MAPPING THE TIME-COURSE AND CONTENT OF VISUAL PREDICTIONS WITH A NOVEL OBJECT-SCENE ASSOCIATIVE MEMORY PARADIGM

BY

CYBELLE MARGUERITE SMITH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor Kara D. Federmeier, Co-chair
Professor Diane K. Beck, Co-chair
Professor Gary S. Dell
Assistant Professor Sepideh F. Sadaghiani
Associate Professor Lili Sahakyan

ABSTRACT

In the current thesis, we present a series of three ERP experiments investigating the time-course and nature of contextual facilitation effects in visual object processing. In all three experiments, participants studied novel object-scene pairs in a paired associate memory paradigm. At test, we presented the scene first, followed after a delay by the test object, which either matched or mismatched the scene. We manipulated two key factors. 1) In all three experiments, we manipulated the severity of contextual mismatch between the presented object and the scene, including categorical violations as well as more subtle visual distortions. In this way, we probed the level of detail at which participants were reactivating the contextually-congruent target object in response to the scene. 2) We manipulated the scene preview timing parameters both between subjects (Experiments 2.1 and 3.1) and within subjects (Experiment 3.2). Our rationale for doing this was as follows. Rather than assuming that contextual facilitation effects reflect an entirely predictive or reactive/integrative process, we tested the hypothesis that contextual facilitation was predictive in nature. If the contextual facilitation was entirely integrative (i.e., people waited until the object was presented before relating it to the scene context), we might expect that the amount of scene preview time would not modulate contextual facilitation effects. What we found instead is that allowing for additional scene preview time leads to enhanced contextual facilitation effects, suggesting that participants are using the additional time that they are observing the scene alone (beyond 200 ms, which is sufficient to extract the gist of the scene) to prepare to process the upcoming object and determine whether it matches the scene. We strengthened our findings by testing this both between subjects using only two time points, and within subjects using a parametric gradation of preview times (which also allowed us to test if our findings generalized to cases of temporal uncertainty). We also took advantage of our use of ERPs to examine dependent measures tied to specific stages of cognition. We particularly focus our analysis and discussion on contextual priming of higher-level visual features, examining how contextual congruency modulates amplitude of the N300 component under various conditions and timing constraints. We also present a set of novel visual similarity analyses relying on V1-like features, which allow us to test for context effects on visual object understanding in a component-neutral fashion. Lastly, we present analyses of context effects on other components of the waveform: the N400, as an index of semantic priming, and the LPC, as an index of response-related processing. Overall, our

findings are consistent with a predictive account, in which participants use scene information to preactivate features of the upcoming object (including higher-level visual form features, as well as semantic features) in order to facilitate visual object understanding. Future work will further disentangle predictive vs. integrative processing accounts of contextual facilitation effects on visual object processing.

ACKNOWLEDGEMENTS

Thank you so much to Kara and the members of my committee for your feedback and support! In particular, outside of my primary advisor (who, duh! see below), this thesis would not have come together nearly as well without the careful feedback, thoughtful advice and critical support at various stages during this project of Diane Beck, our sharp and magnanimous resident Vision Scientist. Diane, you've really helped me in transitioning to a quite different field from where I started – thank you!

Kara, I cannot understate how important a role you've played in forming me as a scientist and a person over the last 5 years – as I've said before, you're brilliant, careful, ethical, kind, and although I will never share your time and financial management skills, I hope to emulate you at both a personal and professional level going forward in my career. You are my Yoda, and also, my "science mom."

Thank you so much also to the many other U of I faculty members that played a critically formative role during my time here. Thank you in particular to Gary Dell, who has been an important part of every major landmark of progress throughout my degree, and who consistently provides brilliant, insightful feedback at the highest level. You and Kara are the ones I most try to emulate when asking questions and giving feedback during lectures/talks because you cut right to the core issues in a kind and constructive way while avoiding all the crap that might sideline lesser scientists. It has been an honor getting to know you, and thank you for all the many ways you've helped me to grow! I can't list all the other numerous faculty that have influenced me in a positive way during my time here, but please know that each of you have really helped with my intellectual and professional development. The U of I Psychology Department is uncommonly devoted to a rigorous and ethical approach to science, and I'm so glad I got to be a part of that! I also want to give a rare shout out to the many supportive administrative staff at the Department (Hi Ashley!) and in the university who also played a big role in helping me along; in particular, recently retired Colleen Vojak and Ken Vickery at the Grad College of External Fellowships gave me critical advice and support that enabled me to win the NSF-GRFP, which substantially improved my quality of life here and has helped me to retain my sanity over the past three years.

Thank you to my current boss, talented and beloved research scientist and lecturer, Hillary Schwarb, under whose graces I'm able to write this right now instead of doing my TA

work. She has been uncommonly supportive and understanding this semester as I finish up my dissertation, even as she's on maternity leave. I really admire you!!

Thank you to my therapist, Dr. Sandra Leister, who is sharp, incredibly good at her job, and steadfastly willing to push my boundaries in various ways that have helped me to achieve a level of success that otherwise would likely not have been possible. I can say without exaggeration that this woman, as a person -- not as some replaceable member of a larger healthcare system – has played as critical of a role as my advisor in enabling me to make it through the last several years and complete my PhD. Thank you!

Thank you also to my labmates and lab alumni and to my incredibly supportive and wonderful network of friends that have helped me so much in making it over the finish line! A special thank you and shout out to Jamie Norton and Ryan Hubbard, whose friendship helped me make it through my low point in grad school and who helped me to re-learn that science can be social, fun, interesting and worthwhile. Thanks also to Willa Michener and Ken Oye, who visited me several times during graduate school and who helped me feel supported and more connected to my family and friends at home at a critical time. Willa, in particular, you really helped bridge the gap between my professional and personal life in a nourishing way by supporting and encouraging me as both a brilliant scientist and a close family friend. I really wish I could mention everyone here but if you are my friend and/or a CABLAB member or alum please know you really did help me a lot and I'm super grateful!

I also particularly want to thank Derar Al-Khawaldeh, my partner and best friend of the last two years. We are no longer romantically involved, but he has supported me emotionally and psychologically through the difficult daily grind of my last few years of grad school at a level that I had never thought would be possible for me in a relationship. I am still extremely lucky to be his close friend, and his compassion, sound advice and sense of humor continue to inspire me and light up my world. You make me a better person.

Thank you so much to my incredible, wonderful, talented, supportive, intellectually diverse, and unbeatable family! I could not have a better family if I had picked you guys out myself, particularly in the Trump era. Some of you have literally saved my life and I'm only here because of you. Thank you to my Mom and Dad and also to my wonderful stepparents Don and Ann. Mom, Dad and Ann all visited me out here in the boondocks of Central Illinois out of pure love – sometimes during extremely frigid weather – on several occasions. Don was incredibly

supportive in other ways and was always thoughtful and understanding as I unloaded all the many stresses of grad school during home visits. Naoka, Rachel and Brandon – you guys were all very emotionally supportive and it really helped me make it through – thank you! Jeff, Sanjay, Bryan and Elise – you guys rock and I am so happy to have you as in-laws and stepsiblings! I can't mention everyone but I'm also very appreciative of my other extended family members and close family friends, including my adorable and endearing nieces and nephews, and my many wonderful aunts, uncles and cousins, all of whom I visited during grad school, and who have also been very supportive. I only wish Grandma were still here but she also deserves thanks for helping me make it. I love all of you guys! I'm looking forward to coming home for Christmas and being the dumb one again.

Lastly, I wanted to send a shout out to my many previous mentors, teachers and friends that have inspired me and helped me along in my career, especially Colin Phillips and the faculty and graduate cohort I had the immense pleasure of meeting and working with during my year at University of Maryland. Colin – you were really my first advisor in many ways and your critical approach to science and the way that you and the other faculty managed to construct a highly collaborative, caring and productive environment, actively building up, promoting and improving an entire department, was truly inspirational. I haven't even heard of another place like that since. Thank you to Toshi Shioda, who kindly went out of his way to invite me to intern at his cancer research lab at MGH as a high school student after substitute teaching my Saturday morning Japanese-as-a-second-language class. You were an important early mentor and friend to me and I really appreciate all the ways you've helped me to achieve more with my life. Thank you to Brian Dunn, my favorite high school teacher who frequently entertained us with stories of world travel from his 20's in between Latin verb conjugation drills, and whose joie de vivre helped me form a sense of what a well-rounded, exciting and fulfilling life might look like. Thank you also to the Stanford alumni association and to Stanford's generous financial aid program, without which I likely couldn't have attended that school, or otherwise would have been saddled with crippling debt like many of my peers. Stanford was an amazing and exciting formative experience that I'll always remember fondly and that has shaped me intellectually in many ways. Thank you to the faculty at Stanford who played a particularly big role in helping me at an early career stage: John Rickford, Meghan Sumner, and Michael C. Frank. Prof. Rickford – you helped me get my feet wet in Linguistics and made me feel included in the

department as an undergrad in a way that would have been uncommon at another place; you also taught me that faculty can be both successful and chill. Meghan – if it wasn't for you I might not have taken the Maryland job, which would have been a huge mistake; you went beyond advising my honors thesis and teaching me the scientific method -- you really took an interest and helped mentor me at multiple levels, in a way that meant something. Thank you! Mike – you were an important mentor to me during my transition to psycholinguistics, and you gave me great advice and advocated for me on numerous occasions; you're also a superstar and an extremely personable guy that I will always admire at both a personal and professional level.

Thank you to everyone! You all really made a difference!!

To my friends and family, and James Marsters

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

The current thesis explores the nature of contextual facilitation effects in visual object recognition, with a focus on scene-object priming. We make use of a scene-novel object picture association paradigm over the course of three ERP experiments. Participants study pairings between natural scenes and line drawings of novel objects in an adaptation of an ERP scene-face associative memory paradigm (Hannula, Federmeier & Cohen, 2006). Participants are then tested on their memory for the paired associations by first viewing a scene context, followed by an object that matches or mismatches the scene. Critically, by manipulating the delay between onset of the (priming) scene context, and the following object target, we can explore to what extent facilitation for associated objects is mediated by preparatory processing. Also, by manipulating the relationship between the object that is presented and that which was previously studied with the scene, we can explore what attributes of the visual object are pre-activated in response to the scene. Thus, the current dissertation explores multiple aspects of how scene information can be used to facilitate visual object understanding, with a focus on the question of which stages of object processing are facilitated, and under what (timing) constraints.

Overview of Experiments and Chapter Guide

In our first experiment (Chapter 2, Experiment 2.1), we test whether novel visual objects can be partially preactivated in response to a mismatching but related context. We initially assess whether participants are able to rapidly learn novel object-scene associations, as well as higher-level categorical associations among object and scene types, and whether this learning transfers to graded memory effects at the target object picture. Moreover, we then assess whether the N300 to objects is facilitated by scenes, as such facilitation would suggest that scenes facilitate object processing at a visuo-structural level even for recently formed scene-object associations, which has been unconfirmed in the literature. We then additionally apply a novel, component-neutral analysis of visual similarity between the contextually congruent target object and the presented object on mismatch trials, as an alternative approach to detecting whether scenes facilitate visual object processing at the level of visual form.

