



Published in final edited form as:

Psychol Sci. 2013 September ; 24(9): 1816–1823. doi:10.1177/0956797613476955.

Differential ERP Signatures Elicited by Semantic and Syntactic Processing in Scenes

Melissa L.-H. Võ and Jeremy M. Wolfe

Harvard Medical School, Brigham and Women's Hospital

Abstract

In sentence processing, semantic and syntactic violations elicit differential brain responses in ERP recordings: An N400 signals semantic violations, while a P600 marks inconsistent syntactic structure. Does the brain register similar distinctions in scene perception? Participants viewed “semantic inconsistencies” created by presenting objects that were incongruent with a scene's meaning and “syntactic inconsistencies” in which an object violated structural rules. We found a clear dissociation between semantic and syntactic processing: Semantic inconsistencies produced negative deflections in the N300/N400 time window, while syntactic inconsistencies elicited a late positivity resembling the P600 found for syntax manipulations in sentence processing. Interestingly, extreme syntax violations such as a floating toast, showed an initial increase in attentional deployment, but failed to produce a P600 effect. We therefore conclude that different neural populations are active during semantic and syntactic processing in scenes and that impossible object placements may be processed categorically different from syntactically inconsistent placements.

Keywords

Semantics; Syntax; Scene Grammar; Scene Perception; ERPs; Visual Perception; Evoked Potentials; Object Recognition; Perception

Introduction

Imagine a world entirely consisting of randomly arranged objects. This would be disconcerting because we have learned over a lifetime that our world is a highly structured, rule-governed place. Objects in scenes, like words in sentences, seem constrained by a “grammar” that we implicitly understand and that allows us to efficiently process scenes. Biederman, Mezzanotte, and Rabinowitz (1982) first applied the terms *semantics* and *syntax* to objects in scenes that underwent different relational violations. According to Biederman's taxonomy, probability, position, and size of objects in scenes are *semantic* relations, since

Corresponding Author: Melissa Le-Hoa Võ, Visual Attention Lab, Harvard Medical School, BWH, 64 Sidney Street, Suite 170, Cambridge, MA 02139, USA, 001-617-768-8812, mlvo@search.bwh.harvard.edu, <http://www.lehoa.macmate.me/MelissaVo/Welcome.html>.

Author Contributions:

M.L.V. developed the study concept. Testing, data collection, data analysis, and interpretation were performed by M.L.V. M.L.V. drafted the paper, and J.M.W. provided critical revisions. All authors approved the final version of the paper for submission.

they require access to object meaning. A green lawn on your office floor instead of a carpet would be a violation of scene semantics. Your laptop, floating above your desk, would be a *syntactic* violation because it defies physical laws of gravity. Biederman et al. found that both semantic and syntactic violations resulted in slower, less accurate object detection (see also Võ & Henderson, 2009).

In this paper, we use the syntax/semantics distinction slightly differently. Here, *syntactic* processing refers to the relationship of objects to the structure of a scene, while *semantic* processing refers to the more global relationship of objects to the scene meaning. Both our definitions and Biederman's are efforts to specify a "grammar" of scenes, but is this semantic/syntactic distinction a merely metaphorical borrowing of terms from linguistics (Henderson & Ferreira, 2004) or does the brain actually code object-scene relationships in this manner? Here, we present evidence that distinct event-related brain potentials (ERPs) are associated with semantic and syntactic processing in scenes and that these potentials resemble the potentials associated with semantics and syntax in language processing.

Ganis and Kutas (2003) used ERPs to investigate the nature and time course of *semantic* context effects on object identification. After a 300ms preview of a scene (e.g. soccer players) either a semantically consistent (soccer ball) or inconsistent (toilet paper) object appeared. They reported a "scene congruity effect" in the N400 time window for incongruous relative to congruous objects, which closely resembled the N400 effect found for violations of semantic expectations using verbal (e.g., Holcomb, 1993; Kutas & Hillyard, 1980) or pictorial information (e.g., Barrett & Rugg, 1990; McPherson & Holcomb, 1999; for a review see Kutas & Federmeier, 2011). While Ganis and Kutas (2003) set up expectations by presenting scene previews and had participants attend to the semantic congruity of objects, Mudrik, Lamy, and Deouell (2010) found that even when the scene and object were presented simultaneously and when the instructions did not direct attention to semantic congruity, a pronounced, slightly more anterior N300/N400 effect emerged. Thus, task-relevance and pre-activation of expectations are not prerequisites for detecting semantic anomalies in scenes. The N400 effect has been observed across many types of stimuli (linguistic, pictures, objects, actions, and sounds), implying that the semantic processing of very different input types might be based on one common mechanism.

