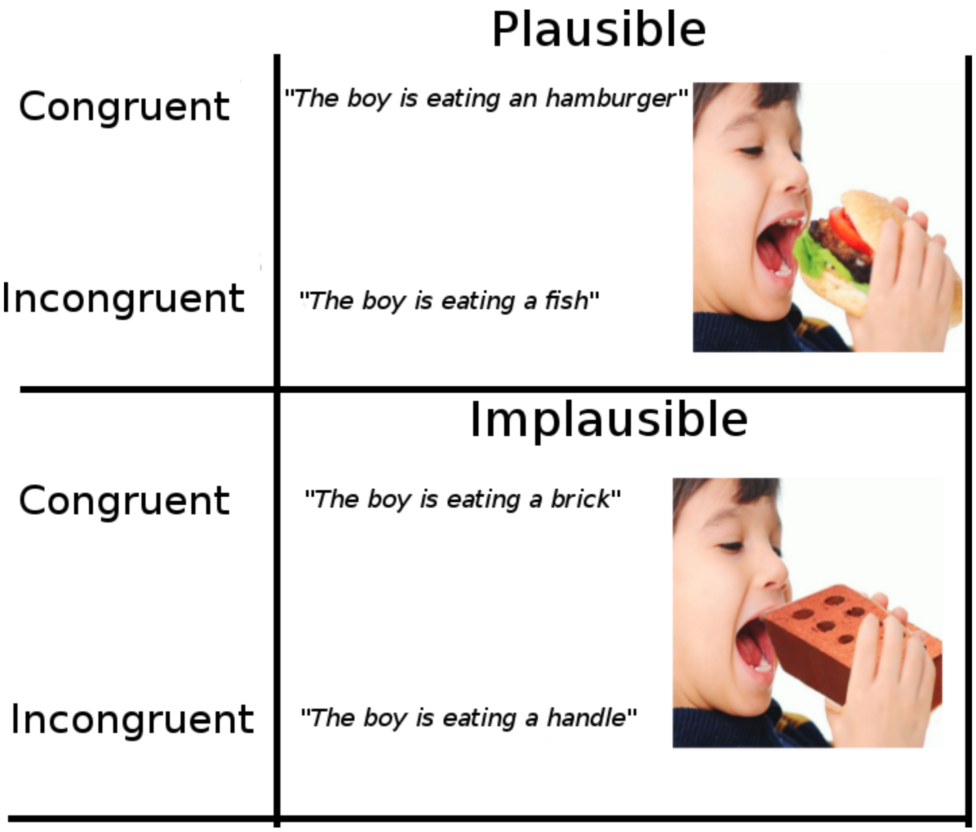
Plausible
Congruent
"The boy is eating an hamburger"
Incongruent
"The boy is eating a fish"
Implausible
Congruent
"The boy is eating a brick"
Incongruent
"The boy is eating a handle"

Figure 1: Experimental design with a full set of crossed pairs of sentence-scene stimulus pairs: Plausibility (Plausible and Implausible) and Congruency (Congruent, Incongruent).

Materials. We used 125 photo-realistic scenes, originally published in Mudrik et al. (2010), and added another 100 scenes based on open-access material from the Internet (e.g., Flickr). The target object was pasted into the scene using the free software GIMP. The size of each scene was scaled to 550 x 550 pixels. Each scene was in two Plausibility conditions (Plausible: a boy eating a hamburger, Implausible: a boy eating a brick. We computed the visual saliency of each image in its plausible and implausible version (both in our new set and in the original set by Mudrik et al. 2010) using the models by Walther and Koch (2006) (WK), and the Adaptive Whitening Saliency model by Garcia-Diaz et al. (2012) (AWS) to make sure that the two versions of the same image did not differ in low-level features. Paired-samples t-tests showed no significant difference of visual saliency between plausible and implausible images using WK[t(148) = -0.02, p = 0.9] or AWS[t(148) = -0.62, p = 0.5]. Moreover, we compared the images on their luminance (L), number of edges (E), and visual clutter (VC, Rosenholtz et al. 2007). Paired t-tests showed again no difference on Plausibility in L[t(148) = 0.36, p = 0.7], E[t(148) = -0.45, p = 0.6] and VC[t(148) = -0.11, p = 0.9]. We also confirmed that in the original set by Mudrik et al. (2010), there was no difference of visual saliency between plausible and implausible images with WK[t(296) = -0.13, p = 0.9] and AWS[t(296) = 0.09, p

= 0.9] model. These results guarantee that the effects observed on the electro-potential activity can be genuinely attributed to the plausibility of the images rather than to their low-level properties.

We crossed Plausibility with Congruency (Congruent, Incongruent) by pairing each scene in the Plausible conditions with two different sentences. The sentence material consisted of 900 unique sentences in total (i.e., 450 scenes paired with 2 different sentences). For example, in a plausible and congruent trial, a participant might read, *the boy is eating a hamburger*, and then view a scene correctly depicting that event. For a plausible and incongruent trial, the participant might read, *the boy is eating a fish* and then view a scene depicting the boy eating a hamburger. The same reasoning was applied to construct the implausible cases [4], see Figure 1. The sentences were written in Portuguese and checked for grammaticality by two independent native-speaking annotators. The target word (e.g., *hamburger* vs. *brick*) was always positioned at the end of the sentence. The annotators also ensured that the target object depicted in the scene was recognized as the target word used in the sentence.

In order to assess how the participants perceived plausibility and congruency during the experiment, we asked them, at the end of each trial, to rate (1) the plausibility of the scene and (2) the congruency between the scene and the sentence, on a scale from 1 to 6 (i.e., from *completely implausible—incongruent* to *completely plausible—congruent*). For the plausibility rating, only the previously shown scene was displayed. For the congruency rating, both stimuli (sentence and scene) were displayed together in the same slide. There was no time limit to answer. For (1), we observed a mean rating of 5.44 for the Plausible scenes, and 1.67 for Implausible scenes. This difference was significant according to a Kruskal-Wallis test ($\chi^2(1) = 2892, p = .001$). Moreover, for (2) we observed a mean rating of 5.19 for the Congruent pairs, and 1.75 for Incongruent pairs. Also this difference was statistically significant according to a Kruskal-Wallis test ($\chi^2(1) = 2259, p < .001$). These results confirmed the efficacy and validity of our experimental conditions. We refer the reader also to Coco and Duran in press, where the same material was used, and identical ratings observed on a pool of 64 participants. Furthermore, we also made sure that the effect of plausibility was not confounded by other sources of information such as the lexical frequency of target words, which is known to mediate stimulus processing (see Rayner and Duffy 1986 for an example in reading research). In particular, we used the SUBTLEX-PT (Soares et al., 2015), which is the largest lexical database for the Portuguese language to date, computed the lexical frequency of plausible (2065 ± 3263) and implausible (1895 ± 4011) words and found no difference between conditions (t = .71, p = .5).

Experimental Procedure. Each trial started with a fixation cross presented for 500 ms in the center of the screen. The sentence was then presented to the participant, one word at time (200 ms) with an inter-stimulus of 200 ms

[4] Note that the incongruent condition could have also been obtained by: (a) fixing the plausibility of the sentence (plausible or implausible) and manipulating the associated scenes (four versions), or (b) mixing plausible sentences with implausible scenes (or vice-versa). However, scene material is much harder to construct, and there are many more ways to construct incongruent than congruent cases, i.e., we would have had an unbalanced design. Thus, we opted against these alternative methods.