In our second experiment (Chapter 3, Experiment 3.1), we drastically shortened the amount of contextual pre-exposure time (i.e. the amount of time scenes were presented in

isolation, prior to superimposing an associated or unassociated object), relative to the first experiment, from 2500 ms to 200 ms. This serves as a first test of whether the contextual facilitation effects found in our first experiment are predictive in nature. To the extent that the pattern of associative memory effects changes across Chapter 2 Experiment 2.1 and Chapter 3 Experiment 3.1, this is consistent with the idea that the effects in our first experiment are driven by preparatory processing.

Our third experiment (Chapter 3 Experiment 3.2) was designed as a within subjects conceptual replication and extension of the first two experiments. We parametrically varied the amount of contextual pre-exposure time at test between 0 (concurrent scene-object presentation) and 2500 ms. In this way, we could not only test whether our findings generalize across slight variations in task demands, but also estimate at what point the predictive benefit on object processing associated with longer scene preview times asymptotes. That is, we can identify roughly how long it takes for viewing a scene in advance to yield maximal contextual benefits on subsequent object processing. Also, again, through the use of ERPs, we can gain an indirect measure of what specific processing stages are likely facilitated by scene congruency under various timing constraints.

In Chapter 4 (discussion), we review the overall pattern of findings from the three experiments together and conclude with suggested future directions.

In the following two sections, we review some relevant background information that supplements the information provided in Chapters 2 and 3.

1. Context-Based Prediction in Language Comprehension

The introductions in Chapters 2 and 3 frame our scientific contributions purely in terms of building upon what we know already about the contextual benefits that congruent scenes afford to visual object processing, specifically. However, it can also be helpful to review what we know about contextual facilitation effects in another domain, and to build parallels across domains, in the interest of building stronger and more generalizable frameworks for understanding cognition. Here, we review findings from the domain of language comprehension, with an eye towards establishing what types of theoretical distinctions should be made, and what kinds of questions

should be answered, in order for us to have an adequate understanding of scene-object contextual facilitation effects at a cognitive level of analysis.

Although for many years theoretical models of language comprehension focused on "bottom-up," stimulus-driven processing of linguistic input, more recent models have highlighted the role that anticipatory, and often "top-down," mechanisms may play in accounting for a variety of effects (see Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005 for a historic overview). Such anticipatory mechanisms may help explain the rapid pace at which language is understood (connected speech rates hover at around 4 syllables / second). Anticipatory processes may in principle operate at multiple levels of linguistic structure. For example, anticipation likely plays a role in choosing an appropriate time to speak when taking turns in conversation. But here we will focus on context-based pre-activation of semantic and word-level features.

One complication when discussing anticipatory processing is that there is often ambiguity in the extent to which anticipatory processes are in fact "top-down." Thus, it's necessary to clarify what we mean by "prediction" with respect to the role of executive function. Many authors contrast priming (as a more bottom-up, passive phenomenon) with prediction (often, as a more strategic process subject to executive control). However, given that our proposed experimental manipulations do not directly speak to the role of executive function in anticipatory processing, the current work will use "prediction" and "preparatory processing" as umbrella terms that encompass both prediction as mediated by more complex executive processes, and simple associative or semantic priming. So long as a preceding context has biased the ease of processing of subsequent stimuli in a stimulus-specific manner (in contrast with general arousal), we will call that "prediction."

Anticipatory processing in language comprehension manifests itself in a variety of ways. Naming latencies and lexical decision times are reduced for words preceded by facilitating contexts (e.g., Hess, Foss, & Carroll, 1995; Kleiman, 1980; McClelland & O'Regan, 1981). In addition, reading times are faster, and skip rates are higher, for words that are predictable given their preceding context (e.g., Ehrlich & Rayner, 1981; McDonald & Shillcock, 2003). Lastly, manipulations of word predictability in context result in highly reproducible neural effects during sentence reading, as assessed using EEG and MEG (e.g., Kutas & Hillyard, 1984; Helenius, Salmelin, Service, & Connolly, 1998).

One neural response that is particularly sensitive to word predictability in context is the N400 Event Related Potential (ERP) component. The N400 is a negative-going waveform appearing roughly 300-500 ms after the onset of a word or other potentially meaningful stimulus. The N400 has been systematically studied for over 35 years, and many of its functional sensitivities, including to semantic, lexical and orthographic properties of words, as well as their interactions with contextual factors, have been extensively replicated (see Kutas & Federmeier, 2011 for review). In part due to its very consistent latency profile, one theory of the N400 suggests that it reflects the process of accessing a semantic long term memory network to temporally bind together semantic features of an incoming stimulus (see, e.g., Federmeier & Laszlo, 2009). N400 amplitude has a tight statistical relationship with the cloze probability of a word given its preceding context (e.g., Wlotko & Federmeier, 2012). The cloze probability of a word is defined as the proportion of the time that human participants generated that word as the first word in a sentential continuation, given its preceding context. N400 responses to highly predictable (high cloze) words are smaller in amplitude than those to low cloze words. N400 amplitude is also reduced in response to purer semantic priming, both as a function of similarity along preset feature dimensions and textual co-occurrence statistics. Thus, there are multiple converging sources of evidence that preceding context facilitates semantic processing of target words, both when presented singly and in sentences.

Newer evidence has also suggested that participants may in some cases anticipate more detailed information about upcoming words, using mechanisms that appear outside the domain of simple priming. For example, Dutch adjectives mismatching the syntactic gender of a highly expected following noun (cloze > 75%) elicit an ERP positivity shortly after the critical adjective's gender inflection (Van Berkum et al., 2005). A similar set of effects have been described for determiners mismatching the syntactic gender of expected upcoming nouns (or, with different polarity, pictures of these) in Spanish (Wicha, Moreno, & Kutas, 2004; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003). Lastly, a more highly contested set of predictive effects have been reported for the English determiners "a" and "an" when the anticipated following noun began with a vowel or consonant, thus matching or mismatching the presented determiner (DeLong, Urbach, & Kutas, 2005; cf. Nieuwland et al., 2017; Yan, Kuperberg, & Jaeger, 2017).

Importantly, recent work on context-based prediction in language comprehension has led to key insights about the nature of prediction that may or may not extend to other domains:

- *Sensitivity to semantic memory structure*. Federmeier and Kutas (1999a), in a study on young adults, found that words that did not fit a preceding sentential context, but nonetheless were semantically related to a predicted word, were partially facilitated by the (incongruent) context. Moreover, these effects became larger in high constraint sentences, where prediction would be hypothetically more advantageous.
- *Subsequent memory effects of prediction.* When words are predicted in a particular sentential context, but not actually presented, they nonetheless generate ERP repetition effects when they are eventually subsequently displayed in the experiment (Rommers & Federmeier, submitted).
- *Effects of constraint and prediction error*. Congruent but unexpected words in highly constraining sentences generate a frontal positivity and concomitant increase in frontal theta band activity, which may be interpreted as neural correlates of prediction error (see Rommers, Dickson, Norton, Wlotko, & Federmeier, 2017).
- *Hemispheric differences*. Several lines of evidence suggest that the left hemisphere is more sensitive to higher level linguistic constraints and also shows signs that expected but never presented words are partially activated, while the right hemisphere pre-activates semantic features in a more bottom-up, long-lasting, and loosely constrained fashion (reviewed in Federmeier, 2007; further discussion in section 4 of the introduction, below).
- *Age-related changes*. Older adults tend to show more limited signs of spontaneous top-down prediction of upcoming words, as suggested by, for example, reduced facilitation for incongruent words that are related to words predictable given a preceding sentential context (Federmeier, McLennan, De Ochoa, & Kutas, 2002).
- *Strategic changes*. The rate at which linguistic contexts are followed by semantically related words can modulate the extent to which N400 effects of semantic pre-activation are observed (Lau, Holcomb, & Kuperberg, 2013). Also, being asked to predict enhances N400 facilitation in processing high cloze sentence-final words (Brothers, Swaab, & Traxler, 2017). Thus, people tend to predict less when their incentive to predict is diminished, and more when it is enhanced.

To what extent does predictive processing in visual object recognition share these features? Some findings suggest that properties of verbal prediction outlined above may indeed extend to other domains. For example, frontal theta also appears in response to prediction errors generated in purely non-linguistic tasks, suggesting that some mechanisms for handling erroneous predictions may hold across domains (e.g., Cavanagh, Frank, Klein, & Allen, 2010). Furthermore, when a visual search display is preceded by a context cue, the proportion of the time that the context cue signals the correct location of a target affects the degree to which the context cue facilitates search (personal communication with Alejandro Lleras). This suggests that the degree to which participants anticipate target locations may be strategically ramped up based on global contextual factors, similar to strategic modulations of prediction effects in language comprehension. Lastly, pictures embedded in sentential contexts often show contextual facilitation effects that parallel those found for words, including interactions with hemisphere (e.g., Federmeier & Kutas, 2002). However, to the extent that sentence-picture priming effects are at least partly linguistically mediated, these findings may not extend to non-verbal contextual primes.

The current thesis will assess what stages of object processing are facilitated by scenes, and speak to whether this facilitation is predictive in nature. In Chapter 2, we will also test whether one characteristic of verbal prediction holds up for purely pictorial stimuli: graded predictive facilitation of contextually incongruent but nonetheless categorically related stimuli. In addition, in the Chapter 3 we will assess the timing properties of prediction more systematically and quantitatively than previous studies in either the verbal or non-verbal domains – thereby directly answering questions about prediction in the domain of visual object recognition and also generating predictions for future language comprehension research.

2. ERP Signatures of Visual Prediction

Evoked activity to visual images can be modulated by the preceding context along several dimensions. In this section, we introduce one ERP component classically associated with visual picture priming that will prove relevant to interpreting our experimental results in all three experiments: The N300.

The N300 (or 'N350') refers to a negative deflection in the ERP waveform between roughly 200 and 400 ms that decreases in amplitude in response to picture priming and related manipulations. The N300 can be elicited in response to both natural, photographic images, and line drawings. It tends to have an overall anteriorly biased distribution, which nonetheless often extends to centro-parietal sites. Although in the current proposal, the N300 will primarily be treated as a unitary complex, there is some evidence for the presence of dissociable subcomponents (e.g., Schendan & Maher 2009).

The N300 is thought to be more tightly yoked to perceptual processing in the visual modality than the N400, which is believed to reflect processing at a more abstract, amodal semantic level. Part of this is due to the N300's classic association with pictorial stimuli. Semantic picture priming paradigms, for example, regularly elicit both an N300 and N400 effect (e.g., McPherson & Holcomb, 1999), while analogous manipulations of word stimuli only elicit an N400[1] (see Kutas & Federmeier, 2011 for review). Both semantically related images (Barrett & Rugg, 1990; McPherson & Holcomb, 1999) and (congruent) masked verbal labels (Chauncey, Holcomb, & Grainger, 2009) have been shown to elicit N300 (and N400) reductions to subsequently presented picture targets. Not all priming manipulations, however, lead to clear N300 effects.