In the language domain, the P600 component has been identified as a marker for syntactic problems that prompt reanalysis of the sentence (e.g., Friederici, Pfeifer, & Hahne, 1993; Osterhout, & Holcomb, 1992). If syntactic violations in scenes are similarly processed, then those violations should elicit similar, late positive brain responses. Previous attempts to find ERP components specific to structural processing in scenes have been mixed. Cohn, Paczynski, Jackendoff, Holcomb, and Kuperberg (2012) compared different types of comic strips involving violations regarding the meaning or sequencing of images. They replicated the N300/N400 effects for semantic violations. While they did not observe P600 responses to structural manipulations, they reported a larger left-lateralized anterior negativity (LAN), which in studies of language has been associated with syntax violations (e.g., Friederici, 2002). Recently, Demiral, Malcolm, and Henderson (2012) manipulated the spatial congruency of objects in scenes while keeping semantic congruency constant. With a 300ms scene preview, spatially incongruent objects elicited an early N300/N400 component, which

was weakened when object and scene were presented simultaneously. However, despite structural manipulations, no P600 or LAN response was observed.

In this study, we directly compared brain responses to semantic and syntactic violations in images of real-world scenes to test whether these forms of scene processing show dissociations of neural mechanisms similar to the language domain. In addition to semantically incongruent objects, we created two types of syntax violations: mild syntax violations based on misplacing objects within scenes and extreme violations, caused by having objects implausibly balance or float in mid air. As we will show, semantic violations were marked by N300/N400 responses, while syntactic inconsistencies elicited a late positivity resembling the P600 found for syntactic processing in language. Interestingly, extreme syntax violations failed to produce a P600 effect.

Methods

Stimulus material

We created 602 colored images of real-world scenes by taking photographs of 152 different scenes in 4 different versions: 1) with a semantically consistent object in a consistent location, 2) semantically consistent object in an syntactically inconsistent location, 3) a semantically inconsistent object in syntactically consistent or 4) inconsistent locations (see Figure 1). The double inconsistency (SEMSYN) was only added to control for possible position effects. Each observer saw all 152 scenes only once throughout the experiment. Scenes were evenly divided amongst the 4 conditions of Figure 1. To better distinguish within the syntax condition, we further divided this condition into 1/3rd of the syntax scenes, in which manipulations were mild (mislocated objects), and 2/3^{rds} where the syntax condition manipulations were more extreme (floating or balancing objects; Figure 2). Fourteen extra scenes were used as fillers for a repetition detection task. Images were not created by post-hoc insertion of objects into scenes. Floating objects, for instance, were actually photographed hovering in mid air (attached to invisible strings) to ensure realistic lighting conditions and minimizing photoshop editing (see online Supplemental Material for more scene examples). In addition, the bottom-up saliency of the critical objects across conditions was assessed using the Itti and Koch (2000) MatLab Saliency Toolbox. Rank order of saliency peaks assigned to the critical object was used to ensure that consistent and inconsistent objects did not differ in mean low-level saliency, $F < 1$.

Participants

Twenty-eight subjects (16 female) participated, ranging in age between 19 and 35 years ($M=25$, $SD=5$). All were paid volunteers who gave informed consent. Each had at least 20/25 visual acuity and normal color vision.

Procedure

Participants were seated in a sound-attenuated, dimly lit room where scenes were presented on a 17-inch monitor with resolution 1,024×768 at 75Hz, viewed at a distance of about 70cm subtending visual angles of 26° (horizontal) and 20° (vertical). Participants were told that they would see a series of scenes containing one critical object marked by a pre-cue.

After a blink phase, a preview of the scene without the critical object was presented for 500ms. Then a red dot appeared at a location in the scene, which indicated where to move the eyes and where to expect the critical object to appear. To avoid eye movement artifacts, participants were instructed not to move their eyes away from the cued location and to confine blinking to the phase preceding each trial. 500ms after cue onset (plus a random jitter between 0 and 300ms to prevent anticipatory effects), the critical object appeared in the scene and remained together with the scene for 2,000ms (see Figure 3).

To keep participants engaged in viewing the scenes without signaling the object-scene inconsistencies, we asked them to view each scene carefully and press a button when a cued object reappeared in the same location in the same scene. All repeated scenes were filler scenes taken equally often from all conditions. These were excluded from subsequent analysis. Each participant saw each scene only once. Conditions were latin-square randomized across participants. At the end of the experiment, participants viewed all scenes again and gave consistency ratings for each object-scene relation (from 1: very consistent to 6: very inconsistent).

EEG Recording and Analysis

EEG was recorded from 64 scalp sites (10-20 system positioning) assigned to nine regions (see Mudrik et al., 2010), a vertical eye channel for detecting blinks, two horizontal eye channels to monitor for saccades, and two additional electrodes affixed the mastoid bone. EEG was acquired with the Active Two Biosemi system using active Ag-AgCl electrodes. All channels were referenced offline to an average of the mastoids. The EEG was recorded at 512Hz sampling rate and offline high-pass filtered at 0.1Hz (24dB/octave) to remove slow drifts. The EEG was subsequently segmented into 1000-ms long epochs averaged separately for each condition time-locked to object onset. Average waveforms were low-pass filtered with a cutoff of 30Hz and each epoch was baseline-adjusted by subtracting the mean amplitude of the pre-stimulus period (−100ms) from all the data points in the epoch. Trials with blinks, eye movements, and muscle artifact were rejected prior to averaging (12% of all epochs).