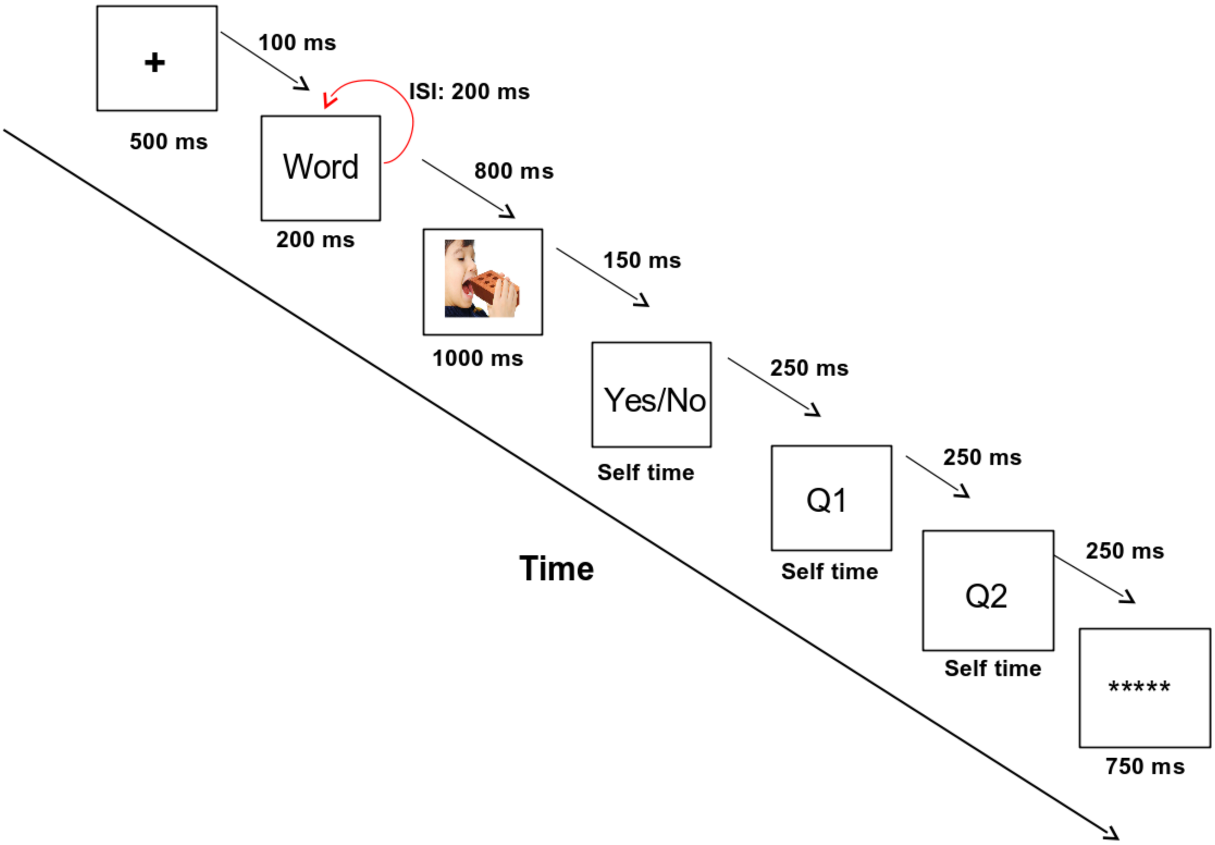

Figure 2: An example of a trial run.

between words. 800 ms after the last target word was shown, the visual scene was displayed for 1000 ms. The duration of the preview time was based on previous work using the same stimuli (Mudrik et al., 2010), which gives enough time to extract scene information and identify the critical target object. After the scene disappeared, the participant was asked on-screen whether the information conveyed by the sentence matches/mismatches the information conveyed by the scene (i.e., a yes/no alternative forced choice). There was no time limit for the participant to provide the response, using key-press. As said in the previous paragraph, the participant was subsequently asked two more questions in two separate screens, where we collected Likert ratings (1-6) about the plausibility of the scene ('This scene is plausible'), and the congruency between the sentence and the scene ('The scene matches the sentence'). For these two additional questions, the stimuli were presented again, and the participant was under no time limit. These additional scores provided us with a subjective, trial-by-trial, measure of plausibility and congruency. The new trial started after a phase for blinking, and upon self-response (see Figure 2). Before the task, the participants practiced six trials to familiarize with the task. Participants sat between 60 and 70 cm from the computer screen. All words were presented in lowercase (Arial; font size 47; black font on white background), at eye-level at the center of the screen, and ranged from 2.2 to 3.8 visual angle.

The 900 total stimuli (225 × 4 conditions) were distributed in 4 lists using a Latin Square Design (225 trial each list). So each list was made of 56 items in each experimental condition, except one condition that had to contain 57 items. Each participant was assigned to one of the lists, and the presentation order was randomized. Presentation software (version 13; nbs. neuro-bs.com/presentation) was used to display the stimuli on a computer screen and to record behavioral responses. The experiment took approximately one hour to be completed.

2.2. Analysis

EEG method. Electroencephalogram (EEG) was acquired through an ActiveTwo Biosemi electrode system from 64 Ag/AgCl active scalp electrodes, mounted in an elastic cap. The electrodes were located at standard left and right hemisphere, positioned over the frontal, parietal, occipital, and temporal areas according to the International 10/20 system guidelines. We had 10 electrodes across the midline, and 27 over each hemisphere. Two additional electrodes (CMS/DRL nearby Pz) were used as an online reference (for further details see `http://www.biosemi.com`; BioSemi, Amsterdam, The Netherlands, and refer to Schutter et al. 2006).

Three other electrodes were attached over the right and left mastoids and below the right eye, to monitor eye movements, blinks, and muscular contractions. We used an ActiveTwo Biosemi amplifier (DC-67 Hz bandpass, 3 dB/octave) to increase the EEG signal, and sampled it at a rate of 512 Hz.

ERP analysis. The electro-physiological data was analyzed using the FieldTrip, open source MATLAB toolbox (`http://fieldtrip.fcdonders.nl/`, Oostenveld et al. 2010)

We focused on EEG responses time-locked to the onset of the scene, over an epoch of 900 ms (from 100 ms prior to the onset of the stimulus till 800 ms after it), where effects of congruency and plausibility are expected to interact. In fact, this is the moment of the verification task, where both stimuli (sentence-scene), with their inherent plausibility value, can be integrated to perform an informed congruency judgment (refer to Figure 2 for an example of a trial run).

On a total of 4275 trials (i.e., 19 participants, 225 items each), we excluded 1216 trials ($\approx$ 28% of the data) which: (a) contained eye-blink, oculomotor, muscle artifacts or were contaminated by electric noise (N = 725, $\approx$ 17% of the data) or (b) when participants incorrectly responded to the congruency verification (N = 491, $\approx$ 12% of the data).

The 3059 trials considered for the analysis were zero-phase (forward and reverse) low-pass filtered off-line (30 Hz) and baseline corrected to the mean-amplitude of 100 ms pre-stimulus period, independently for each participant. We computed evoked-potentials, (i.e., ERP) by averaging single-trial EEG, and obtained grand-average for each experimental condition by averaging all by-participants ERPs.

Electrodes	Region
Fp1, AF7, AF3, F3, F5, F7	Left Frontal
F1, Fpz, Afz, Fz, F2	Mid Frontal
FT7, FC5, FC3, C3, C5, T7, TP7, CP5, CP3	Left Central
FC1, C1, CP1, CPz, FC2, FCz, Cz, C2, CP2	Mid Central
P1, Iz, Oz, POz, Pz, P2	Mid Parietoccipital
P3, P5, P7, P9, PO7, PO3, O1	Left Parietoccipital
Fp2, AF8, AF4, F4, F6, F8	Right Frontal
FT8, FC6, FC4, C4, C6, T8, TP8, CP6, CP4	Right Central
P4, P6, P8, P10, PO8, PO4, O2	Right Parietoccipital

Table 1: Division of the electrodes into regions.

We statistically analyzed the mean-amplitudes of the ERP from 100 to 500 ms after scene onset in windows of 100 ms each. This resulted in the following windows of interest: (a) 100-200 ms for early effects of congruency observed in a similar picture-word verification task (e.g, Dikker and Pylkkanen 2011), (b) 200-300 ms for effects of plausibility (e.g., Mudrik et al. 2010, 2014), (c) 300-400 ms for effects of image congruency (e.g., West and Holcomb 2002; Sitnikova et al. 2008) and (d) 400-500 ms, for effects of semantic integration and expectancy violation [5].

We assessed statistical differences using 4-way ANOVAs (IBM SPSS) with Plausibility (Plausible, Implausible), Congruency (Congruent, Incongruent), Region (Frontal, Central, Occipito-parietal) and Laterality (Left, Midline, Right) as within-factors (please refer to Table 1 for a summary of the electrodes distribution in Regions, and to Mudrik et al. 2014 for an identical grouping). We applied Greenhouse Geisser adjustments to correct for violations of sphericity and Bonferroni correction for multiple comparisons. Post-hoc analyses (Tukey HSD) were also conducted to evaluate the source of the relevant interactions. We visualize the ERP grand-mean amplitude, for Plausibility and Congruency in two multi-panel plots Region x Laterality (i.e., 9 plots each). Moreover, in order to display possible interactions between congruency and plausibility, we visualize in a bar-plot all four experimental conditions across the three regions of the head, at the different latencies analyzed.

(In Appendix A, we report EEG responses time-locked to the onset of the last word up to the onset of the

[5] We also tried a larger time-window from 300-500 ms, also reported in previous literature (e.g., Mudrik et al. 2010). But, we found that such an aggregation confounded the exact latency of an important interaction between congruency and plausibility, i.e., between 400-500 ms. So, we decided to consider smaller windows instead, and have a finer temporal resolution about the effects of our experimental conditions.