Hamm, Johnson, and Kirk (2002), for example, examined the pattern of occurrence of N300 and N400 effects across several semantic priming manipulations in a passive viewing task using word cues followed by target line drawings. A given picture could be preceded by a congruent or mismatching basic level or subordinate level name (basic and subordinate level names were blocked). The authors then examined four types of contrasts: 1. subordinate-between (the cue word 'robin' followed by a target picture of a collie – the word 'collie' followed by a picture of a collie), 2. subordinate within (the word 'poodle' followed by a picture of a collie - the word 'collie' followed by a picture of a collie), 3. basic (the word 'bird' followed by a picture of a collie – the word 'dog' followed by a picture of a collie), and 4. basic-subordinate combination (the word 'bird' followed by a picture of a collie – the word 'collie' followed by a picture of a collie). Critically, while all four priming manipulations elicited a reduced N400 to the primed picture, N300 priming was observed more selectively. Specifically, while 'collie'

[1] N300-like effects have also been reported for words, however, using masked repetition priming: repeated words, orthographic neighbors, and, more-so than control neighbors, pseudo-homophones (e.g. 'bakon-BACON' > 'bafon-BACON') all can reduce amplitude of an anterior negativity known as the 'N250' (reviewed in Grainger & Holcomb, 2009). To the extent that these effects reflect an orthographic processing stage of word recognition, however, they are still consistent with a visual processing account of N300-like deflections.

primed the N300 to a picture of a collie more-so than 'robin' or 'bird,' 'poodle' primed the N300 just as well as 'collie,' suggesting that words denoting visually similar templates primed equally well. Interestingly, 'dog' did not prime the N300 more than 'bird,' possibly due to there being a reduced incentive to predict specific visual forms in blocks where only basic-level category primes were used.

Several other studies have suggested that more elaborate visual contexts, such as natural scenes and videos, may also prime the N300, but the presence of immediately following anterior N400 effects have made results difficult to interpret. The appearance of contextually incongruent objects in static images (Mudrik, Lamy, & Deouell, 2010; Võ & Wolfe, 2013) and video clips (Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008) have been claimed to elicit a larger amplitude N300 and N400 response relative to contextually congruent objects. However, in some cases, only a (somewhat more anterior than is typical for words) N400 effect is elicited (e.g., Ganis & Kutas, 2003). Still frames taken from videos of natural, descriptive gestures that matched or mismatched a preceding cartoon have also been suggested to elicit an N300 (as well as an N400) effect (Wu & Coulson, 2011). In the current thesis, we have developed a novel analysis technique using visual similarity effects (described in more detail in the following chapters), which can get around problems with component-based analyses, including difficulty isolating the N300 and N400 responses, which partially overlap in both time and space.

Schendan and colleagues conducted another series of experiments that further shed light on the functional sensitivity profile of the N300, and point to an intermediate visual processing stage dissociable from the multimodal, abstract semantic processing that occurs during the N400 time window. Schendan and Kutas (2002) presented sequences of fragmented images at progressively decreasing levels of fragmentation and asked participants at each stage to name the objects or else indicate that they were unidentifiable. They demonstrated that the N300 is larger for unidentifiable than identifiable fragmented images, but not when the images are so fragmented that their contours are unrecoverable. Schendan and Kutas (2003) used a repetition paradigm to examine differential processing and memory representations for pictures of objects in canonical vs. unusual views. They found that the N300 is more positive to stimuli presented in canonical than unusual views. Thus, the more easily an object can be identified, perhaps based on global visual contours, the smaller its N300.

Schendan and Kutas (2007) explored repetition priming on the N300 using a repeated picture naming implicit memory paradigm. During both study and test, the participant viewed line drawings of nameable objects and pushed a button as soon as they knew the name of the object presented, and then later indicated how confident they were and named the object. During the study phase, images were presented for 5000 ms, while during the test phase, they were only

presented for 17 ms. At test, line drawings were always 'segmented' – that is, dashed lines were used to represent contours of the object. However, earlier at study, participants had viewed either an identical segmented image ('Same'), a complementary segmented image (i.e. one with dashes where blanks would be at test and vice versa; 'Complementary Fragment'), an intact image (with fully connected lines, 'Intact'), or they had not studied that image before ('Unstudied'). In a separate experiment, they also tested the additional study conditions of 'Half Fragment' (where segments were half as long) and 'Intact-Segmented' (where isoluminant green and red segments were alternated) – for this second experiment, all line drawings were presented in green, with the exception of the red segments in the 'Intact-Segmented' study images. They found that the N300 showed a similarly large repetition priming effect across all conditions, suggesting that the N300 reflects processing of global shape contours, rather than 'local features.'

Several attempts have been made to source localize the N300, and to theoretically characterize it within a broader visual processing network. Schendan and colleagues (Schendan & Kutas, 2007; Schendan & Maher, 2009) situate the N300 in a "two-state interactive account of visual object knowledge activation" (later extended to a multi-state account in Schendan & Ganis, 2012). They propose that after an initial feed-forward activation of occipito-temporal cortex, reflected in the P150/N170 response, occipito-temporal cortex is once again activated during the N300 time window, but this time is more subject to executive control (specifically, they propose that the ventrolateral pre-frontal cortex, VLPFC, should play a particularly important role in regulating N300 activation). Attempts at ERP source localization are indeed consistent with an occipito-temporal source for the N300 (Schendan & Maher, 2009), but the link to the VLPFC is somewhat weaker and more indirect. fMRI experiments using paradigms known to generate N300 effects can also inform attempts at source localization. In one fMRI study, (Ganis, Schendan, & Kosslyn, 2007) asked participants to categorize more and less impoverished images of objects and pseudo-objects. They found that ventro-occipital cortex was more highly activated by more impoverished than less impoverished images. Also, frontal areas including VLPFC were more active on trials where objects were marked as categorized after a longer period of time (which correlated with a higher degree of object fragmentation), than a shorter period. Overall, data have been consistent with an account in which the N300 reflects top-down activation of intermediate visual areas (perhaps encoding global shape information) in order to categorically assess visual input.

CHAPTER 2: NEURAL SIGNATURES OF LEARNING NOVEL OBJECT-SCENE ASSOCIATIONS

Abstract

Objects are perceived within rich visual contexts, and statistical associations may be exploited to facilitate their rapid recognition. Recent work using natural scene-object associations suggests that scenes can prime the visual form of associated objects, but it remains unknown whether this relies on an extended learning process. We asked participants to learn categorically structured associations between novel objects and scenes in a paired associate memory task while event-related brain potentials (ERPs) were recorded. In the test phase, scenes were first presented (2500 ms) followed by objects that matched or mismatched the scene; degree of contextual mismatch was manipulated along visual and categorical dimensions. Matching objects elicited a reduced N300 response, suggesting visuo-structural priming based on recently formed associations. Amplitude of an extended positivity (onset ~200 ms) was sensitive to visual distance between the presented object and the contextually associated target object, most likely indexing visual template-matching. Results suggest recent associative memories may be rapidly recruited to facilitate object recognition in a top-down fashion, with clinical implications for populations with impairments in hippocampal-dependent memory and executive function.

Key Words: visual object recognition; N300; template matching; paired associate learning; statistical learning

Introduction

Rapidly perceiving and categorizing visual objects is an important survival skill, so much so that humans and other animals capitalize on statistical regularities in the environment in order to enhance its efficiency (Friston, 2005; Ranganath & Ritchey, 2012). It has long been known that objects presented in congruent visual contexts are recognized more accurately and rapidly and processed more efficiently than those in incongruent (but equally cluttered) contexts (Palmer, 1975; Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989; Boyce & Pollatsek, 1992; Davenport & Potter, 2004; Ganis & Kutas, 2003). Although much research has

focused on what aspects of visual environments are used as cues for object recognition (reviewed in Oliva & Torralba, 2007), a smaller body of work has addressed what aspects of visual objects are brought online or primed in response to contextual cues (e.g., Biederman et al., 1982; Henderson & Hollingworth, 1999; Bar, 2004; Truman & Mudrik, 2018; Schendan & Kutas, 2007; Brandman & Peelen, 2017). These studies suggest that both semantic and structural information about visual objects are primed by scene contexts (e.g., Gronau, Neta, & Bar, 2008; Brandman & Peelen, 2017; Mudrik, Lamy, & Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014; Truman & Mudrik, 2018). However, these studies also largely relied on naturally occurring statistical associations that may be learned over long periods of time. Thus, it is possible that scene-based structural priming of visual objects is contingent on an extended learning process, which has neurophysiological and clinical implications. To test this, we primed novel objects using newly associated visual scene contexts and checked for activation of object-related category-level and lower-level visual information.

Context effects on visual object processing are abundant and suggest both semantic and form-based facilitation. Interpretations of ambiguous, degraded, briefly presented, and masked objects are influenced by their visual context (Bar & Ullman, 1996; Freeman et al., 2015; Barenholtz, 2013; Palmer, 1975; Davenport & Potter, 2004; Brandman & Peelen, 2017). For example, an identical visual percept of a blurred object can be categorized as either a hairdryer or a drill depending on whether it is embedded in a bathroom or workshop scene (Bar, 2004). Similarly, briefly presented objects are more likely to be misidentified as a visually similar object following presentation of a context associated with the incorrect item (Palmer, 1975). Recent evidence has suggested that even racial judgments of faces on an Asian-White continuum are biased by presentation in an American or Chinese setting (Freeman, Ma, Han, & Ambady, 2013; Freeman et al., 2015). However, such effects of scene context on object categorization may be driven by semantic effects on higher level decision processes and do not necessarily reflect form-based priming (Henderson & Hollingworth, 1999; Truman & Mudrik, 2018).

The argument that scenes facilitate object identification by supporting form-based matching processes has been primarily based on indirect neurophysiological evidence (Bar, 2003, 2004, 2007; Brandman & Peelen, 2017; Truman & Mudrik, 2018). Of relevance to the current study, an event-related potential (ERP) component known as the N300 (also referred to as the N350, Ncl, or simply the N3 complex) is elicited by visual objects, and its amplitude is

often, but not always, modulated by scene congruency (Võ & Wolfe, 2013; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008; Mudrik et al., 2010; cf. Ganis & Kutas, 2003). The N300 is a fronto-central negativity peaking between roughly 200 and 400 ms, believed to index an object model selection process based on global shape information (Schendan & Kutas, 2002, 2003, 2007), although slight variations in scalp topography across manipulations and over time suggest it may encompass multiple related processes (Truman & Mudrik, 2018; Schendan & Ganis, 2012). Although semantic priming consistently reduces the amplitude of a later and more centrally distributed component, the N400, amplitude of the N300 is reduced more selectively by form-based priming (Hamm, Johnson, & Kirk, 2002; Kovalenko, Chaumon, & Busch, 2012). N300 amplitude is also modulated by factors pertaining to isolated object images, such as the noise level at which a visually degraded object is recognized, and viewpoint canonicity (Doniger et al., 2000; Schendan & Kutas, 2003). fMRI data also suggest that contextual congruency of objects modulates activity in a network including the lateral occipital complex (LOC), which is known to store high-level form representations of visual objects and is believed to be the human analogue of monkey inferotemporal (IT) cortex (Tootell, Tsao, & Vanduffel, 2003, Gronau et al., 2008; Grill-Spector et al., 1998). Indeed, N300 scalp topography is consistent with generators in the ventral visual stream, including the LOC (Schendan & Ganis, 2012). Moreover, recent fMRI and MEG evidence has shown that cross-decoding accuracy of object animacy in LOC is enhanced by simultaneous presentation of disambiguating scenes, and that peak cross-decoding accuracy for objects embedded in scenes is at around 320-340 ms (Brandman & Peelen, 2017). Because of its numerous links to high-level visual processes, including perceptual closure and viewpoint accommodation, as well as form-based visual priming, differences between contextually congruent and incongruent objects on the N300 component have been taken as an index of scene priming on visual form-based matching processes during object recognition.