Results

Behavioral results

Repetition detection task performance—Participants' task was to inspect every scene and press a button when they found an exact repetition of a scene they had previously seen. The overall error rate averaged 3%. Erroneous responses were excluded from the ERP analyses.

Inconsistency ratings—Inconsistent objects were rated higher on the inconsistency scale than the consistent controls: consistent=1.26, syntax-mislocated=3.13, syntax-physics=4.38, semantics=4.33, all t -values > 11.0. Mislocated objects were rated as less inconsistent than syntax-physics and semantics, $t(27)=5.21$, $p<.01$ and $t(27)=7.12$, $p<.01$, respectively. Syntax physics and semantics did not differ from each other, $t<1$.

ERP results

We were interested in replicating N300/N400 incongruity effects previously found for semantically inconsistent objects in scenes. In addition, we wanted to test whether syntactically inconsistent objects would elicit a late positivity, resembling the P600 response known from sentence processing. Figure 4 shows scalp distributions of ERP difference waves (consistent subtracted from semantic, syntactic-physics, and syntactic-mislocated) for three post-object onset time windows N300:250-350ms, N400:350-600ms, and P600:600-1000ms (see Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008).

These scalp plots clearly show a dissociation between semantic and syntactic processing for mislocated objects. Figure 5 presents grand average waveforms for semantic, syntax-physics, and syntax-mislocated conditions together with the consistent control. Semantic inconsistencies triggered a negative response in the N300 and N400 time ranges. In contrast, mislocated objects elicited a late positive response that is significant in the P600 range (statistics below) and was preceded by a trend for a positive response in the N400 time window. The extreme syntactic inconsistencies show significant negativity only in the N300 range and not in the P600 range associated with other syntax violations.

To test whether the observed ERP effects differed significantly from the consistent control, we measured mean amplitudes for each time window and each condition and submitted these to paired T-tests (see Figure 6). We confined our analyses to the mid-central region (averaged across FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2), which has previously shown strong “N390 scene congruity effects” (Ganis & Kutas, 2003) as well as P600 effects in sentence processing (e.g., Osterhout & Holcomb, 1992).

N300: 250-350ms

Both semantic and syntactic-physics manipulations elicited significantly more negative responses than consistent controls, $-1.65\mu\text{V}$, $t(27)=3.67$, $p<.01$, and $1.06\mu\text{V}$, $t(27)=3.08$, $p<.01$, respectively, while mislocated objects did not show any effects, $t<1$.

N400: 350-600ms—In the N400 time window, we only observed pronounced negative responses for semantic inconsistencies, $-1.50\mu\text{V}$, $t(27)=2.70$, $p<.01$, while syntax-physics did not show effects, $t<1$, and mislocated showed a tendency for a greater positive response, $+1.33\mu\text{V}$, $t(27)=2.03$, $p=.05$.

P600: 600-1000ms—Brain responses in the P600 time window, however, showed a strong positive response to mislocated objects, $+2.33\mu\text{V}$, $t(27)=3.00$, $p<.01$, whereas neither semantic nor extreme syntax violations elicited responses significantly different from the consistent controls, $t<1$ and $t(27)=1.23$, $p=.22$, respectively.

Discussion

The aim of our study was to investigate whether semantic and syntactic processing of objects in scenes draw on qualitatively different neural mechanisms. We tested this using electrophysiological markers known to distinguish semantic and syntactic processing in the language domain. We replicate previous findings of an early N300 component, suggesting

initial difficulties in perceptual processing of inconsistent objects (e.g., Eddy, Schmid, & Holcomb, 2006; McPherson, & Holcomb, 1999). This was followed by an N400 for semantic inconsistencies, which might signal increased post identification processes based on semantic knowledge (e.g., Ganis & Kutas, 2003; Mudrik et al., 2010).

Most importantly, we show a clear dissociation between semantic and syntactic processing by finding a late positivity in response to syntactic scene violations resembling the P600 syntactic effect found in language. This late positivity was only observed for mild syntax violations caused by merely mislocated objects. More extreme syntax violations elicited no such effect. Returning to Figure 2a, we may speculate that while a pot in a kitchen is nothing unusual, the unexpected structural relationship between the object and its location within the scene may subsequently trigger scene reanalysis resulting in a P600 deflection.

Violations of physics, like the floating beer bottle of Figure 2b, may be odd enough to impede initial perceptual processing as seen in the N300 effect, but such violations may be too odd to permit resolution by the later reanalysis of the scene that yields a P600. This would follow findings with linguistic stimuli where extremely ungrammatical sentences do not elicit at P600 response (see Hopf, Bader, Meng, & Bayer, 2003). This might also explain Demiral et al.'s (2012) lack of P600 responses, while they instead observed early N300-N400 effects. In addition to repeating a limited set of scenes, mild and extreme syntax manipulations were confounded by contrasting, e.g., a plane versus a bus in the sky. Based on our data, we would argue that a floating bus constitutes such an extreme syntactic violation that reanalysis becomes ineffective.