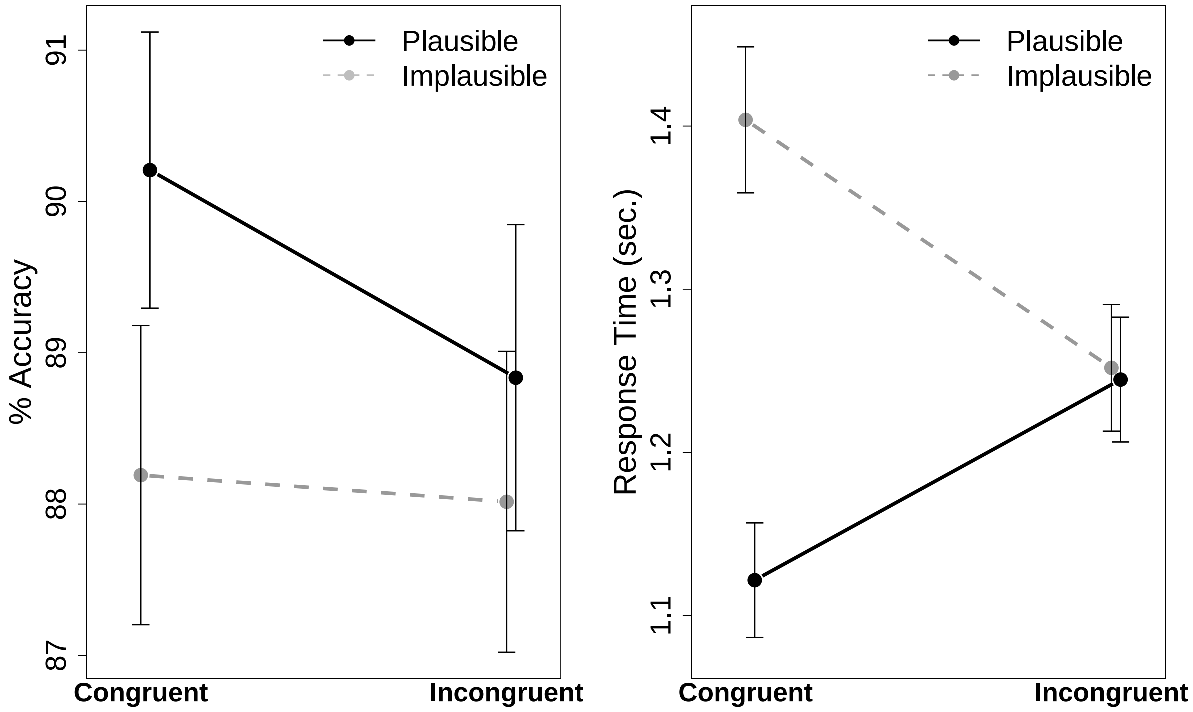

Figure 3: Interaction plot (means and standard error) for: percentage accuracy of the verification responses (left panel), and the associated response time on correct trials only (right panel) across experimental conditions: Congruency (Congruent, Incongruent) displayed along x-axis, and Plausibility displayed as lines (Plausible: black-solid; Implausible: gray-dashed).

scene (i.e., 800 ms) to: (1) replicate previous literature about the effects of word plausibility during language
205 comprehension (i.e., N400 for implausible words); and (2) demonstrate that any effect of word plausibility on the EEG responses is exhausted by the time the scene is presented. In Appendix B, instead, we present corroborating results using a non-parametric cluster-based permutation approach.)

3. Results and Discussion

3.1. Behavioral responses

210 In Figure 3, we visualize the observed data (accuracy and response time) across the different experimental conditions. ANOVA (F1, i.e., by-participants) on response accuracy showed only a significant main effect of Plausibility $[F(1,18) = 4.6, p < .05]$, whereby participants committed more errors for implausible than plausible stimuli. The main effect of Congruency was not significant $[F(1,18) = 1.07, p = .3]$, nor the interaction between Plausibility and Congruency $F < 1$. On response time (correct trials only), we confirmed a main effect of plausibility,

whereby participants took longer responding to implausible than plausible stimuli $[F(1,18) = 8.16, p < .01]$, and a significant interaction between Plausibility and Congruency $[F(1,18) = 5.03, p = .04]$. Post hoc-analyses revealed that when scenes were congruent with the content of the sentence, participants took longer to respond when the stimuli were implausible (p =.003). The same was not true when the pairs of stimuli were incongruent.

These behavioral results are largely consistent with previous literature, which found that implausible stimuli were harder to process than plausible stimuli (e.g., Davenport and Potter 2004). Moreover, the interaction between congruency and plausibility on the response time was an exact replication of results from another study, which used the same verification paradigm, experimental conditions and stimuli, but used an action-dynamics (i.e., computer-mouse tracking) approach (Coco and Duran, in press); see General Discussion for a comparison of the results obtained in the two studies. It is also interesting to notice the lack of interaction between Congruency and Plausibility on response accuracy. This may have resulted from the fact that participants were under no time pressure to provide the verification response.

3.2. ERP mean-amplitude responses at scene processing

In Figure 4 and 5, we visualize the ERP mean-amplitude time-locked to the onset of the scene, contrasting the two conditions of Congruency (Congruent, Incongruent) and Plausibility (Plausible, Implausible) respectively. The plot suggests that implausible stimuli, as well as incongruent pairs, elicited stronger negative shifts in the ERP than plausible stimuli or congruent pairs. However, in order to obtain statistically informed insights on the temporal dynamics of these effects, and uncover any possible interaction between plausibility and congruency, we examined four windows: (a) 100-200 ms, (b) 200-300 ms, (c) 300-400 ms, and (d) 400-500 ms. In Figure 6, we visualize the ERP responses of each window of interest for the four experimental conditions across the regions of the head.

100-200 ms. We find significant main effects of Congruency $[F(1,18) = 5.8, p = .03]$, Region $[F(2,36) = 40.3, p < .001$, and Laterality $[F(2,36) = 17.9, p < .001]$ (the reader is referred to Table 1 for an overview of the Regions). In particular, incongruent pairs elicited a stronger negativity on the ERP than congruent pairs. Frontal areas displayed a stronger negativity than Central and Parieto-occipital areas. ERPs were more negative at the mid-line sites than on the left and right hemispheres. We also found a Laterality by Region interaction $[F(2.02, 36.34) = 24.4, p < .001]$ because at central regions the ERPs were more negative at the mid-line sites than on the left and right hemispheres ($p < .001$ for both contrasts), while at frontal and parieto-occipital regions the ERPs were more widespread. Crucially, we also observed three-way interactions between Congruency, Region, and Laterality $[F(4,72) = 3.0, p = .036]$ and Congruency, Plausibility, and Region $[F(1.32, 23.81) = 4.0, p = .048]$. Post-hoc analysis revealed that, at frontal and central sites, incongruent pairs of stimuli elicited a larger negativity than congruent pairs, but such effect was restricted to implausible stimuli ($p < .001$ at frontal sites, and $p = .015$ at central sites).

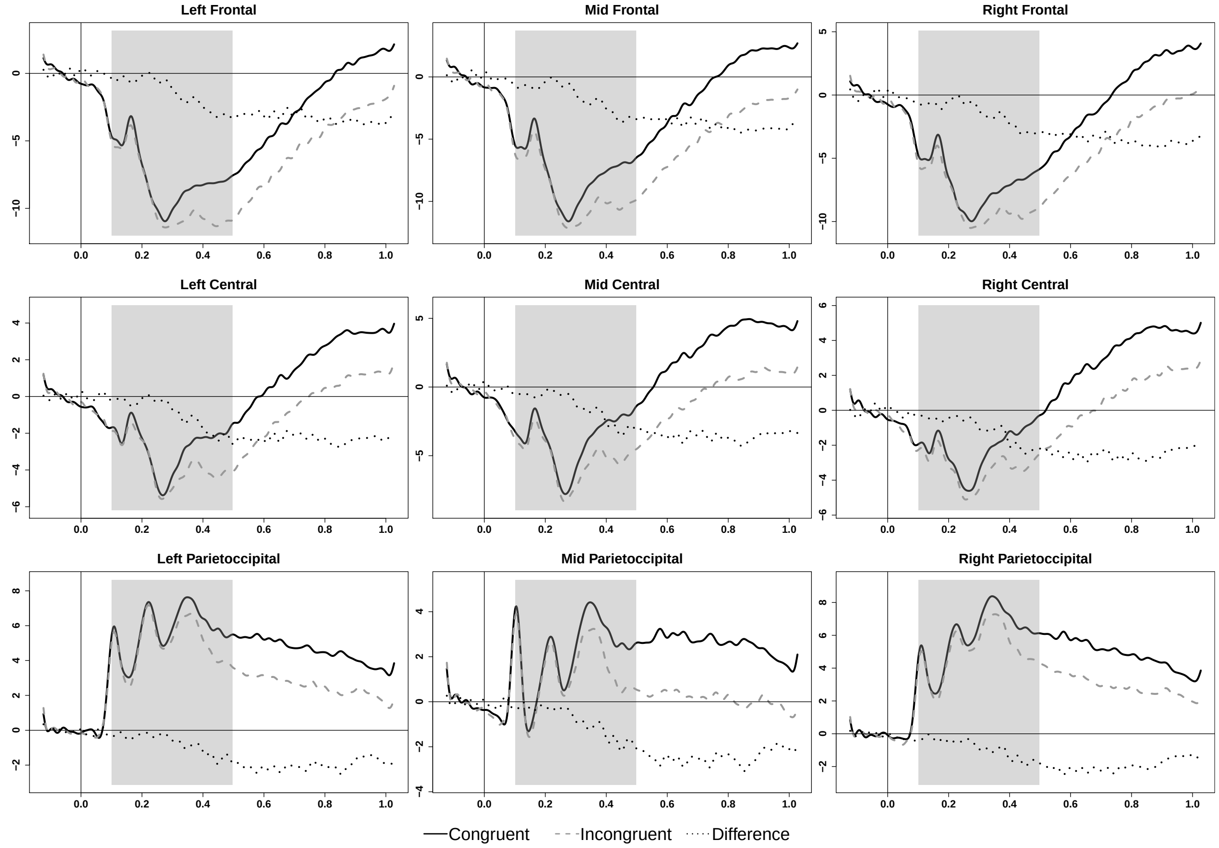

Figure 4: ERP responses at the onset of the visual scene when Congruent (tick line) or Incongruent (dashed-line) with the preceding sentence. We also plot the difference between the two waves (dotted-line). We mark with gray shaded rectangles the temporal latencies that were statistically analysed.