In the current study, we will not only assess object N300 scene-congruity effects, but will also more directly confirm the engagement of perceptual matching processes based on a component-neutral item-based analysis. In our component-neutral analysis, visual similarity between the presented (distorted or mismatching) object and the object expected based on the preceding context is used to predict ERP amplitude. To the extent that visual similarity to the contextually-congruent target object predicts ERP amplitude, this suggests that visual information about the target object has been brought online in response to the scene. In this way,

we can confirm visual form priming via two different and complementary approaches: a component-based analysis drawing from the N300 literature, and a component-neutral analysis that makes a more direct inference based on the results of a single data set.

It is not yet known whether form-based priming of object representations by scenes is contingent on an extended learning process, which would suggest a crystalized associative representation less subject to the influence of task demands and disruption by disturbances in short-term memory and executive function. Prior evidence is suggestive that at least some aspects of object-scene priming may rely on labile attentional and hippocampally-dependent processes. For example, context-object associations can be rapidly implicitly learned and applied to facilitate tasks such as visual search (reviewed in Jiang & Chun, 2003). Using an associative memory paradigm, Hannula and colleagues have shown that arbitrary face-scene pairs can be rapidly learned and yield scene congruency effects on face viewing times within a single experimental session, as long as the hippocampus is intact (Hannula, Ryan, Tranel, & Cohen, 2007). Moreover, an ERP adaptation of Hannula and colleagues' face-scene associative memory task revealed that faces matching recently associated scenes elicit a less negative scalp potential at 300 ms, which has timing and scalp topography consistent with N300 facilitation (Hannula, Federmeier, & Cohen, 2006). The current study extends Hannula and colleagues' design to novel object-scene pairs, in an attempt to elicit this early ERP congruency effect using a wider class of object stimuli, and to explicitly link it to visual form analysis. If, indeed, form-based priming of visual objects can be based on recently formed associations, this raises the possibility that visual object recognition in the natural world may be subtly disrupted in elderly or clinical populations that are less able to rapidly form and recruit novel associations.

In the current study, participants associate line drawings of novel objects with natural images of visual scenes. To facilitate learning, novel objects are grouped into higher-level categories, each of which is exclusively associated at study with exemplars of a particular scene category (beaches, highways, etc.) for any given participant. In the test phase, participants are shown scenes from the scene-object pairs they recently studied, followed after 2500 ms by an object that a) exactly matches what had originally been studied on that specific scene, b) is a distorted version of the object they studied on that scene, c) is the wrong object altogether but belongs to the same scene-congruent higher level category (i.e. an object that has been studied on beaches in general but now is displayed on the wrong specific beach), or d) is the wrong object

and belongs to a scene-incongruent category (i.e. a highway-associated object on a beach). If form-based priming of objects by scenes can be induced by recently learned associations, we should expect the N300 to be reduced in amplitude to objects matching the scene, relative to mismatches. We should also expect the ERP response to show graded sensitivity to the degree of structural mismatch between presented incongruent objects and the scene-congruent target object, which we assess in a separate analysis. In addition to the N300, we will also examine context-based facilitation of the N400 and late positive complex (LPC) components, which have been linked to semantic memory and extended visual and task-relevant processing, respectively.

Methods

Participants

Data are reported from 24 participants (mean age 21 (18-28); 10 males), all native English speaking University of Illinois undergraduates, who were compensated with payment. Three additional participants were replaced due to excessive trial loss or, in one case, use of substances on our exclusionary list. All participants provided written informed consent, according to procedures established by the IRB at the University of Illinois. Handedness was assessed using the Edinburgh inventory (Oldfield, 1971). All participants were right-handed; mean score: .86, where 1 denotes strongly right-handed and -1, strongly left-handed. 10 reported having left-handed family members. No participants had major exposure to languages other than English prior to the age of 5. All had normal or corrected-to-normal vision, and none had a history of neurological or psychiatric disorders or brain damage or was using neuroactive drugs. Participants were randomly assigned to one of 24 experimental lists.

Materials

Overview

Objects were novel, designed to resemble biological organisms (“germs”), or mechanical devices (“machines”). Major object categories were paired with familiar scene categories (e.g., beaches,

highways), such that each type of scene was associated with one type of germ and one type of machine for each participant. At test, participants either saw exactly the same object that they had studied, an object from the same category subtype (e.g., a distortion), an object from a different subcategory (that had not been studied on that particular scene but had been studied on the same *type* of scene), or an object from a different major category (that thus would never have appeared on any scene of that type). Across the experiment, all object types were paired with all scene types, and objects and scenes were never repeated in the study phase. Details of stimulus development and counterbalancing follow.

Scenes

Stimuli consisted of pairings between photographs of natural scenes and line drawings of novel objects (see Figure 2.1). Scenes depicted one of 6 categories: beaches, city streets, mountains, forests, highways, and offices. Scenes were drawn from a pool of 288 images, 48 per category, that were previously normed as being highly representative of their respective scene types, and which had been rescaled to 800 x 600 pixels (see Torralbo et al., 2013 for norming details).

Objects

Line drawings of novel object prototypes for biological organisms ("germs") or mechanical devices ("machines") were created by an artist with the aid of Adobe Photoshop to maintain a consistent set of visual textures. Within the two classes of germs and machines, drawings were further organized into 6 major categories, each with 3 subcategories. Thus, there were 18 total subcategories of germs and machines, respectively, each with a single representative prototype image. Major categories shared aspects of their visual structure and texture, as well as homologous parts. Each subcategory prototype image was scanned and roughly centered, then further manipulated using Photoshop and the animation software Unity to derive individual exemplars. For each subcategory prototype, we altered the prototype image along 3 continuously changing parameter scales, with the prototype image at the center of each scale. We then took snap shots of the distorted images at various points along each parameter scale to generate 3 continuously varying sets of 8 exemplars (24 exemplars / object subcategory), with 4 exemplars

on either side of the central prototype within each set. We never displayed the original prototype images in any experiment. For germs, parameter changes induced global distortions, such as gradually twisting the body or changing its width, while for machines, which were assumed to have a more rigid body, parameter changes affected the size and relative positions of component parts, which were first extrapolated from the prototype images using Photoshop layers to avoid unnatural gaps. Often, several parameters would change at once along each scale to heighten the visual dissimilarity among exemplars. See Figure 2.2 for an illustration of how exemplars were derived from the prototype images. All novel object line drawings were resized to 274 x 274 pixels.

Counterbalancing and Experimental Conditions

24 experimental lists of stimuli were created, grouped into 6 sets of 4. Each list was assigned to a single participant. Lists within each set maintained a fixed study trial relationship between the 6 major categories of novel objects within each class (germs or machines), and the 6 scene categories. Thus, for any given list (and participant), each germ or machine type would only ever appear on a single scene category (e.g. "beaches") at study, and this relationship would hold for all 3 subtypes of the object major category. The mapping between object and scene categories was then systematically rotated across the 6 sets, so that over the entire experiment every object category was associated with every scene category at study.

Each list consisted of 288 study and 288 test pairs of novel object and scene stimuli. Stimuli within each list were organized into 18 blocks of 16 study pairs followed by 16 test pairs. Blocks alternated between all germs and all machines; the first block for each list was always germs. The same set of 288 unique scenes was used across all lists; scenes were presented exactly once at study and once in the corresponding test phase in each list. Each set of 4 lists had 288 unique test objects that were used across all 4 lists; these objects were randomly drawn (without replacement) from the total set of possible exemplars, within constraints of the experimental design. Test objects were never repeated within a list, but were sometimes repeated across lists, such that some test object images were used more often over the full counterbalancing than others. 793 novel object images out of 864 possible exemplars were presented as test objects over the course of the experiment.

Within each block, an approximately equal number of pairs involving each scene category and object major category were presented at both study and test. All 16 scenes presented at study within a block were presented in a different pseudorandom order in the corresponding, and immediately following, test phase. In the test phase, 4 of the 16 test trials were assigned to each of 4 experimental conditions (See Figure 2.3):

1. **Exact match**: test object exactly matched the object paired with the scene at study
2. **Distortion**: test object was a different exemplar of the same subtype of object presented with the scene at study
3. **Within-category mismatch**: test object was a different subtype of object within the same major category as the object presented with the scene at study
4. **Between-category mismatch**: test object was a different major category of object from that presented with the scene at study (and thus, belonged to a category of object that would never have been studied with that scene category previously; e.g., an object category that had only ever appeared on offices at study, now appearing on a beach at test).

To illustrate, a typical study phase might contain scene-object (S-O) pairings as follows: S(category = 1, exemplar = a)-O(major category = 1, subtype = 1, exemplar = a), S(2,a)-O(2,1,a), S(3,a)-O(3,1,a), S(3,b)-O(3,2,a), S(3,c)-O(3,3,a), etc. For the corresponding test phase, a match trial might be S(1,a)-O(1,1,a), a distortion trial might be S(1,a)-O(1,1,b), a within category mismatch trial might be S(1,a)-O(1,2,b), and a between category mismatch trial might be S(1,a)-O(2,1,b). Because there were only 16 trials per study phase, only 16 of 18 object categories (6 major categories x 3 subtypes / major category) were presented in each study phase. Half of the between category mismatches in the test phase were created by swapping two object subcategories from different major categories that had been presented in the preceding study phase (e.g. study: S(1,a)-O(1,1,a), S(2,a)-O(2,1,a); test: S(1,a)-O(2,1,b), S(2,a)-O(1,1,b)). The other half were created by introducing an object category that had not been presented in the preceding study phase (e.g. study: S(1,a)-O(1,1,a), never present object in category (2,1) that block; test: S(1,a)-O(2,1,a)). Within category mismatches were always created by swapping object subcategories within a major category that had been presented in the preceding study phase (e.g. study: S(1,a)-O(1,1,a), S(1,b)-O(1,2,a); test: S(1,a)-O(1,2,b), S(1,b)-O(1,1,b)). Other

than in the exact match condition, the exact object exemplar presented at test never matched that presented in the preceding study phase.