Even more than in linguistics, the distinction between scene syntax and semantics is not always clear-cut. Biederman et al. (1982) offered a first, thought-provoking classification of object-scene inconsistencies according to which only physical constraints like gravity would be considered syntactic. Here we propose that *syntactic* processing actually goes beyond physical impossibility to include evaluation of relative object positions within a scene. The computer mouse, sitting on top of the screen, is physically legal but structurally unexpected, making it a syntactic violation by our usage. *Semantic* processing on the other hand examines object meaning relative to the semantic scene category. The mouse, in the preceding example, is in the wrong place, but in the right scene, making it semantically congruent. In our data, we found different neural signature for “plausible object in the wrong place” and “implausible object in this scene” which are similar to the signatures seen for “plausible word in the wrong place” and “implausible word in this sentence”. This moreover suggests that there might be some commonality in the mechanisms for the processing of meaning and structure encountered in a wide variety of cognitive tasks. As Chomsky (1965, 2006a,b) might put it, the general principles of language are not entirely different from the general principles of thought, including thoughts about visual scenes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by ONR N000141010278 to JMW and DFG:VO1683/1-1 to MLV. We thank John Gabrieli and Marianna Eddy at MIT for their invaluable support of this project.

References

- Barrett SE, Rugg MD. Event-related potentials and the semantic matching of pictures. *Brain and Cognition*. 1990; 14(2):201–212. [PubMed: 2285513]
- Biederman I, Mezzanotte RJ, Rabinowitz JC. Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*. 1982; 14(2):143–177. [PubMed: 7083801]
- Cohn N, Paczynski M, Jackendoff R, Holcomb PJ, Kuperberg GR. (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*. 2012; 65(1):1–38. [PubMed: 22387723]
- Chomsky, N. *Aspects of the theory of syntax*. MIT Press; Cambridge, Mass: 1965.
- Chomsky, N. From Language and Problems of Knowledge.. In: Beakley, B.; Ludlow, P., editors. *Philosophy of Mind: Classical Problems/Contemporary Issues (201-203)*. MIT Press; Cambridge: 2006a.
- Chomsky, N. On Cognitive Structures and Their Development: A Reply to Piaget.. In: Beakley, B.; Ludlow, P., editors. *Philosophy of Mind: Classical Problems/Contemporary Issues (751-755)*. MIT Press; Cambridge: 2006b.
- Demiral TB, Malcolm GL, Henderson JM. ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia*. 2012; 50(7):1271–1285. [PubMed: 22391475]
- Eddy M, Schmid A, Holcomb PJ. Masked repetition priming and event-related brain potentials: A new approach for tracking the time course of object perception. *Psychophysiology*. 2006; 43(6):564–568. [PubMed: 17076812]
- Friederici AD. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*. 2002; 6(2):78–84. [PubMed: 15866191]
- Friederici AD, Pfeifer E, Hahne A. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*. 1993; 1:183–192. [PubMed: 8257874]
- Ganis G, Kutas M. An electrophysiological study of scene effects on object identification. *Brain research Cognitive brain research*. 2003; 16(2):123–144. [PubMed: 12668221]
- Henderson, JM.; Ferreira, F. Scene perception for psycholinguists.. In: Henderson, JM.; Ferreira, F., editors. *The interface of language, vision, and action: Eye movements and the visual world*. Psychology Press; New York: 2004. p. 1-58.
- Holcomb PJ. Semantic priming and stimulus degradation—Implications for the role of the N400 in language processing. *Psychophysiology*. 1993; 30(1):47–61. [PubMed: 8416062]
- Hopf J-M, Bader M, Meng M, Bayer J. Is human sentence parsing serial or parallel? Evidence from event related brain potentials. *Brain Research. Cognitive Brain Research*. 2003; 15(2):165–177. [PubMed: 12429368]
- Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*. 2000; 40:1489–1506. [PubMed: 10788654]
- Kutas M, Federmeier KD. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*. 2011; 62(1):621–647.
- Kutas M, Hillyard SA. Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*. 1980; 11(2):99–116. [PubMed: 7272388]
- McPherson WB, Holcomb PJ. An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*. 1999; 36(1):53–65. [PubMed: 10098380]
- Mudrik L, Lamy D, Deouell LY. ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*. 2010; 48(2):507–517. [PubMed: 19837103]

- Osterhout L, Holcomb PJ. Event related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*. 1992; 31:785–806.
- Sitnikova T, Holcomb PJ, Kiyonaga KA, Kuperberg GR. Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*. 2008; 20(11):2037–2057. [PubMed: 18416681]
- Võ ML-H, Henderson JM. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*. 2009; 9:1–15. [PubMed: 19757963]