200-300 ms. Significant main effects of Plausibility $[F(1,18) = 5.35, p = .03]$, Laterality $[F(2,36) = 24.19, p < .001]$ and Region $[F(2,36) = 60.56, p < .001]$, as well as, interactions between Plausibility and Region $[F(1.18,21.27) = 8.1, p < .001]$ and Region and Laterality $[F(2.4,43.34) = 18.76, p < .001]$ were observed. In particular, implausible stimuli elicited a larger negativity than plausible stimuli and ERPs were more negative over frontal areas, especially at the mid-line sites. Moreover, post-hoc analyses showed that implausible stimuli were associated with a larger negativity than plausible stimuli, especially in frontal as compared to central and parieto-occipital sites ($p < .001$ for both contrasts).

300-400 ms. Significant main effects of Congruency $[F(1,18) = 12.2, p < .005]$, Plausibility $[F(1,18) = 6.0, p = .025]$, Laterality $[F(2,36) = 15.3, p < .001]$ and Region $[F(1.09,19.63) = 66, p < .001]$ were observed. In particular, implausible stimuli, or incongruent pairs elicited a larger negativity than plausible or congruently matching stimuli. The interaction between Congruency and Region $[F(1.13,20.38) = 3.7, p = .035]$ was also significant,

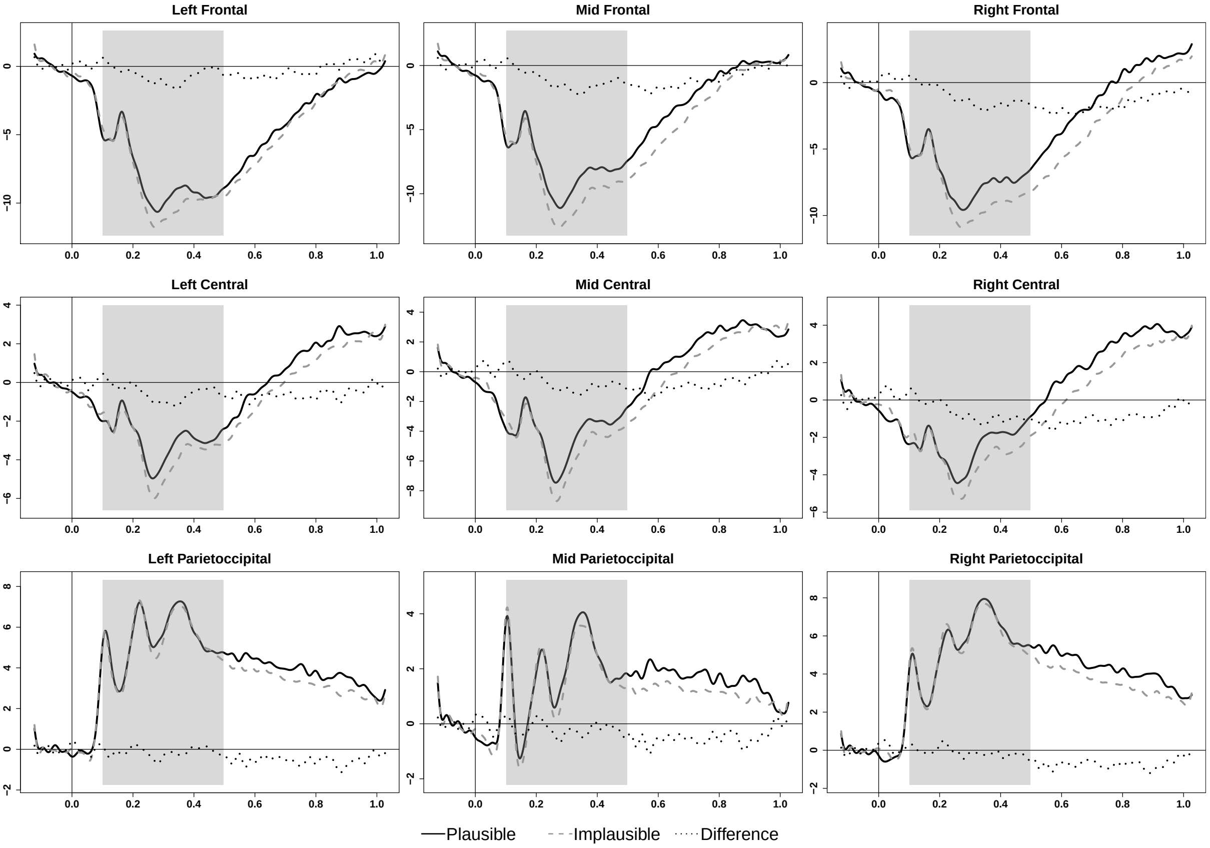

Figure 5: ERP responses at the onset of the visual scene with Plausible (tick line) or Implausible (dashed-line) content. We also plot the difference between the two waves (dotted-line). We mark with gray shaded rectangles the temporal latencies that were statistically analysed.

as well as the interaction between Plausibility and Region $[F(1.21,21.74) = 13.8, p < .001]$. Post-hoc analysis revealed that the effect of congruency was widespread (all $Ps < .005$), while the effect of plausibility had a more fronto-central distribution ($p < .001$ for both contrasts). The interaction between Congruency and Plausibility was not significant ($p > .2$).

400-500 ms. Significant main effects of Congruency $[F(1,18) = 27.2, p < .001]$, Laterality $[F(2,36) = 10.3, p < .001]$, and Region $[F(1.08,19.45) = 40.8, p < .001]$, as well as the interaction between Region and Laterality $[F(4,72) = 21.6, p < .001]$. The effect of Congruency was modulated by Region $[F(1.12,20.19) = 6.6, p < .005]$ and by Laterality $[F(2,36) = 3.6, p = .037]$. Likewise, Plausibility interacted with Region $[F(1.21,21.83) = 4.3, p = .021]$ and Laterality $[F(2,36) = 4.4, p = .020)]$, but was not significant as a main effect. In particular, implausible and incongruent stimuli, elicited a stronger negativity than implausible and congruently matching pairs at mid-right, fronto-central sites ($p < .001$ for all contrasts). Crucially, the interaction between Congruency