Each list contained 72 trials per condition. Across lists, each test object and each scene appeared in each of the four experimental conditions an equal number of times. In fact, identical test object-scene pairs were used across the first 3 experimental conditions; only in the fourth experimental condition (between category mismatch) was it necessary to shuffle the specific pairings of test object and scene. Trial order within each block was pseudo-randomized, such that no more than two trials corresponding to each condition were presented in a row, and no more than 3 trials mapping on to a 'same' response (that is, trials in the exact match or distortion conditions) occurred in a row at test. Response hand was counterbalanced across lists.

Procedure

Participants passively studied the paired scenes and novel objects and then were tested by being asked to indicate, for each in a new set of pairs, whether the object matched the presented scene. Study and test phases were organized into 18 study-test blocks, between which the participant was encouraged to take a break. All breaks were self-paced.

In each study phase, 16 scene-object pairs were presented, each beginning with a white fixation cross on a black background presented for 350-550 ms (duration jittered to reduce the impact of anticipatory slow potentials on the timelocked waveform). Next, the scene alone was presented centrally for 2500 ms on a black background. Right after the scene appeared, participants were allowed to move their eyes to take in the scene; however, 1800 ms into scene presentation, the fixation cross brightened to indicate that the participant should fixate in the center of the screen in preparation for object presentation. 700 ms later (2500 ms after scene onset), a white square containing the object appeared in the center of the screen, super-imposed on top of the scene, for 2500 ms. A screen with the word "***BLINK***" was then displayed for 2000 ms (preceded and followed by 50 ms of blank screen), in order to encourage the participant to blink between trials.

In the test phase immediately following each study phase, 16 scene-object pairs, repeating all 16 scenes from the study phase, were displayed. Participants were asked not to move their eyes for the entire test trial duration. Similar to the study trials, each scene was first

displayed by itself for 2500 ms on a black background, followed by 2500 ms during which the test object was displayed centrally on top of the scene, again embedded in a white square. Participants were asked to wait until the object-scene pair was replaced by a question mark in the center of the screen to respond; the question mark remained on screen until a response was made. Participants had 3 response options: the object on this scene is (1) the same object that I studied with this scene, (2) not the same object, but "could have gone with" this scene, (3) not the same object and could not have gone with this scene. Participants were told that if the object was only slightly visually distinct from what they remembered studying (e.g. had a different body position or proportions), they should still respond (1). They were also told that an object and scene 'could go together' if they believed that pair looked similar to other study items and could hypothetically be presented in an upcoming trial of the experiment, even if they knew that they hadn't studied it. Participants were never told that there was a structured relationship among the object and scene categories. Participants were explicitly instructed that each test phase only covered materials studied in the immediately preceding study phase and that testing was non-cumulative across blocks. Prior to the main experiment, participants were given a practice block of 4 study and 4 test trials, which used different but qualitatively similar object and scene images to those in the main study.

During recording, participants were seated in a comfortable chair at a viewing distance of approximately 100 cm from the computer display. The visual angle of the scenes was 13.6° by 9.3° and that of the object images was 4.6° by 4.2°. The recording session lasted approximately 90 minutes. Afterwards, participants filled out an exit survey about the strategies they had used to complete the task, including several open response questions. In order to determine how participants were using the response scale, they were then given several visual examples of corresponding study and test scene-object pairs, and asked to indicate how they would respond to that test item, given that they remembered studying that study item. Lastly, for a subset of the object images (one from each of the 6 major categories of germs and machines), participants indicated which scene categories the object had been associated with during the experiment by circling one or more of 6 scene category labels.

EEG Data Acquisition and Preprocessing

The electroencephalogram (EEG) was recorded from 26 silver/silver-chloride electrodes evenly spaced over the scalp. The sites are midline prefrontal (MiPf), left and right medial prefrontal (LMPf and RMPf), left and right lateral prefrontal (LLPf and RLPf), left and right medial frontal (LMFr and RMFr), left and right mediolateral frontal (LDFr and RDFr), left and right lateral frontal (LLFr and RLFr), midline central (MiCe), left and right medial central (LMCe and RMCe), left and right mediolateral central (LDCe and RDCe), midline parietal (MiPa), left and right mediolateral parietal (LDPa and RDPa), left and right lateral temporal (LLTe and RLTe), midline occipital (MiOc), left and right medial occipital (LMOc and RMOc), and left and right lateral occipital (LLOc and RLOc). The midline central (MiCe) electrode was placed where the "Cz" electrode would appear using the international 10-20 system. Eye movements were monitored via a bipolar montage of electrodes on the outer canthus of each eye. Blinks were detected by an electrode below the left eye. Impedances were kept below 5 KΩ. Signals were amplified with a .02–250 Hz bandpass using a BrainVision amplifier and digitized at 1000 Hz. Data were referenced online to the left mastoid and rereferenced offline to the average of the left and right mastoids. Each trial consisted of a 1000 ms epoch preceded by a 200 ms prestimulus baseline. Trials contaminated by eye movements, blinks or other recording artifacts were rejected offline. Artifact rejection procedures using subject-specific threshold parameters resulted in average trial loss of 11.3% for the exact match condition, 11.4% for the distortion condition, 12.6% for the within-category mismatch condition, and 11.8% for the between category mismatch condition. A digital lowpass Butterworth IIR filter with a 30 Hz half-amplitude cut-off and 12 dB/octave roll-off was applied prior to statistical analysis. Prior to permutation-based cluster analysis, data were further down-sampled to 100 Hz.

Analysis

Behavioral analyses of response distributions were conducted using logistic regression modeling in R. Statistical analyses of individual trial EEG data were conducted using mixed effects models built with the lme4 package in R. Models initially included crossed random effects of subject and item, and by-subjects random slopes of each predictor of interest, but the random effects structure was sometimes scaled back to address convergence issues. EEG dependent measures for analysis (mean amplitudes over particular stretches of time and space post test object onset)

were determined in two ways: 1. a priori windows in time and space determined based on the prior literature and 2. a data-driven method, permutation-based cluster analysis. For component-based analyses, we used time windows and electrode selections based on the prior literature: 250-349 ms over fronto-central sites to capture the N300, 350-499 ms over centro-parietal sites for the N400, and 500-699 ms and 700-899 ms over posterior sites to capture early and late time windows of the Late Positive Complex (LPC). Because the N300 was the primary measure of interest, and because its distribution has been variably characterized over the literature, we took the conservative approach of measuring effects at all fronto-central sites (16 total); the N400 and LPC were characterized at 8 sites each, focused around each component's typical distribution and reducing topographic overlap with other components. Key predictors of EEG amplitude included match condition and visual similarity to the target object.

Results

Behavioral – Online Accuracy

Participants discriminated well among test objects that matched and mismatched the presented scene. As instructed, they tended to respond that both 'exact match' and 'distortion' condition objects matched the scene (response option 1 = 'match'), and that 'within-category mismatch' and 'between-category mismatch' condition objects did not (response options 2 = 'possible mismatch' and 3 = 'impossible mismatch'). Accuracy of appreciating the match between object and scene was computed after collapsing the exact match and distortion conditions (which were treated as a 'match') and the two mismatch conditions along with the two different mismatch responses ('possible' and 'impossible'). Mean accuracy was 81.5%, range 64.2-97.9%. Participants were also sensitive to the type of mismatch, and were more likely to respond that the test object could not have gone with the test scene ('impossible' response) for the between-category mismatch condition than the within-category mismatch condition. This was characterized using a logistic regression model predicting the probability of response 'impossible mismatch' with fixed effect of condition (exact match or distortion vs. within category mismatch vs. between category mismatch; the model failed to converge when the exact match and distortion conditions were treated as separate predictors). The model included subject random

intercepts and by-subjects random effects of condition. Nested model comparisons confirmed that 'impossible' responses were more likely for the between-category mismatch than the within-category mismatch condition: intercept = -.635, β = 1.78, SE = .217, z = 8.208, χ_1^2 = 32.36, p < .001. Figure 2.4a shows the mean response distribution across subjects for each condition.

Behavioral – Posttest Categorization

By the end of the experiment, participants demonstrated explicit knowledge of the categorical mapping among objects and scenes. Participants were more likely to indicate that objects were associated with a scene category when the two had been paired at study. Figure 2.4b shows the normalized confusion matrix indicating the probability that a scene category, if circled, belonged to the correct scene category for the depicted object. This was assessed with a logistic regression model predicting the probability of circling a scene category. The model included a fixed effect of match (between the scene and object), crossed random intercepts for subject, object (response item) and scene (response choice), and a by-subjects random effect of match. Nested model comparisons were used to confirm that scenes matching the object were more likely to be circled: intercept = -4.406, β = 7.814, SE = .918, z = 8.513, χ_1^2 = 40.13, p < .001.

ERP Analysis: Match Condition

Our component-based match condition analysis assessed whether scenes induced N300 visual form priming of associated objects in the test phase, and also examined semantic (N400) and decision-related (LPC) processing. In an analysis of the trial-by-trial ERP response to object images in the test phase, we targeted the temporal and topographic distribution of the N300, N400, and LPC (using an apriori split of the 400 ms LPC window into early and late parts, 200 ms each, keeping window size more comparable across the analyses), and examined differences across conditions (see Figure 2.5 for component timing and scalp distributions). Only behaviorally correct (response option 1 = 'match' for match and distortion and response options 2 = 'possible mismatch' or 3 = 'impossible mismatch' for within and between category mismatches) and artifact free trials were included. Linear mixed effects models were fit to the individual trial data, including fixed effects of:

(1) condition (contrasting match, distortion, within, and between category conditions)
(2) response type (for within and between category mismatch conditions only, since only one response type was considered correct for match and distortion conditions); it should be noted that condition (within vs. between) was moderately associated with response ('possible' vs. 'impossible' mismatch); Cramér's V = .27.
(3) the interaction between condition and response

Models also included crossed random intercepts of subject, item (object+scene), and channel, and by subjects random slopes of condition, response and their interaction.

Because some of the between category mismatches were also newly presented within the study-test block (whereas all other item types were within-block repetitions), we also examined the role of item recency in differentiating the between category mismatches from the other conditions. To do this, we compared our model above to one with an additional fixed effect, which contrasted condition 4 swap trials (generated by swapping object images presented in the immediately preceding study phase) and condition 4 new trials (generated by presenting an object image that was not presented in the immediately preceding study phase).

All models were fit using maximum likelihood estimation. Fixed effects were initially tested using likelihood ratio tests with nested model comparisons. Follow up comparisons of condition means were conducted using the contest function in the lmerTest package in R, with family-wise error rate corrected p-values and the Satterthwaite approximation for the degrees of freedom; except where otherwise indicated, condition contrasts collapse across response type using linear combinations of beta weights but were conducted on a model that included a fixed effect of response type. The distortion condition was included in all models, but never differed reliably from the match condition. For ease of reporting, we thus describe only contrasts among the match condition and the two violation types. 95% confidence intervals on beta weights and contrasts were computed using the bootMer function in the lme4 package in R (N = 2000 iterations).