SEMANTICS

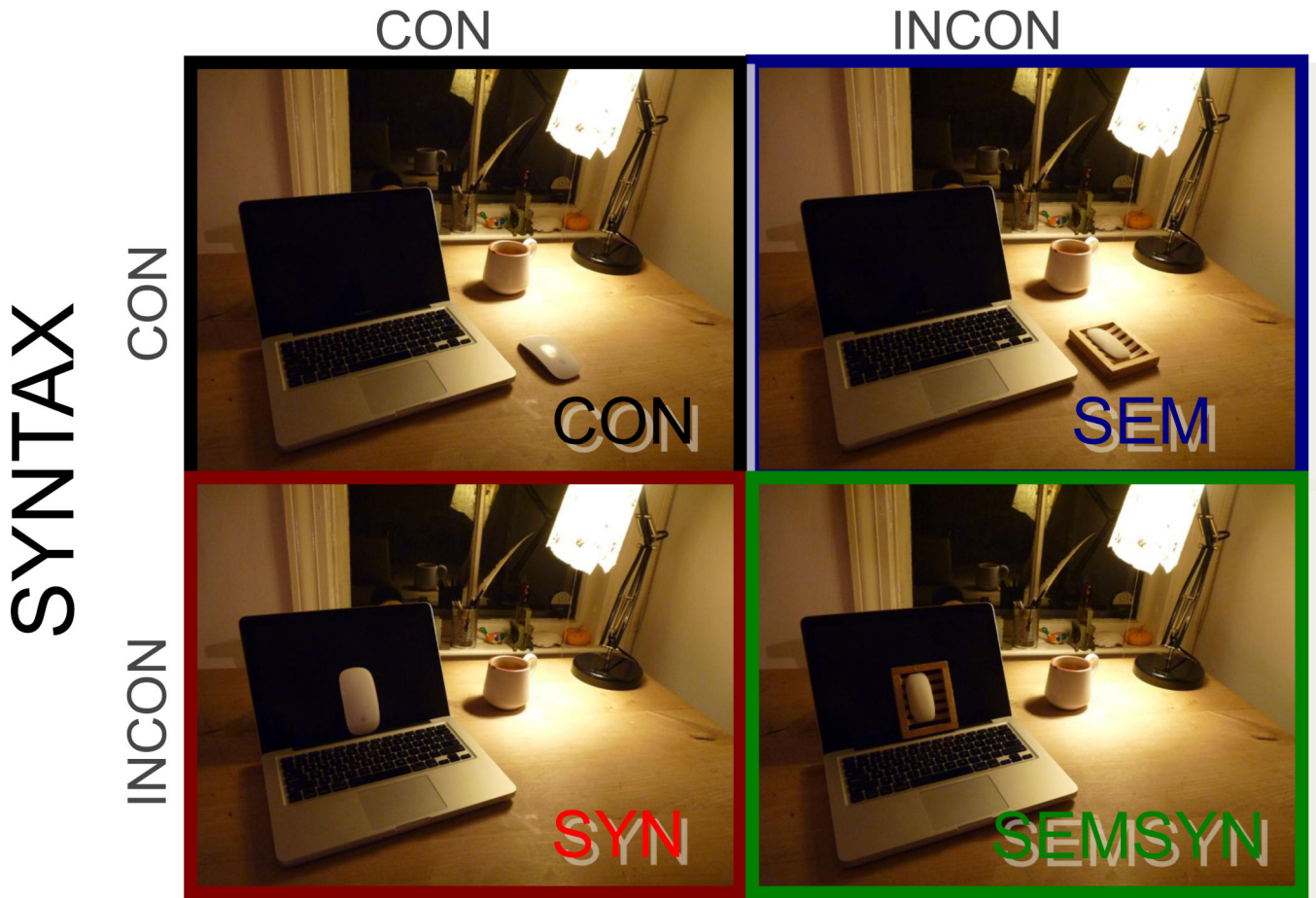


Figure 1. Four exemplar images of a scene based on semantic and syntactic manipulations. CON: Semantically consistent object in a consistent location (computer mouse next to computer), SYN: semantically consistent object in a syntactically inconsistent location (mouse on computer screen), SEM: a semantically inconsistent object (bar of soap) in syntactically consistent or SEMSYN: inconsistent locations (bar of soap on computer screen).

Mislocated



Physics



Figure 2. Example scenes showing two types of syntax violations: Mild violations were created by MISLOCATED otherwise semantically consistent objects (left panel); Violations of PHYSICS (right panel) were created by either showing objects hovering in mid air or critically balancing. In the PHYSICS cases, semantic understanding of the object is not needed in order to detect syntactic violation.