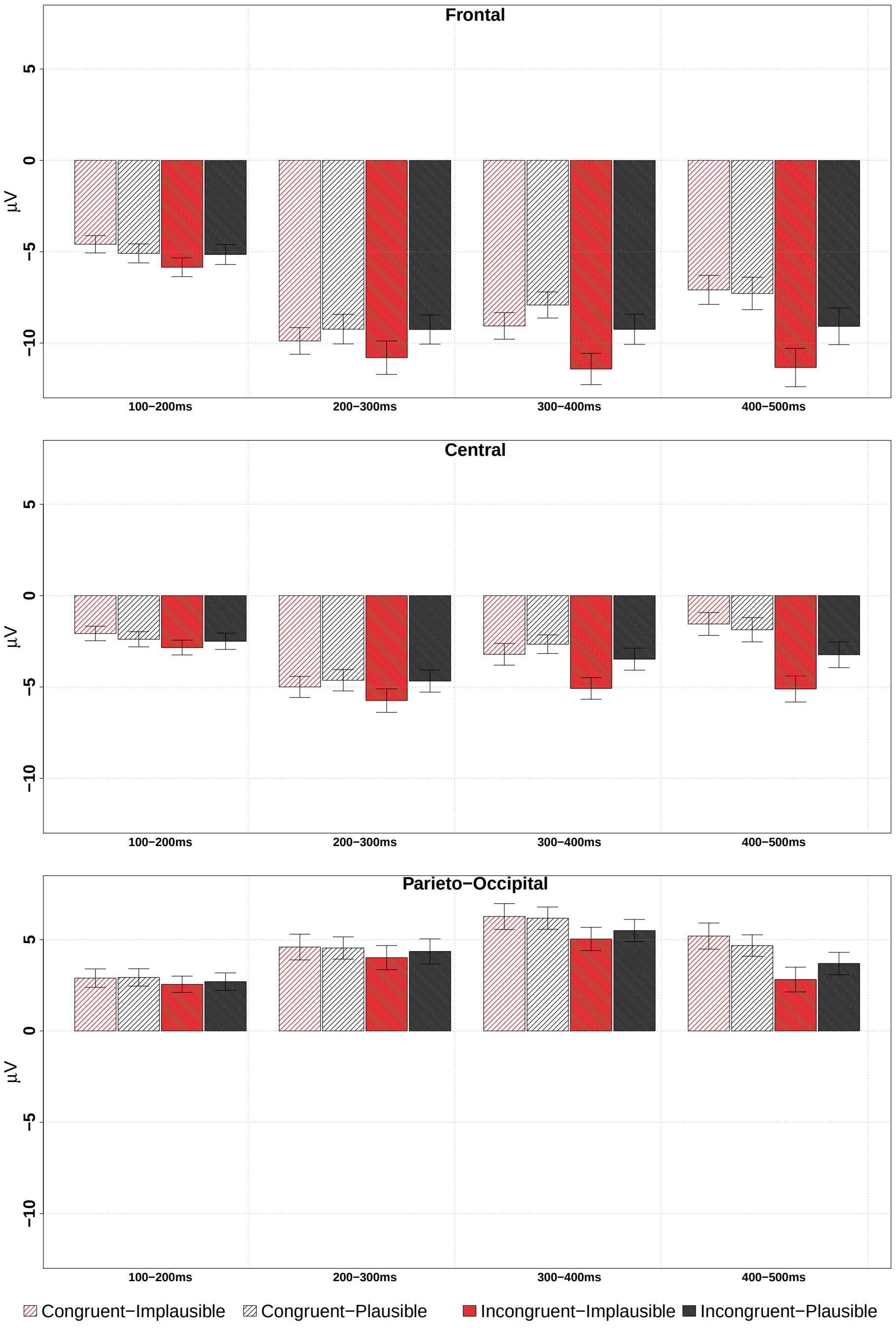

Figure 6: Bar-plot of ERP responses (mean and standard error) for the different latencies examined (100-200 ms, 200-300 ms, 300-400 ms, 400-500 ms) with Congruency represented using line density (Congruent - sparse; Incongruent - dense), and Plausibility represented using colors (Plausible - black, Implausible - red), divided in panels by head regions (Frontal, Central, Parieto-Occipital) from top to bottom.

and Plausibility $[F(1,18) = 6.1, p = .024]$ was also significant, as well as the four-way interaction Congruency by Plausibility by Region by Laterality $[F(3.43,61.77) = 2.9, p = .037]$. Post-hoc analyses revealed that for congruently matching stimuli, the effect of plausibility was restricted to the left parieto-occipital sites ($p = .022$); whereas, incongruent stimuli with an implausible content elicited a larger negativity than stimuli with a plausible content (i.e, the effect was widespread; $p < .005$ for all contrasts).

4. General Discussion

Our cognitive system heavily relies on expectancy mechanisms to facilitate the processing of incoming information and forward appropriate responses (Rao and Ballard, 1999; Bar, 2007; Friston, 2010; Wacongne et al., 2012; Clark, 2013; Koster-Hale and Saxe, 2013; Pickering and Clark, 2014; Lupyan and Clark, 2015; Kuperberg, 2016).

Research using electro-physiology has shown that stimulus processing is mediated by, at least, two types of expectancy: (1) the likelihood of a precise stimulus within its local context (e.g., an implausible word in a sentence, or an odd object in a scene; Kutas and Hillyard 1980; Hagoort et al. 2004; Ganis and Kutas 2003; Mudrik et al. 2010); and (2) the global consistency of consecutive stimuli (e.g., a narrative of scenes with an incongruent ending; West and Holcomb 2002; Sitnikova et al. 2008; Cohn et al. 2012).

By using a cross-modal (sentence-scene) verification paradigm, the current study examined how the simultaneous violation of both local (stimulus *plausibility*) and global (contextual *congruency*) expectations would impact processing costs. Our main result was that, indeed, the simultaneous violations of both expectancies resulted into the largest processing costs. However, we also observed a degree of independence in the neural mechanisms employed to resolve violations of plausibility and congruency. In fact, violations of plausibility gave rise to temporally and spatially localised processing costs, while costs were longer-lasting and more widespread when congruency was violated.

In particular, incongruent pairs of stimuli elicited a larger negative shift of electro-potential activity than congruent pairs, as early as 100 ms after the scene onset. This result replicated previous work using cross-modal verification tasks (e.g., Teder-Sälejärvi et al. 2005; Dikker and Pylkkanen 2011; Brunellière et al. 2013). Critically, however, such an effect was particularly prominent for incongruent pairs that had implausible content. This result suggests that the content of the stimuli is made available by the cognitive system as soon as the congruency operation begins taking place.

A main effect of plausibility was observed only at 200 ms and 300 ms, as indicated by the larger negativity for implausible than plausible stimuli. Also this result replicated previous literature on the topic (e.g., Mudrik et al. 2010, 2014). Congruency alone, instead, did not have a significant effect at 200 ms, but it started exerting its effect again at 300 ms and 400 ms, with a larger negativity for incongruent than congruent stimuli. This result suggests that at 200 ms the semantics of the scene start being evaluated. Then, at 300 ms, while scene plausibility is resolved, processing resources are allocated again to the congruency violation. At this point in time separate sources are probably recruited, with plausibility mainly restricted to fronto-central area, while congruency being more widespread.

Our study differs from Mudrik et al. 2010, 2014, where the same set of visual scenes was originally used, in that content congruency had longer-lasting and wider-spread effects than stimulus plausibility (compare Figure 4 in this study with Figure 4 of Mudrik et al. 2014). We tentatively suggest that the nature of our verification task might have required a different allocation of cognitive resources, whereby more prominence was given to violations occurring on the congruency of the stimuli, which is the core goal of the task, rather than plausibility, which mainly relates to their local content.

Finally at 400 ms, in contrast to previous literature, we find stimulus plausibility to be significant only in interaction with congruency: incongruent and implausible stimuli are more costly to process than their congruent counterparts. This result suggests that the mechanisms of semantic integration indexed by the N400 component are sensitive to both global and local effects of expectancy in a modality independent manner. The N400 may therefore be a proxy for general violations of former expectations. In fact, as shown by Dyck and Brodeur 2015, when the semantics of a target object does not violate the context of the scene, but has an ambiguous identity, N400 effects are observed only when the the context of the scene does not help the disambiguation of the object (i.e., it is neutral). So, it is not the semantics of the stimulus, but its overall expectancy value that might be indexed by the N400 component.

On a similar line of arguments, we believe that our results provide additional evidence in support of contextual matching models (e.g. Bar and Ullman 1996; Bar 2004; De Cesarei and Loftus 2011; Mudrik et al. 2014; Trapp and Bar 2015). Such models assume that an experience-based schematic prediction of a scene (and its context) is generated from its low spatial-frequency, and top-down control exerted upon this pre-activated schema to filter out irrelevant information. Thus, object identification becomes compromised when the object is incongruent with pre-activated contextual information (e.g., a brick in the mouth of a boy). Our results add to this claim that context matching might be directly mediated by other information sources, such as language in our cross-modal verification task. By reading a sentence, in fact, the participant is pre-activating a conceptual schema (or message) of a possible context, which gets perceptually ’filled’ by a subsequent scene. In this scenario, object identification becomes even harder when the object violates its embedding scene context and it is incongruent with the conceptual schema activated by the linguistic information (see Figure 6 to visualize the significant interactions found between congruency and plausibility at 100 ms and 400 ms, which supports this suggestion).

Our findings also provide insights about the role played by local and global expectancy mechanisms in cognitive architectures centered around predictive principles of error-correction (e.g., Bar 2007; Friston 2010). In particular, our data suggest a hierarchical pipeline in which first-order connections, coding for the coherence of stimulus representation (congruency), are followed by second-order modulatory connections that instead relate to the precision of their content (plausibility). Thus, when errors occur on first-order connections (mis-match between stimuli), processes of correction are longer-lasting than when errors occur on second-order connections (implausible content), which are instead more locally resolved. The reader is referred to a recent paper by (Kanai et al., 2015) inspiring these ideas.

We can also attempt to formulate a pipeline about the interplay of global and local violations of expectancy, which goes from stimulus processing (up-stream) to the behavioural verification response (down-stream). We do so by comparing the results of the current study with another of our studies, which used the same cross-modal verification task and stimuli, but looked at the moment-by-moment, arm-reaching response, observed when the verification is acted out (Coco and Duran, in press). Upon stimulus processing, the violations of both expectancy mechanisms triggered the largest processing costs on the neural responses. However, in Coco and Duran (in press) we observed that correct verifications of implausible but congruently matching stimuli were associated with more complex behavioral responses than incongruent pairs. This result corroborates with the long reaction time that we found in both studies for this condition (refer to section 3.1 of this study). We argue that by accepting as congruent (and therefore as true) implausible information (and therefore implicitly false), our prior knowledge about event information is violated, i.e., boys do not eat bricks.