N300

To examine the N300, voltages were separately averaged across time for each trial from 250-349 ms at each of 16 frontal and central sites. There was a main effect of condition (χ^2_3 =14.91, p < .01), which persisted when recency was accounted for by including between category mismatch swap versus new as a fixed effect (χ^2_3 =13.79, p < .01; including this effect did not substantially improve model fit, χ^2_1 < 1). Follow-up comparisons revealed that between category mismatches were more negative than matches (diff = -1.79 μV, 95% CI = [-2.74, -.83], $F_{(1,25.5)}$ = 13.6, p < .01). There was also a tendency for between category mismatches to be numerically more negative than within category mismatches (diff = -1.08 μV, 95% CI = [-2.25, 0.15], $F_{(1,21.8)}$ = 3.19, p < .1). Within category mismatches did not reliably differ from matches (diff = -.70 μV, $F_{(1,20.2)}$ = 1.49, p > .1). The main effect of response type and its interaction with condition for within and between category mismatches were not statistically reliable (|t|'s < 1).

N400

To examine the N400, voltages were separately averaged across time for each trial from 350-499 ms at 8 central and parietal sites. There was a main effect of condition (χ^2_3 =14.05, p < .01), which remained as a numeric trend when recency was accounted for by including between category mismatch swap versus new as a fixed effect (χ^2_3 =7.80, p < .1; including this additional factor improved model fit, χ^2_1 = 4.22, p < .05). Follow-up comparisons revealed that between category mismatches were more negative than all other conditions (match: diff = -1.63 μV, 95% CI = [-2.59, -.68], $F_{(1,24.8)}$ = 11.3, p < .01; within category mismatch: diff = -1.33 μV, 95% CI = [-2.53, -.15], $F_{(1,20.2)}$ = 4.90, p < .05). When swapped was compared with new between category violations, new items were found to be more negative (diff = -.79 μV, 95% CI = [-.03, -1.55], $F_{(1,2546)}$ = 4.07, p < .05). Both new and swapped between category violations were more negative than matches (swap – match: diff = -1.23 μV, 95% CI = [-2.33, -.22], $F_{(1,33.5)}$ = 5.45, p < .05; new – match: diff = -2.02 μV, 95% CI = [-3.09, -1.00], $F_{(1,33.0)}$ = 15.4, p < .001). However, only new between category violations were reliably more negative than within category violations (diff = -1.73 μV, 95% CI = [-2.97, -.56], $F_{(1,24.1)}$ = 7.42, p < .05). Within category violations did not reliably differ from matches (diff = -.29 μ V, $F_{(1,19.8)}$ < 1). For the two violation types, the main effect of response type and its interaction with condition were not statistically reliable (|t|'s < 1).

LPC

To examine the LPC, voltages were separately averaged across time for each trial from 500-699 ms (early LPC) and 700-899 ms (late LPC) at each of 8 posterior sites. In the early time window there was no reliable effect of match condition (χ_3^2 =2.60, p > .1). However, there was a reliable effect of response type (possible vs. impossible mismatch) for within and between category violations, such that trials indicated as being a possible mismatch were less positive (β = -1.42 (μV), 95% CI = [-2.60, -.06], χ_1^2 = 4.62, p < .05). There was no reliable interaction between response type and condition (|t| < 1).

In the late time window, there was a main effect of condition (χ_3^2 =13.49, p < .01), which persisted when recency of the between category mismatches was accounted for (χ_3^2 = 14.60, p < .01; including swapped versus new as a factor did not substantially improve model fit, χ_1^2 = 1.12, p > .1). Between category mismatches were more positive than matches (diff = 1.32 μV, 95% CI = [.20, 2.48], $F_{(1,24.1)}$ = 5.30, p < .05) and within category mismatches (diff = 1.47 μV, 95% CI = [.19, 2.82], $F_{(1,21.3)}$ = 4.96, p < .05). There continued to be a reliable effect of response type (possible vs. impossible mismatch), with possible mismatch trials less positive (β = -2.88 (μV), 95% CI = [-4.09, -1.67], χ_1^2 = 17.97, p < .001). There was no reliable interaction between response type and condition (|t| < 1).

Summary

Our component-based analysis revealed N300 priming of contextually congruent objects. Match effects were first observed in the N300 window, with increased negativity for between category mismatches relative to matches and distortions, and within category violations falling numerically in between. This pattern of effects was also seen on the N400, and the N400 additionally showed sensitivity to recency, with new items more negative than recently seen ones (replicating many prior studies). Finally, the LPC was sensitive to participants' judgments, being more positive to trials judged to be impossible mismatches, and, in the later part of the window, to between category violations overall.

Although there were *apriori* reasons to think that the N300, N400, and LPC might show effects of the experimental manipulations (cf. Hannula, Federmeier, & Cohen, 2006), we also

were interested in characterizing the pattern obtained when no apriori choices were made about either time window or scalp channels for analysis. We examined this with an exploratory cluster analysis of match condition for behaviorally correct test object trials, conducted in the time domain using the ft_timelockstatistics function in fieldtrip. Details of the analysis and results can be found in Appendix A. Importantly, the results of the analysis converged with the component-based approach, showing effects differentiating match trials from mismatch trials beginning around 230-250 ms over the front of the head (like the N300), becoming more broadly distributed, and ending, around 400-430 ms over central/posterior sites (like the N400). Differences between the two mismatch trials types were also found from 760-990 ms over the back of the head (like the LPC).

ERP Analysis: Target Similarity

We also conducted a component-neutral analysis to confirm that visual form information about the contextually congruent object is brought online in response to the context scene. In doing so, we can make a more direct inference about the role of scene-induced visual information in object processing in this experimental context. We assessed whether low-level visual features of the scene-congruent object were accessed in memory even when it was not displayed. In focusing on the N300, our component-based analyses targeted what are typically thought to be intermediate stages of visual form analysis (Schendan & Ganis, 2015). Here, however, because we are not targeting the sensitivity of any particular waveform feature, we examined low-level visual feature similarity, providing a conservative test of visual form reactivation and taking advantage of the fact that the modeling of such low-level features is well-established (Pinto, Cox, & DiCarlo, 2008). Thus we used V1-like low-level visual features to derive a similarity metric to predict ERP amplitude. We adapted our match condition analysis models, adding a new fixed effect: visual distance from the target object. We then tested whether this predictor improved model fit using nested model comparisons. Models did not distinguish among the two types of between category violations (new vs. swap). Including a by-subjects random slope for visual distance from the target object resulted in a failure to converge, so the same random effects structure was used as for modeling the effects of match condition.

Visual distance was computed as follows. First, V1-like features were generated for each object image using the model in Pinto, Cox, and DiCarlo (2008). Next, features with variance close to zero were removed to avoid numeric issues while scaling, and feature values were mean-centered and scaled to unit variance. Next, PCA was applied to reduce the dimensionality of the feature space to 766 (from >80,000) while maintaining an explained variance ratio of .999. Visual distance was defined as the Euclidean distance between the presented object image and the target (studied) object image in this feature space, for each trial. Visual distance was grand mean centered prior to model fitting.

Including visual distance to the target object as a continuous predictor substantially improved model fit for all components (N300: χ_1^2 = 67.0, p < .001, β= -3.615x10^{-3}, 95% CI = [-4.45, -2.71]x10^{-3}; N400: χ_1^2 = 88.3, p < .001, β= -5.612x10^{-3}, 95% CI = [-6.74, -4.45]x10^{-3}; LPC-early: χ_1^2 = 31.5, p < .001, β= -3.211x10^{-3}, 95% CI = [-4.32, -2.09]x10^{-3}; LPC-late: χ_1^2 = 56.6, p < .001, β= -4.464x10^{-3}, 95% CI = [-5.64, -3.31]x10^{-3}). Negative beta values indicate that for all components, the more visually similar the displayed object was to the target object, the more positive the waveform. The distribution of visual distances across trials, broken down by condition, and including only behaviorally correct and artifact-free trials, is displayed in Figure 2.7. Table 2.1 lists the mean number of trials per subject for each condition and visual distance bin in Figure 2.7. The ERP waveforms corresponding to these trials, broken down by visual distance bin, are displayed in Figure 2.8, averaging across match condition.

Visual distance effects within component time windows and conditions

We also examined whether the effect of visual distance to target was modulated by category-level information. Importantly, this allowed us to observe whether visual similarity effects were apparent even within the distortion condition, suggesting a purer effect of visual similarity per se that is not contingent on category boundaries. Figure 2.9 shows estimated linear trends of visual distance by component and match condition, collapsing across response. We tested for interactions between visual distance to target, condition, and response, within behaviorally correct trials. The three-way interaction among match condition, response, and (mean centered) visual distance to target, as well as all lower order interactions, were added as fixed effects to the original visual distance to target models, while maintaining an identical random effects structure.

Interactions were tested using nested model comparisons. The effect of visual distance to target is reported separately for each condition and response, including 95% bootstrapped confidence intervals (N = 2000 iterations). All point estimates and confidence intervals on linear combinations of beta weights are reported below as 1000 times the original estimates.

N300

There were significant interactions of condition x visual distance (χ_2^2 = 27.59, p < .001), and condition x response x visual distance (χ_2^2 = 15.66, p < .001). Visual distance to target effects were significant in the distortion condition, and on within category violation trials that were responded to as 'possible' mismatches, but not on other mismatch trials (distortion: -5.105 [-6.119, -4.077]; 'possible' within category mismatch: -3.903 [-5.593, -2.220]; 'impossible' within category mismatch: .680 [-1.095, 2.384]; 'possible' between category mismatch: 4.068 [-2.462, 11.085]; 'impossible' between category mismatch: -1.169 [-4.960, 2.452]). Within category mismatches showed larger effects of visual distance to target when they were responded to as being 'possible' (see non-overlapping 95% CIs).

N400

There were significant interactions of condition x visual distance (χ_2^2 = 10.64, p < .01), and condition x response x visual distance (χ_2^2 = 17.39, p < .001). The effect of visual distance was significant at α = .05 uncorrected for all condition by response combinations, except for between category mismatches that were responded to as 'possible' (distortion: -4.903 [-6.233, -3.555]; 'possible' within category mismatch: -10.421 [-12.689, -8.205]; 'impossible' within category mismatch: -4.515 [-7.132, -2.205]; 'possible' between category mismatch: 5.804 [-2.598, 14.189]; 'impossible' between category mismatch: -4.883 [-9.290, -.253]). Within category mismatches showed larger effects of visual distance to target when they were responded to as being 'possible'; 'possible' within category mismatches also showed larger visual distance effects than the distortion condition (see non-overlapping 95% CIs).

LPC (Early)

There was a significant interaction of condition x visual distance (χ_2^2 = 11.12, p < .01), and a numeric trend towards an interaction of condition x response x visual distance (χ_2^2 = 5.77, p < .1). The distortion and within category mismatch conditions showed reliable effects of visual distance to target, but not the between category mismatch conditions (distortion: -2.250 [-3.566, -.823]; 'possible' within category mismatch: -7.001 [-9.195, -4.706]; 'impossible' within category mismatch: -3.316 [-5.694, -.917]; 'possible' between category mismatch: .940 [-6.978, 8.542]; 'impossible' between category mismatch: -2.489 [-6.568, 1.480]). 'Possible' within category mismatches showed larger target dissimilarity effects than the distortion condition (see non-overlapping CIs), and were numerically larger than 'impossible' within category mismatches.