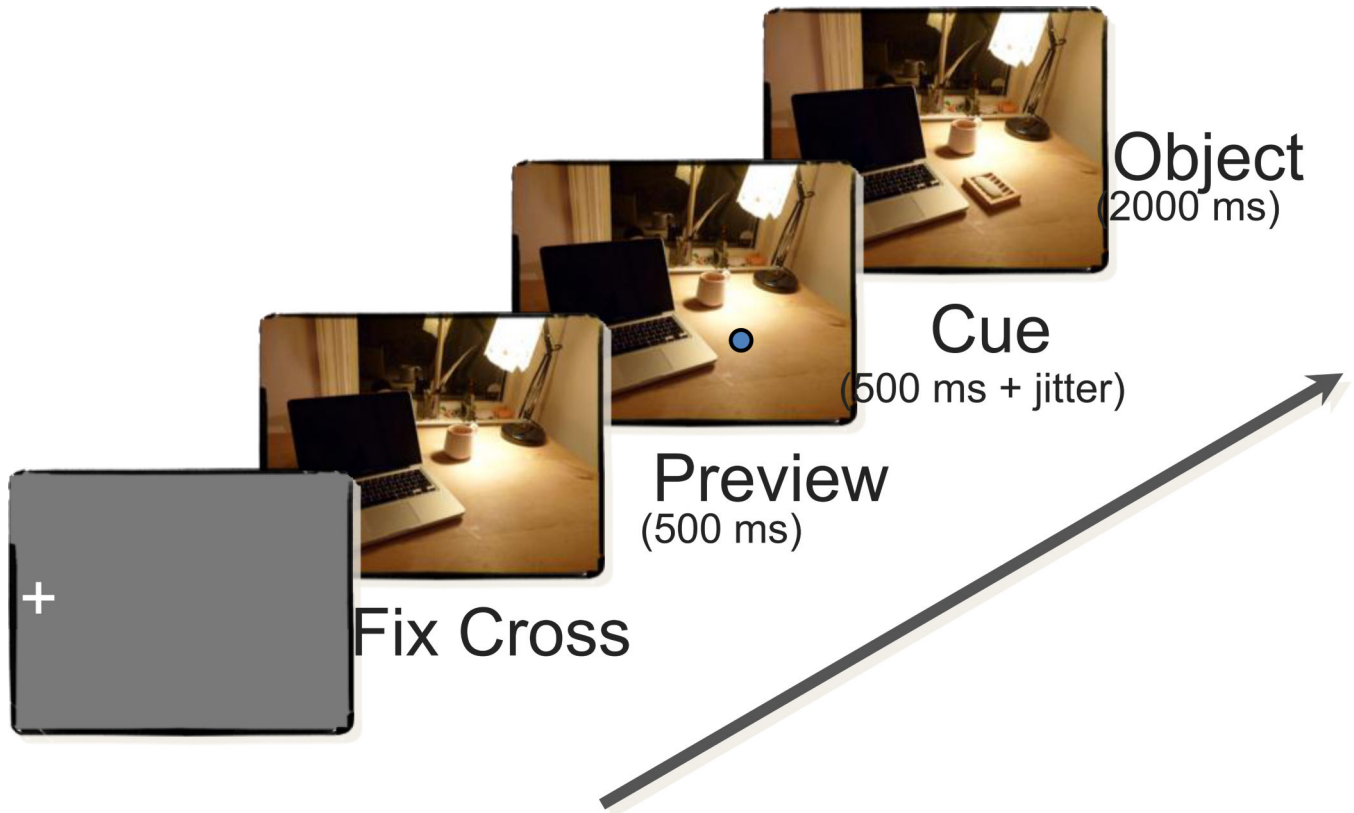


Figure 3.

Trial sequence starting with the presentation of a fixation cross that indicated when blinking was encouraged. Once ready, subjects initiated the trial per button press, which triggered the presentation of a preview scene without the critical object for 500ms, followed by a 500ms cue plus jitter. Upon presentation of the cue, participants moved their eyes to the cued location. Finally the object appeared at the cued location and remained visible on the screen together with the scene for 2000ms.