Our study opens several important questions to be addressed by future research. In particular, it is important to assess whether a richer linguistic context, i.e., more than a single sentence, as well as a different order of modality presentation (i.e., scene-first), would change the dynamics of expectancy violations associated with stimulus plausibility and contextual congruency. As suggested above, the nature of task might have given more prominence to violations of congruency rather than stimulus plausibility. Moreover, congruency led always to a ’Yes’ response, while incongruency to a ’No’ response. Thus, the neural differences attributed to congruency processing may have other nature, such as response preparation. Thus, by using a different task, it might be possible to assess whether congruency is still implicitly computed, and if so, how would it interact with stimulus plausibility. Another possible avenue would be to examine the effect of congruency across different dimensions. Here, the plausibility between stimuli was kept invariant (e.g., a plausible sentence was always paired with a plausible scene), and congruency was manipulated as a content mis-match (e.g., eating a hamburger vs a fish). This implied that participants already expected an implausible scene after reading an implausible sentence. Hence, plausibility violations, as well as interactions between plausibility and congruency might have triggered less pronounced effects. Thus, it would be of theoretical interest to investigate other cases of incongruency by crossing, for example, stimuli with a different plausibility value (e.g., a plausible sentence with an implausible scene); and compare it to the case investigated in this study (i.e., stimuli share a similar plausibility).

In summary, our study provides novel insights on expectancy mechanisms during stimulus processing in a cross-modal verification task by disentangling the differential role of plausibility and congruency on processing costs. The evidences provided here suggest that interdependent mechanisms are utilized by the cognitive system to integrate the semantic content of cross-modal stimuli and establish their mutual congruency.

5. Acknowledgment

We thank Liad Mudrik for having shared the stimuli with us. We also would like to thank Sergio Della Sala, Rita Jeronimo, George L. Malcolm, Mante Nieuwland, Ana L. Raposo, and two anonymous reviewers for their important feedback on previous versions of this manuscript. We also acknowledge Frederico J. Marques, who sustained MIC in the early stages of this project. Rest in peace. MIC was supported by Fundação para a Ciência e Tecnologia (SFRH/BPD/88374/2012) and Leverhulme Trust (ECF-2014-205). SA and KMP were supported by Max Planck Institute for Psycholinguistics, Donders Institute for Brain, Cognition and Behaviour and Fundação para a Ciência e Tecnologia (PTDC/ PSI-PCO/110734/2009; UID/BIM/04773/2013 CBMR 1334; PEst-OE/EQB/LA0023/2013; and UID/PSI/00050/2013).

Bar, M., 2004. Visual objects in context. Nature Reviews Neuroscience 5 (8), 617–629.

Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. Trends in cognitive sciences 11 (7), 280–289.

Bar, M., Ullman, S., 1996. Spatial context in recognition. Perception 25 (3), 343–352.

Biederman, I., Glass, A. L., Stacy, E. W., 1973. Searching for objects in real-world scenes. Journal of experimental psychology 97 (1), 22.

Boyce, S. J., Pollatsek, A., 1992. Identification of objects in scenes: the role of scene background in object naming. Journal of Experimental Psychology: Learning, Memory, and Cognition 18 (3), 531.

Brunellière, A., Sánchez-García, C., Ikumi, N., Soto-Faraco, S., 2013. Visual information constrains early and late stages of spoken-word recognition in sentence context. International Journal of Psychophysiology 89 (1), 136–147.

Camblin, C. C., Gordon, P. C., Swaab, T. Y., 2007. The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. Journal of Memory and Language 56 (1), 103–128.

Carpenter, P., Just, M., 1975. Sentence comprehension: a psycholinguistic processing model of verification. Psychological Review 82 (1), 45.

Clark, A., 2013. Whatever next? predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences 36 (03), 181–204.

Clark, H., Chase, W., 1972. On the process of comparing sentences against pictures. Cognitive Psychology 3 (3), 472–517.

Coco, M. I., Duran, N. D., in press. When expectancies collide: Action dynamics reveal the interaction between stimulus plausibility and congruency. Psychonomic Bulletin & Review.

Coco, M. I., Malcolm, G. L., Keller, F., 2014. The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. The Quarterly Journal of Experimental Psychology 67 (6), 1096–1120.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., Kuperberg, G. R., 2012. (Pea) nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. Cognitive Psychology 65 (1), 1–38.

Davenport, J., Potter, M., 2004. Scene consistency in object and background perception. Psychological Science 15, 559–564.

De Cesarei, A., Loftus, G. R., 2011. Global and local vision in natural scene identification. Psychonomic bulletin & review 18 (5), 840–847.

De Graef, P., Christiaens, D., D'Ydewalle, G., 1990. Perceptual effects of scene context on object identification. Psychological Research 52, 317–329.

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.

Dikker, S., Pylkkanen, L., 2011. Before the n400: Effects of lexical–semantic violations in visual cortex. Brain and Language 118 (1), 23–28.

Dyck, M., Brodeur, M. B., 2015. Erp evidence for the influence of scene context on the recognition of ambiguous and unambiguous objects. Neuropsychologia 72, 43–51.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11 (2), 127–138.

Ganis, G., Kutas, M., 2003. An electrophysiological study of scene effects on object identification. Cognitive Brain Research 16 (2), 123–144.

Garcia-Diaz, A., Fdez-Vidal, X. R., Pardo, X. M., Dosil, R., 2012. Saliency from hierarchical adaptation through decorrelation and variance normalization. Image and Vision Computing 30 (1), 51–64.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K. M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304 (5669), 438–441.

Hagoort, P., van Berkum, J., 2007. Beyond the sentence given. Philosophical Transactions of the Royal Society B: Biological Sciences 362 (1481), 801–811.

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.

Henderson, J., Weeks, P., Hollingworth, A., 1999. The effects of semantic consistency on eye movements during complex scene viewing. Journal of Experimental Psychology: Human Perception and Performance 25, 210–228.

Kanai, R., Komura, Y., Shipp, S., Friston, K., 2015. Cerebral hierarchies: predictive processing, precision and the pulvinar. Philosophical

Transactions of the Royal Society of London B: Biological Sciences 370 (1668), 20140169.

Knoeferle, P., Urbach, T., Kutas, M., 2011. Comprehending how visual context influences incremental sentence processing: Insights from ERPs and picture-sentence verification. Psychophysiology 48 (4), 495–506.

Koster-Hale, J., Saxe, R., 2013. Theory of mind: a neural prediction problem. Neuron 79 (5), 836–848.

Kuperberg, G. R., 2016. Separate streams or probabilistic inference? what the n400 can tell us about the comprehension of events. Language, Cognition and Neuroscience, 1–15.

Kutas, M., 1993. In the company of other words: Electrophysiological evidence for single-word and sentence context effects. Language and Cognitive Processes 8 (4), 533–572.

Kutas, M., Federmeier, K., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review of Psychology 62, 621–647.

Kutas, M., Hillyard, S. A., 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207 (4427), 203–205.

Kutas, M., van Petten, C., Kluender, R., 2006. Psycholinguistics electrified ii (1994-2005). ma gernsbacher & m. traxler (eds.), handbook of psycholinguistics.

Lau, E. F., Holcomb, P. J., Kuperberg, G. R., 2013. Dissociating N400 effects of prediction from association in single-word contexts. Journal of Cognitive Neuroscience 25 (3), 484–502.

Lau, E. F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics:(de) constructing the n400. Nature Reviews Neuroscience 9 (12), 920–933.

Loftus, G., Mackworth, N., 1978. Cognitive determinants of fixation location during picture viewing. Journal of Experimental Psychology: Human Perception and Performance 4, 565–572.

Lupyan, G., Clark, A., 2015. Words and the world predictive coding and the language-perception-cognition interface. Current Directions in Psychological Science 24 (4), 279–284.

Maris, E., 2004. Randomization tests for erp topographies and whole spatiotemporal data matrices. Psychophysiology 41 (1), 142–151.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of eeg-and meg-data. Journal of neuroscience methods 164 (1), 177–190.

Marslen-Wilson, W., Tyler, L. K., 1980. The temporal structure of spoken language understanding. Cognition 8 (1), 1–71.

Menenti, L., Petersson, K. M., Scheeringa, R., Hagoort, P., 2009. When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. Journal of Cognitive Neuroscience 21 (12), 2358–2368.

Mudrik, L., Lamy, D., Deouell, L., 2010. ERP evidence for context congruity effects during simultaneous object - scene processing. Neuropsychologia 48, 507–517.

Mudrik, L., Shalgi, S., Lamy, D., Deouell, L. Y., 2014. Synchronous contextual irregularities affect early scene processing: Replication and extension. Neuropsychologia 56, 447–458.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2010. Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. Computational intelligence and neuroscience 2011.

Pickering, M. J., Clark, A., 2014. Getting ahead: forward models and their place in cognitive architecture. Trends in Cognitive Sciences.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M., 2000. The time-course of intermodal binding between seeing and hearing affective information. Neuroreport 11 (6), 1329–1333.

Rao, R. P., 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.

Rayner, K., 2009. Eye movements and attention in reading, scene perception, and visual search. The quarterly journal of experimental psychology 62 (8), 1457–1506.

Rayner, K., Duffy, S. A., 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition 14 (3), 191–201.

Rosenholtz, R., Li, Y., Nakano, L., 2007. Measuring visual clutter. Journal of vision 7 (2), 17–17.