LPC (Late)

There was a significant interaction of condition x visual distance (χ_2^2 = 15.92, p < .001), but the condition x response x visual distance interaction did not reach significance (χ_2^2 = 4.56, p = .103); we still report each condition by response contrast separately for consistency. Visual distance to target effects were significant in the distortion condition, and on within category violation trials that were responded to as 'possible' mismatches, but not on other mismatch trials (distortion: -6.077 [-7.434, -4.633]; 'possible' within category mismatch: -4.119 [-6.358, -1.789]; 'impossible' within category mismatch: -.889 [-3.287, 1.514]; 'possible' between category mismatch: 2.039 [-6.135, 10.464]; 'impossible' between category mismatch: -2.215 [-6.453, 2.281]).

Summary: Visual distance effects within component time windows and conditions

In summary, we found that visual distance to target effects are generally apparent in both the distortion and within category mismatch conditions, but more often did not reach significance in the between category mismatch conditions, despite trending in the same numeric direction. We also found that within category mismatches generally showed larger effects of visual distance to match when they were responded to as being 'possible' (vs. 'impossible') scene-object pairs, particularly on the N300 and N400 components. On the N400 and early LPC components,

'possible' within category mismatches showed even larger effects of visual distance than the distortion condition.

Additional Control Analyses: Object Prototypicality

We also tested whether visual distance to target effects could be explained by a partial confound with object prototypicality. Objects that were closer in visual distance to the target also tended to be closer to their own category prototype (r = .155). We found that visual distance to target effects were more robust than and could not be explained away by (cue-independent) prototypicality effects, although both effects tended in the same numeric direction (see Appendix C for detailed results). Finally, we tested whether the presented object is visually compared to the target object prototype, vs. the target object exemplar, for each ERP component (see Appendix D). We found that exemplar-specific visual information about the target object is brought online and compared to the presented object, as evidenced by sensitivity to the degree of mismatch between the current object and the target exemplar, even when mismatch between the current object and the target prototype is accounted for, during the N300, N400 and LPC time windows.

Discussion

Although prior research has indicated that scenes can prime the visual form of associated objects (Bar, 2004; Brandman & Peelen, 2017; Truman & Mudrik, 2018), most studies have relied on natural statistical associations that may be learned over many years. We tested whether scene-object visual form priming extends to recently learned scene-object associations (< 2 hours), using a set of categorically organized novel objects in an explicit paired association memory task. We examined two EEG measures of form-based priming: the N300 (comparing match and mismatch trials in the test phase) and target similarity effects (within mismatch trials, regressing a continuous measure of visual similarity to the contextually associated, but not presented, object). In both cases, results suggest that scenes can indeed prime the visual form of even recently associated objects.

Behaviorally, participants were able to successfully associate novel objects and scenes, and in a later offline post-test they demonstrated explicit knowledge of the higher category-level pairing between object types and scene types (which persisted throughout the experiment but which they were never explicitly informed of). A priori component-based and data-driven cluster-based analyses converged to reveal an N300 facilitation (reduced negativity beginning around 200-250 ms, with a frontal scalp distribution) to objects matching their associated scenes, relative to mismatches, in the test phase. Both matching objects and close distortions elicited a more positive waveform than objects that mismatched the scene type. Objects that mismatched the specific scene, but which were congruent with its higher-level category (beach, mountain, etc.), showed an intermediate level of facilitation in the N300 time-window, particularly at central sites. This may reflect overlapping generators with the subsequent N400, which also showed more facilitation for within- than between-category mismatches, and which is generally known to differentiate near and distant semantic violations (Federmeier & Kutas, 1999, 2001). Although the N300 and N400 effects we observe overlap, the early part of the effect is temporally and topologically more aligned with an N300 than an N400. Given prior work linking the N300 specifically to visual form-based priming (Hamm, Johnson, & Kirk, 2002; Kovalenko, Chaumon, & Busch, 2012), these results suggest that scenes may enhance accessibility of the visual form of even recently associated objects, at least when heightened accessibility is useful to the task at hand.

In a second set of analyses, we focused on distortion and mismatch trials. We revealed a novel effect: an enduring sensitivity across the ERP waveform to (low-level) visual similarity to the target (contextually congruent) item, onsetting at roughly 200 ms. The more visually similar the presented object was to the contextually congruent object, the more positive the waveform, across the N300, N400, and LPC time windows. This effect was not simply driven by degree of match at the category-level: it was apparent on an item-level basis within the distortion condition, holding category identity constant. Attesting to the effect's reliability, it was separately observed in the within-category mismatch condition, and there was a numeric trend in the same direction for between-category mismatches. Given the extended timing of the effect, we believe it at least partially reflects a template matching process (Mostert, Kok, & De Lange, 2015; Kok, Failing, & de Lange, 2014; Kok, Mostert, & De Lange, 2017; Summerfield et al., 2006; Summerfield & De Lange, 2014), in which the current object is visually compared to a

memory template of the target object, evoked by the context scene. We differentiate "template matching," which may reflect a task-specific perceptual decision-making process in which a single memory representation (that of the template) is prioritized for comparison, from the generic perceptual matching processes that occur in order for an object to be recognized. Although early sensitivity to target similarity may partly reflect visual form priming itself, later sensitivity is more likely to reflect the formation of a perceptual judgment (Mostert, Kok, & De Lange, 2015). Interestingly, in the LPC time window at central and posterior sites, target similarity effects had the opposite polarity of differences due to mismatch type (that is, more severe category violations were more positive). This suggests distinct temporally overlapping processes of visual comparison and rule-based decision-making, in keeping with findings that the LPC is sensitive to both perceptual analysis and decision-related processing (e.g. Falkenstein, Hohnsbein, & Hoormann, 1994; Schendan & Kutas, 2002, 2003). Regardless of whether early sensitivity to target similarity is ultimately better explained as direct visual form priming of the current object, or as an index of decision-related processing, it corroborates early availability of the memory template of the target object, making visual form priming at 200-350 ms of the presented object more plausible given that sensitivity to even low-level visual feature similarity is apparent at the same time.

Furthermore, comparing effects of visual similarity to the target exemplar vs. the target prototype speaks to the types of strategies participants may have used to complete our task. Hypothetically, participants could have remembered only the abstract/amodal conceptual object category that was paired with the scene and simply assessed whether the presented object belonged to the appropriate category. Thus, they could have brought online only a coarse visual representation of the target category (e.g. a prototype schematic) without finer visual details of the target object exemplar. However, sensitivity to target similarity goes beyond what would be expected if each image were compared to a prototype; rather, exemplar-specific details of the target image are compared to the current image, beginning as early as the N300 time window. Even when it was not necessary for the task at hand, participants brought online a fine-grained memory representation of the contextually associated target that included exemplar-level visual information that could be dissociated from the target prototype using gabor filters. Although we only report visual distance effects using a single feature space, future work could compare multiple measures of target similarity (e.g. shape-based, frequency-based, abstract/semantic) to

more fully probe the nature of the target object memory representation brought online in response to scene contexts.

Our results also suggest that task-relevant category-level information may modulate the visual matching process or the extent to which it is engaged. The effect of visual similarity to target tended to be larger in magnitude for distortions and within-category mismatches than for between-category mismatches. Also, within-category mismatches that were reported as being incongruent but able to "go with the scene" elicited a stronger visual similarity effect than those reported as mismatching the scene type. Indeed, 'possible' within category mismatches showed even larger effects of visual distance than distortions on the N400 and early LPC. These interactions may be partially driven by changes in the range of visual distance to target across match conditions and responses. For example, matching processes may be the most evident when the presented image is distinct enough from the target for the difference to be detectable, but not so different that there is little representational overlap. Nonetheless, given that within and between category mismatches had largely overlapping distributions of visual distance in our experiment, it may also be that 'impossible' mismatches were more likely to be rejected from consideration in a top-down / rule-based fashion. To the extent that scene-congruent visual objects are more likely than scene-incongruent objects to be matched against a memory template for an expected scene-congruent object, this corroborates aspects of Bar's (2004) theory of object recognition, which postulates that scenes constrain the object categories that are considered for assignment to a visual stimulus. Within this framework, each scene type in our experiment served as a 'context frame' associated with a subset of the object stimuli. When the presented object matched the context frame, the participant considered whether it might also belong to the target object category. However, when the presented object mismatched the context frame, this matching process was engaged to a lesser extent. Future work should further explore whether the target similarity effects observed here are task-specific and if they can be modulated by additional factors known to affect anticipatory visual processing.

Taken together, our results corroborate the hypothesis that recently formed arbitrary associations between contextual cues and object representations can facilitate visual object recognition via visuo-structural priming (as dissociated from amodal semantic or decision boundary-based effects, which may co-occur). Although we used scene-novel object associations and an explicit memory task, similar contextual priming effects have been observed using other

types of visual sequences, even when contextual associations are unrelated to the task at hand. For example, Turk-Browne and colleagues had participants make orthogonal judgments to a continuous sequence of faces and places, and found that statistical regularities induce hippocampally encoded predictive cuing of upcoming stimuli, and facilitate visual object recognition as indexed by behavioral responding (Turk-Browne, Scholl, Johnson, & Chun, 2010). Kok and colleagues have similarly found expectation-based pre-activation of task-irrelevant visual features, using cross-modal cuing (Kok et al., 2017). Moreover, some theories suggest that top-down effects on perception, implicit statistical learning, and explicit paired associate learning may rely on partially overlapping mechanisms (Pearson & Westbrook, 2015). To the extent that statistical associations are rapidly implicitly learned and regularly used to facilitate processing of anticipated upcoming visual input, our findings may extend to visual object processing under the task demands of normal daily life. That is, recent episodic memories of the objects present in an environment may lead to visual form priming that facilitates object recognition when the environment is reinstated. In turn, populations with short-term memory impairments as the result of a disorder or as a function of normal aging may also experience disruptions in object recognition relative to healthy young adults. A growing body of literature linking the hippocampus and prefrontal cortex, areas particularly susceptible to damage and disruption (Anand & Dhikav, 2012; Baars & Gage, 2010; Fabiani, 2012), to visual prediction and mismatch detection, underscores this possibility (hippocampus: Chen, Olsen, Preston, Glover, & Wagner, 2011; Chen, Cook, & Wagner, 2015; Duncan, Ketz, Inati, & Davachi, 2012; Hindy, Ng, & Turk-Browne, 2016; Kok & Turk-Browne, 2018; frontal cortex: Bar et al., 2006; Summerfield et al., 2006; Summerfield & Koechlin, 2008).