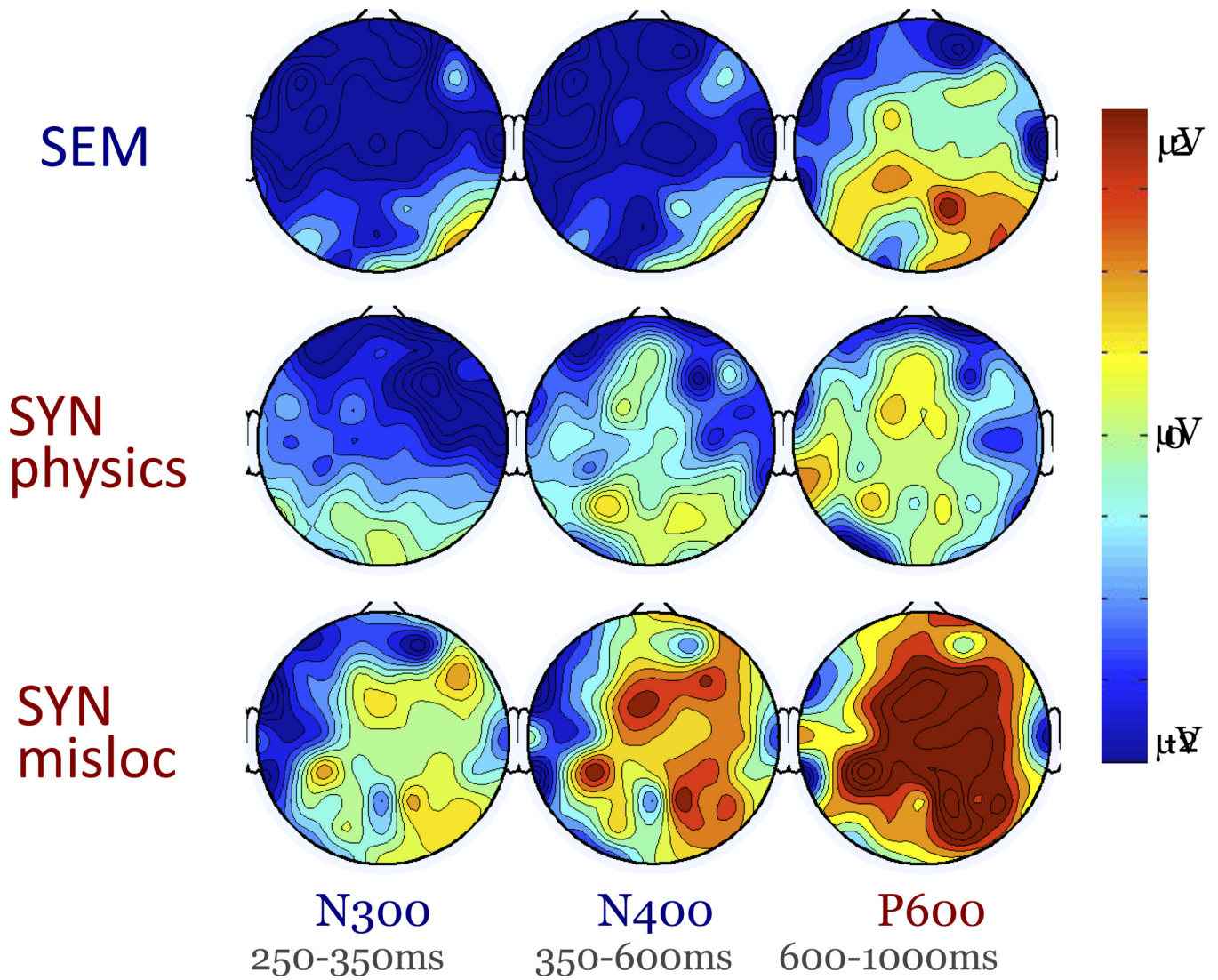


Figure 4. Scalp distributions of ERP effects (inconsistent minus consistent) for N300, N400, and P600 latency windows as a function of semantic, syntax-physics, and syntax-mislocated object-scene violations. Cooler colors indicate negative brain responses compared to the control, warmer colors indicate more positive responses.

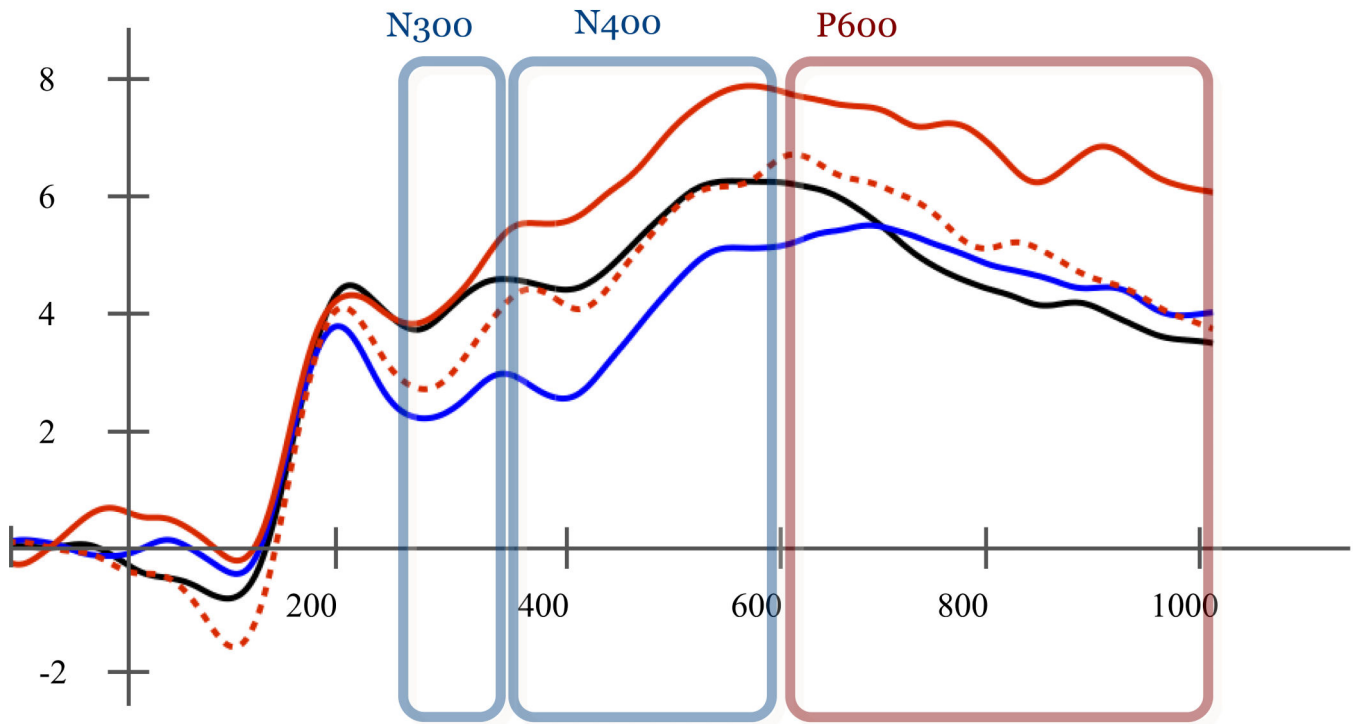
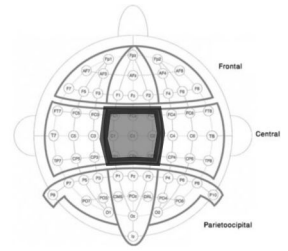