Schutter, D. J., Leitner, C., Kenemans, J. L., Van Honk, J., 2006. Electrophysiological correlates of cortico-subcortical interaction: A cross-frequency spectral eeg analysis. Clinical Neurophysiology 117 (2), 381–387.

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. Journal of Cognitive Neuroscience 20 (11), 2037–2057.

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., de Almeida, J. J., Comesaña, M., Perea, M., 2015. On the advantages of word frequency and contextual diversity measures extracted from subtitles: The case of portuguese. The Quarterly Journal of Experimental Psychology 68 (4), 680–696.

Sun, H.-M., Simon-Dack, S. L., Gordon, R. D., Teder, W. A., 2011. Contextual influences on rapid object categorization in natural scenes. Brain research 1398, 40–54.

Teder-Sälejärvi, W., Russo, F., McDonald, J., Hillyard, S., 2005. Effects of spatial congruity on audio-visual multimodal integration. Cognitive Neuroscience, Journal of 17 (9), 1396–1409.

Trapp, S., Bar, M., 2015. Prediction, context, and competition in visual recognition. Annals of the New York Academy of Sciences 1339 (1), 190–198.

Van Berkum, J., Hagoort, P., Brown, C., 1999. Semantic integration in sentences and discourse: Evidence from the n400. Journal of cognitive neuroscience 11 (6), 657–671.

Van Petten, C., Luka, B. J., 2012. Prediction during language comprehension: Benefits, costs, and erp components. International Journal of Psychophysiology 83 (2), 176–190.

Võ, M.-H., Wolfe, J., 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. Psychological science 24 (9), 1816–1823.

Wacongne, C., Changeux, J.-P., Dehaene, S., 2012. A neuronal model of predictive coding accounting for the mismatch negativity. The Journal of Neuroscience 32 (11), 3665–3678.

Walther, D., Koch, C., 2006. Modeling attention to salient proto-objects. Neural networks 19 (9), 1395–1407.

West, W. C., Holcomb, P. J., 2002. Event-related potentials during discourse-level semantic integration of complex pictures. Cognitive Brain Research 13 (3), 363–375.

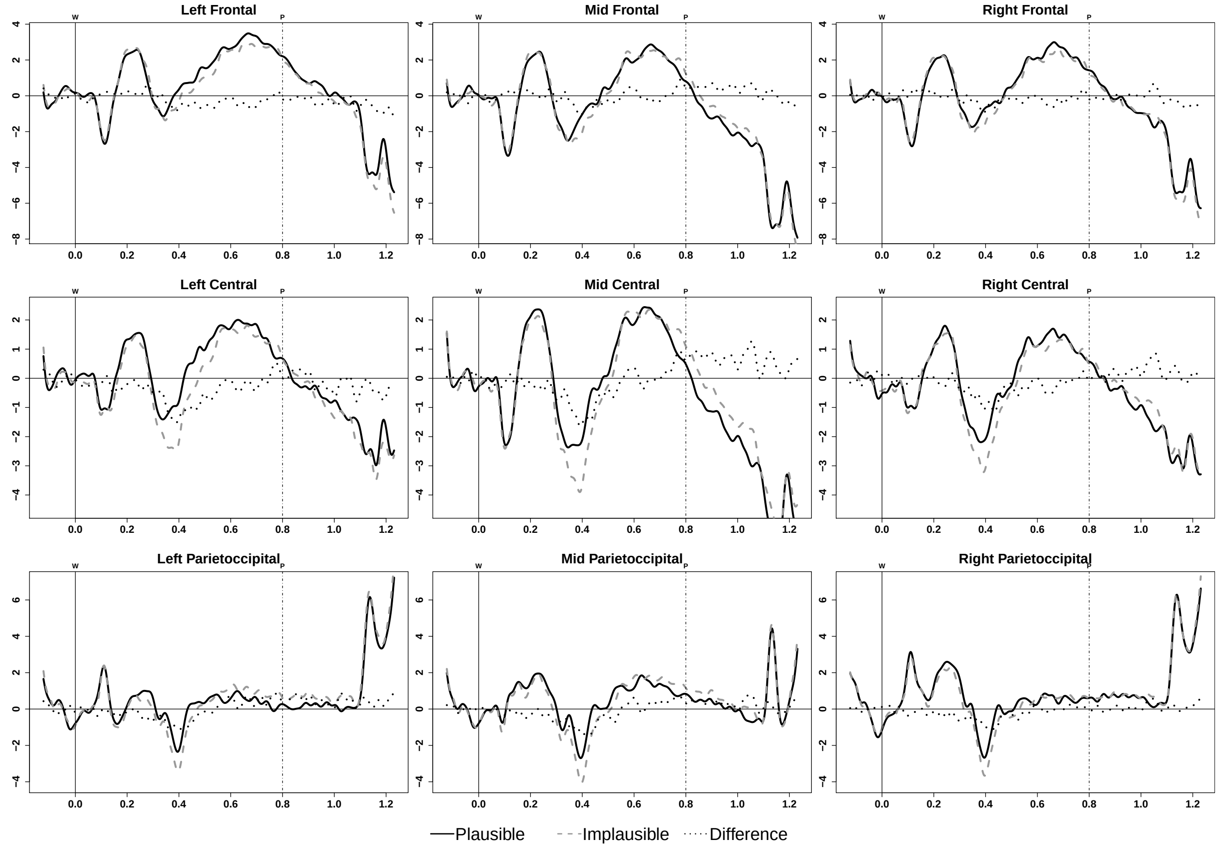

Figure A.7: ERP responses time-locked to the target word onset. We represent the Plausible condition with a tick line and the Implausible condition with a dashed-line, and plot the Difference between the two ERP waves using a dotted-line. We mark in the plot the onset of the (W)ord and the (P)icture with vertical dashed-lines.

Appendix A. ERP grand-mean analysis time-locked to the onset of the last (critical) word

We analyze EEG responses time-locked to onset of the last word in the sentence on which plausibility is manipulated. We focus on the following windows of interest: (a) 100-200, (b) 300-400, (c) 400-500, (d) 700-800. In Figure A.7, and visualize the ERP mean-amplitude contrasting the two conditions of Plausibility (Plausible, Implausible).

100-200 (ms):. No significant main effect of Plausibility, nor two-way interactions of Plausibility with Laterality, F ¡ 1, Region $[F(1.3,23.39) = 3.25, p = 0.07$ or three-way interactions were found (Plausibility x Region x Laterality) $[F(4,72) = 1.71, p = .16]$.

300-400 (ms):. Significant main effect of plausibility $[F(1,18) = 6.19, p = .023]$, with implausible words eliciting a larger negative shift than plausible words. However, we did not observe any interaction of Plausibility with

Laterality: F ¡1, Region, $[F(1.32, 23.68) = 2.81, p = .1]$ nor their three way interaction (Plausibility x Region x Laterality) was significant, $[F(2.46, 44.28) = 1.63, p = .2]$.

400-500 (ms):. We found a significant main effect of plausibility $[F(1,18) = 5.14, p = .036]$ going in the same direction of the previous window, i.e., implausible words elicited a larger negativity than plausible words. But again, no significant interaction of Plausibility with Laterality, $[F(2,36) = 1.17, p = .32]$, Region, $F(1.42, 25.57) = 1.98, p = .17]$, nor the three way interaction Plausibility x Region x Laterality was significant, F ¡ 1.

700-800 (ms):. No significant main effect of Plausibility $[F(1,18) = .41, p = .5]$. We only observed a significant two-ways interaction of Plausibility with Laterality $F(2,36) = .4, p = .04]$. Post-hoc analysis showed that implausible stimuli triggered more positive ERPs than plausible stimuli, but this difference was restricted to midline sites (p ¡. 005); no differences between plausible and implausible sentences were observed on right and left sites.

Brief Discussion:. We replicated the classic effect of implausible words eliciting stronger negative shifts than plausible words (see Kutas and Hillyard 1980 for seminal work). Differently from previous results, however, the effect of plausibility kicked in earlier, i.e., already at 300ms. This might be a consequence of the verification task employed that required participants to engage into a deeper semantic processing of the target word to accurately respond to the congruency judgment. Importantly, the effect of plausibility was exhausted by the end of the window. If anything, we observed implausible words triggering more positivity than plausible words, only at the mid-line sites. This result reassures that the effect of plausibility observed on ERPs time-locked to the scene onset (reported in the main text) is genuinely related to the visual information of the scene (and its congruency value with respect to the content of the sentence); and, it is certainly not influenced by the plausibility of the word preceding it.

Appendix B. Cluster-based non-parametric permutation tests time-locked to scene onset

We present results largely corroborating with the ANOVA analyses reported in the main text, but using a non-parametric cluster-based permutation approach (Maris, 2004). Conceptually, for every sample (channel-time), a t-statistics is computed between experimental conditions. An electrode channel is considered significant in a given temporal window, if there are at least 2 neighboring channels simultaneously significant at $\alpha < 0.025$. Once the cluster is formed, its significance is evaluated (sum of t-values within the cluster) over 1000 random permutations of the data, where experimental conditions of interest (e.g., the two levels of Plausibility) are compared (Maris and Oostenveld, 2007). This test controls for the false positive rate usually associated with multiple comparisons, and it provides a rough idea about how the effects distribute over the head-scalp.

We visualize the distribution of brain activity over the head-scalp using heat-maps from stimulus onset till 800 ms after, in temporal windows of 50ms each; and mark with asterisks the electrodes of the cluster which were

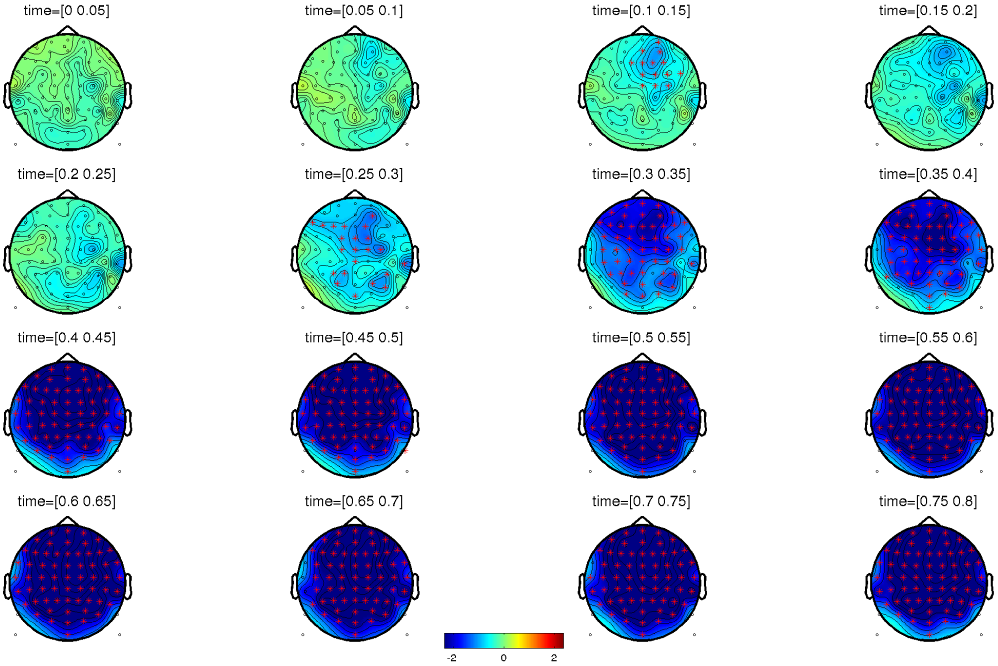

Figure B.8: Heat-map for scalp distribution over time (from stimulus onset till 800 ms after it in slices of 50 ms each) for the contrast (**Incongruent-Congruent**). We mark with asterisks the electrodes where the two conditions significantly differed.

significant at $\alpha < 0.025$. We focus on three types of contrast: (1) Congruent versus Incongruent, (2) Plausible versus Implausible, and (3) the interaction between Congruency and Plausibility.

In Figure B.8, we show how the scalp-distribution between congruent and incongruent verifications differ over time. The results show that, as early as 100 to 150 ms, incongruent trials elicited a significantly stronger negativity than congruent trials, hence confirming the results obtained with the ANOVAs on ERP mean amplitude. This effect is localized in the fronto-central regions (e.g., at electrodes Fz, FCz, F2, F4). The difference between incongruent and congruent trials starts again between 250-300 ms mostly in the central region, then such difference becomes progressively stronger, lasting until the end of the temporal region of interest.

In Figure B.9, we find a significant difference between plausible and implausible scenes at 200-250ms, which has a mainly a fronto-central distribution. This effect is perfectly in line with previous literature (e.g., Mudrik et al. 2014), with a maximum peak amplitude occurring slightly later than the effect of congruency.

In Figure B.10, we examined the interaction between Congruency and Plausibility, and found significant differences at 400-450ms and 500-550ms. This interaction is explained by the fact that incongruent and implausible stimuli elicit a stronger negativity than congruent and implausible stimuli.

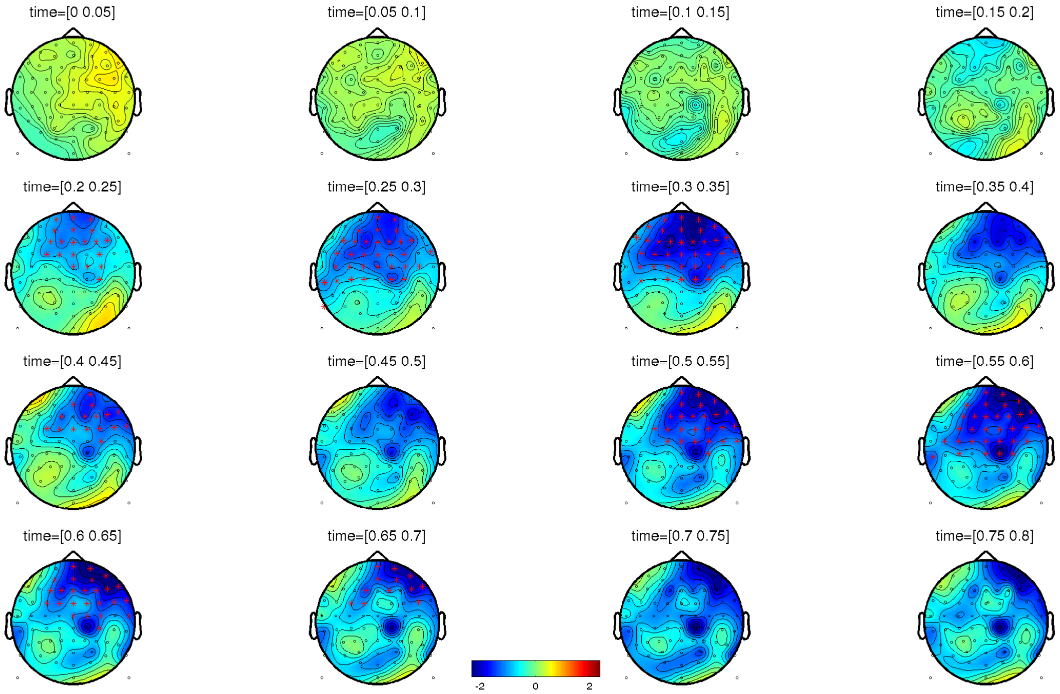

Figure B.9: Heat-map for scalp distribution over time (from stimulus onset till 800ms after it) for the contrast (**Implausible-Plausible**). We mark with asterisks the electrodes where the two conditions significantly differed.

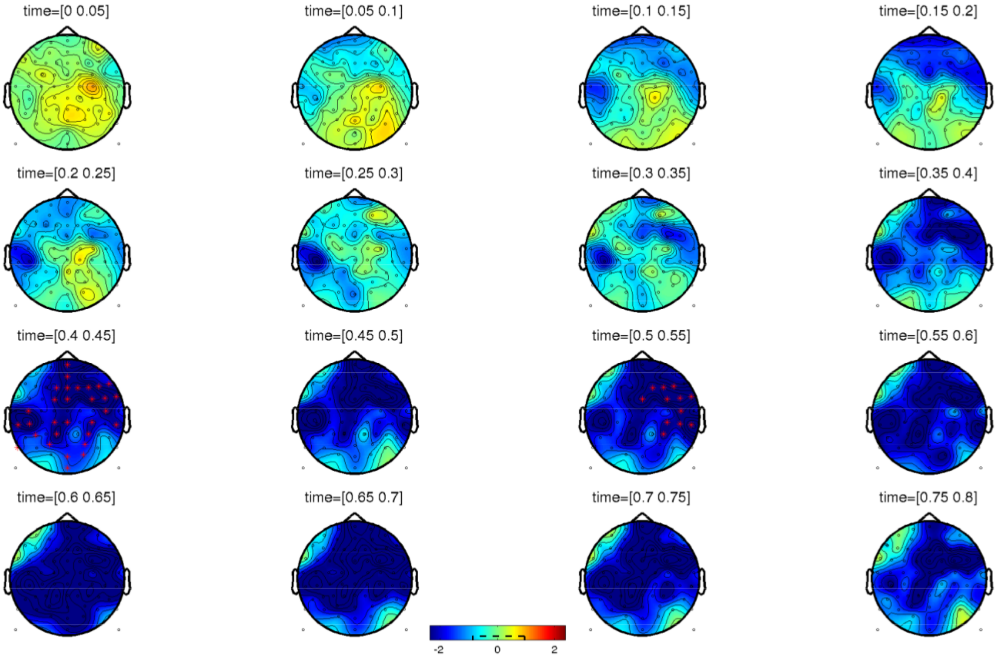

Figure B.10: Heat-map for scalp distribution over time (from stimulus onset till 800ms after it) for the interaction between Congruency and Plausibility. The interaction contrast is constructed by comparing differences between levels across conditions, i.e., the difference between Plausible and Implausible in the Congruent condition with the same difference but in the Incongruent condition. We mark with asterisks the electrodes where the interaction term was significantly different.