Figure 5. Grand average ERP waveforms measured at the mid-central region of consistent controls (black) and semantic violations (blue) together with syntax-physics violations (red dotted), and syntax-mislocation violations (red solid).

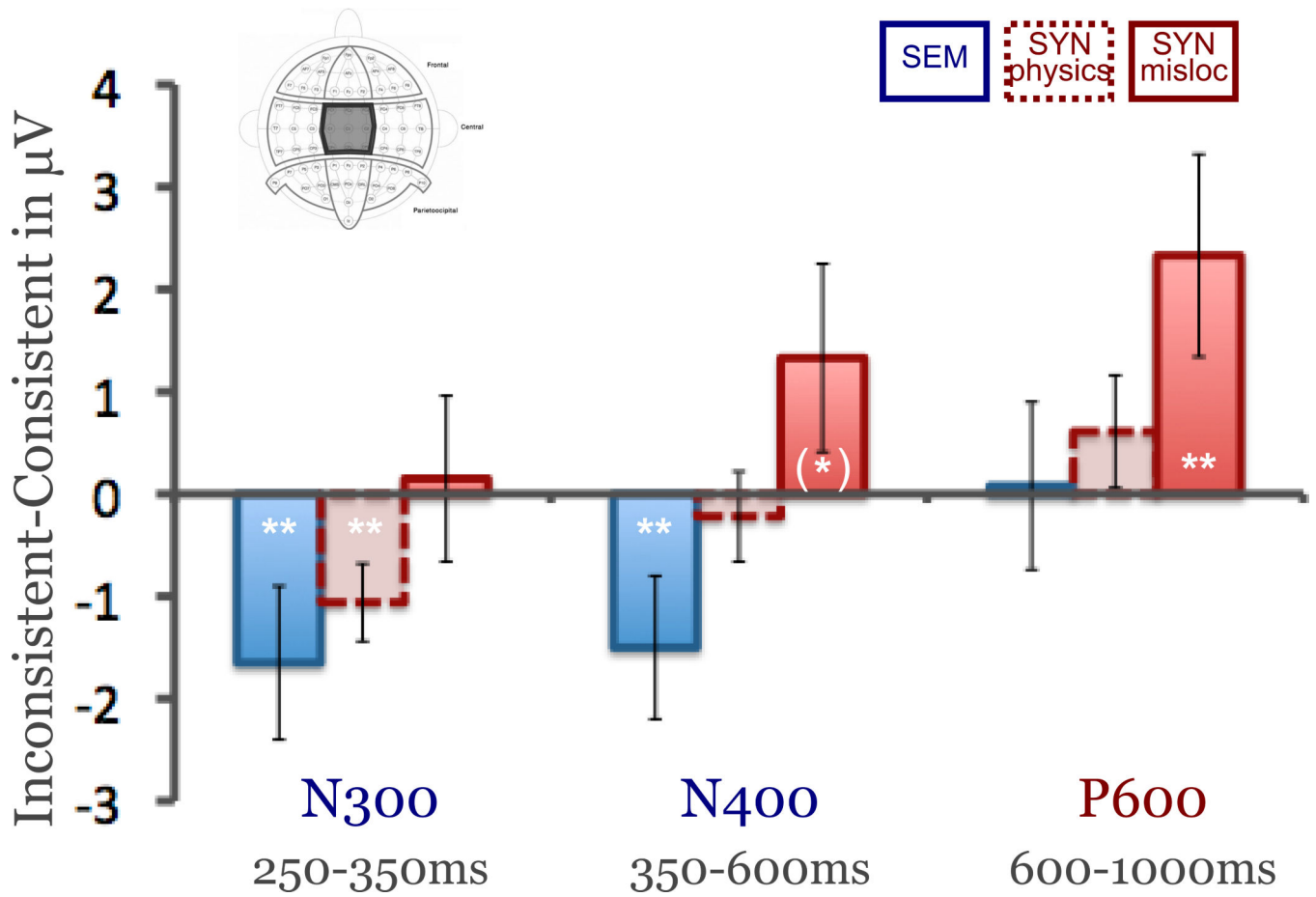


Figure 6. Mean voltage of effects (inconsistent minus consistent control) for N300, N400, and P600 latency windows as a function of semantic (blue), syntax-physics (red dotted), and syntax-mislocated (red solid) object-scene violations. Bars depict standard errors. Stars indicate statistically significant differences between inconsistency and control.