In summary, neural evidence suggests that scenes can prime the visual form of even recently associated objects. Gabor-filter based visual features, similar to the empirically inferred neural representation in areas V1/V2, are reasonably well correlated with the memory representation of the object that is brought online in response to a context scene. Rapid statistical learning of object-scene associations could be exploited in future research to more carefully control for the strength of contextual associations between objects and scenes. Future work should focus on expanding the generalizability of our approach by bridging the gap between explicit paired associate learning and implicit statistical learning in the context of daily life. Also, the types of similarity analyses we used in the present study could be extended to further refine

our understanding of the nature of memory representations of visual objects, by comparing multiple measures of visual similarity. Lastly, clinical implications of the current results could be verified by comparing behavioral and neural indices of scene-object priming across the lifespan and in disordered populations.

Table

Table 2.1. Mean number of behaviorally correct and artifact-free trials per subject included for each condition and visual distance bin.

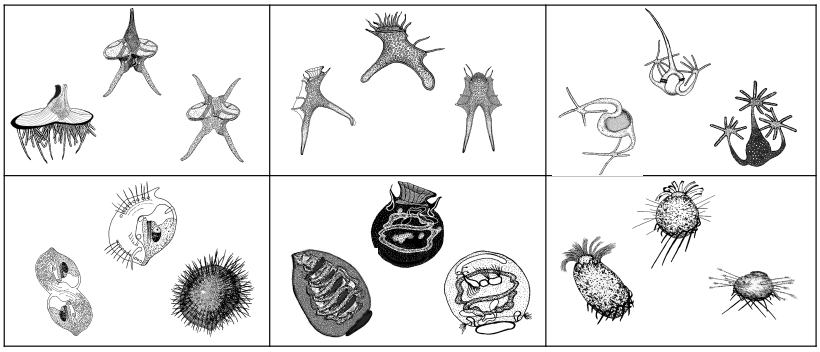
V1-Like Feature Euclidean Distance to Target	Match Condition: Match	Distortion	Within Category Mismatch	Between Category Mismatch
0 (Match)	48.6	0	0	0
(0,100]	0	4.7	0	0
(100,200]	0	17.3	0	0
(200,300]	0	15.5	9.6	7.4
(300,400]	0	5.7	24.9	27.5
(400,500]	0	2	9.2	18.1
>500 (Diff)	0	2.3	7.5	8.4

Figures

Figure 2.1. Scene and novel object categories. a) Examples of the 6 scene categories used in the experiment. b) Prototypical images for the 18 subcategories of germs, grouped into 6 major categories. c) Prototypical images for 18 subcategories of machines, grouped into 6 major categories.

a)

b)



c)

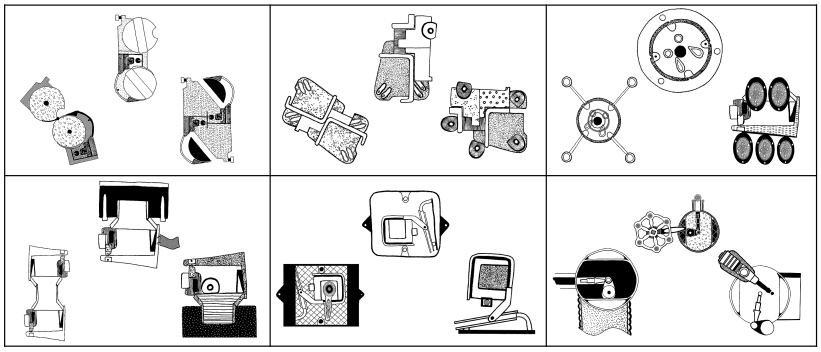


Figure 2.2. Illustration of how exemplars were derived from prototypical images for germs and machines. Germs were assumed to have flexible, and machines rigid, bodies.

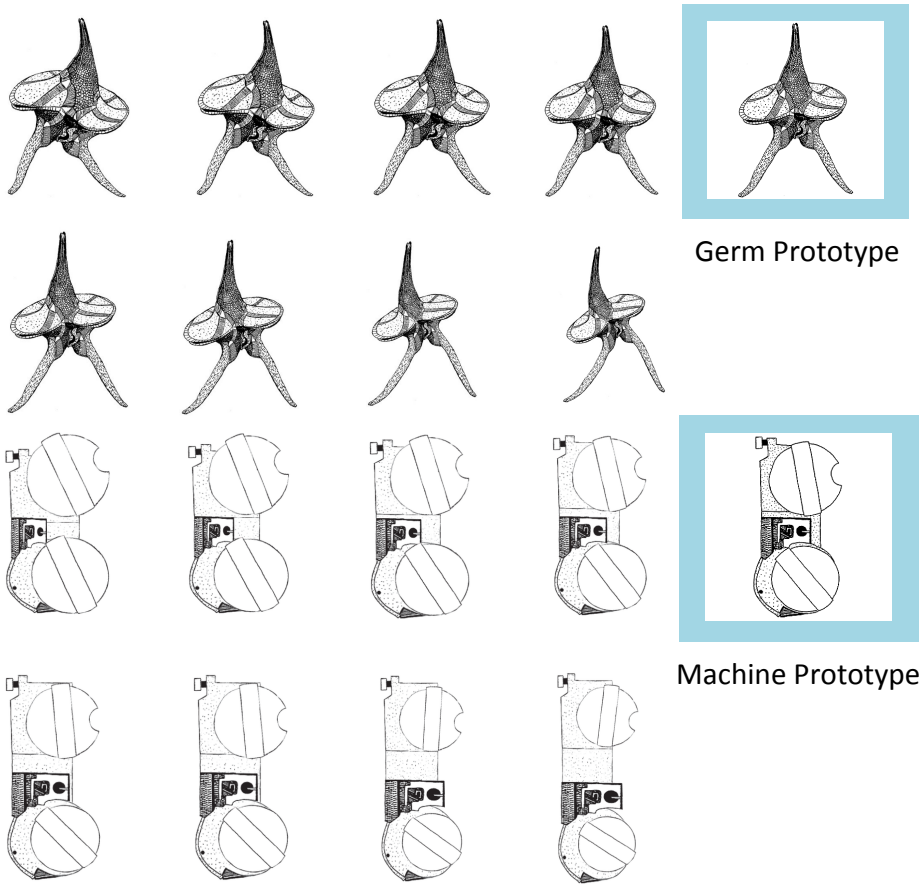


Figure 2.3. Experimental conditions.

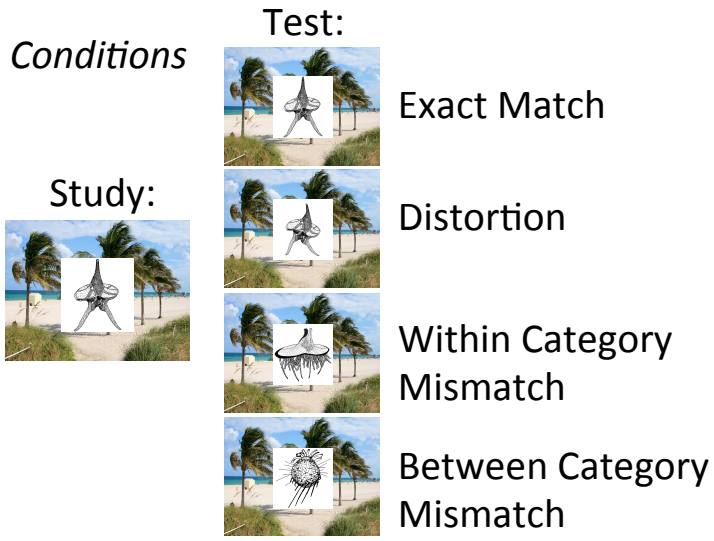
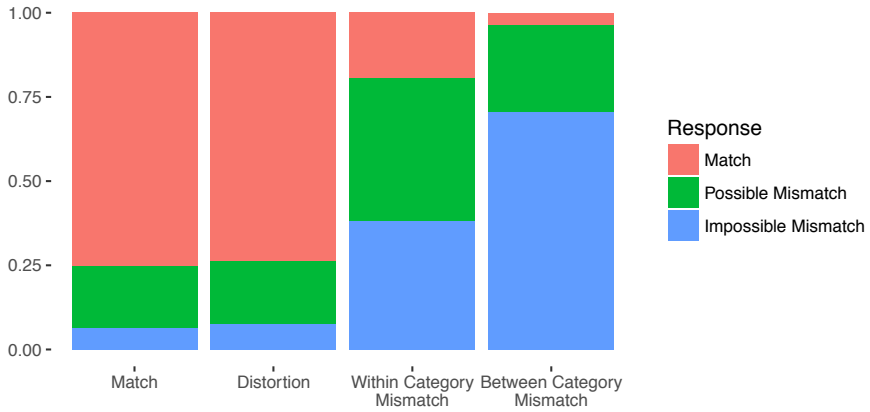


Figure 2.4. Behavioral results. a) Proportion of responses by condition in the online memory task. Participants were reliably sensitive to condition, and responded to distortions similarly to the exact match condition, as instructed. b) Confusion matrix of scene-object category associations indicated at post-test. Scenes circled by participants were the associated scene category for the displayed object 76-90% of the time, demonstrating explicit knowledge of the scene – object category mapping.

a)



b)

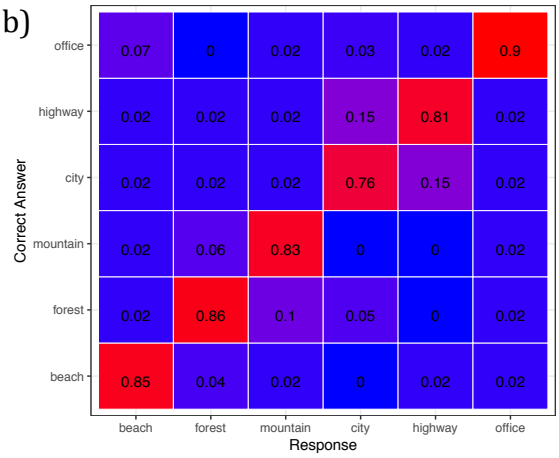


Figure 2.5. Timing and electrode sites used to characterize the components of interest.

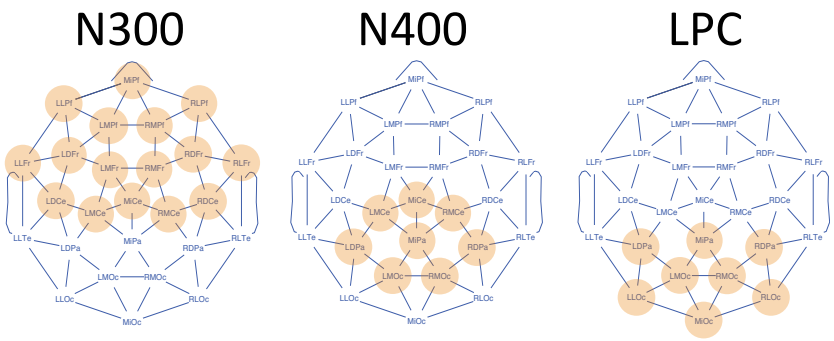


Figure 2.6. Match vs. mismatch conditions at 12 representative sites (scalp locations indicated at bottom right). An additional 15 Hz low pass filter was applied after averaging for display purposes.

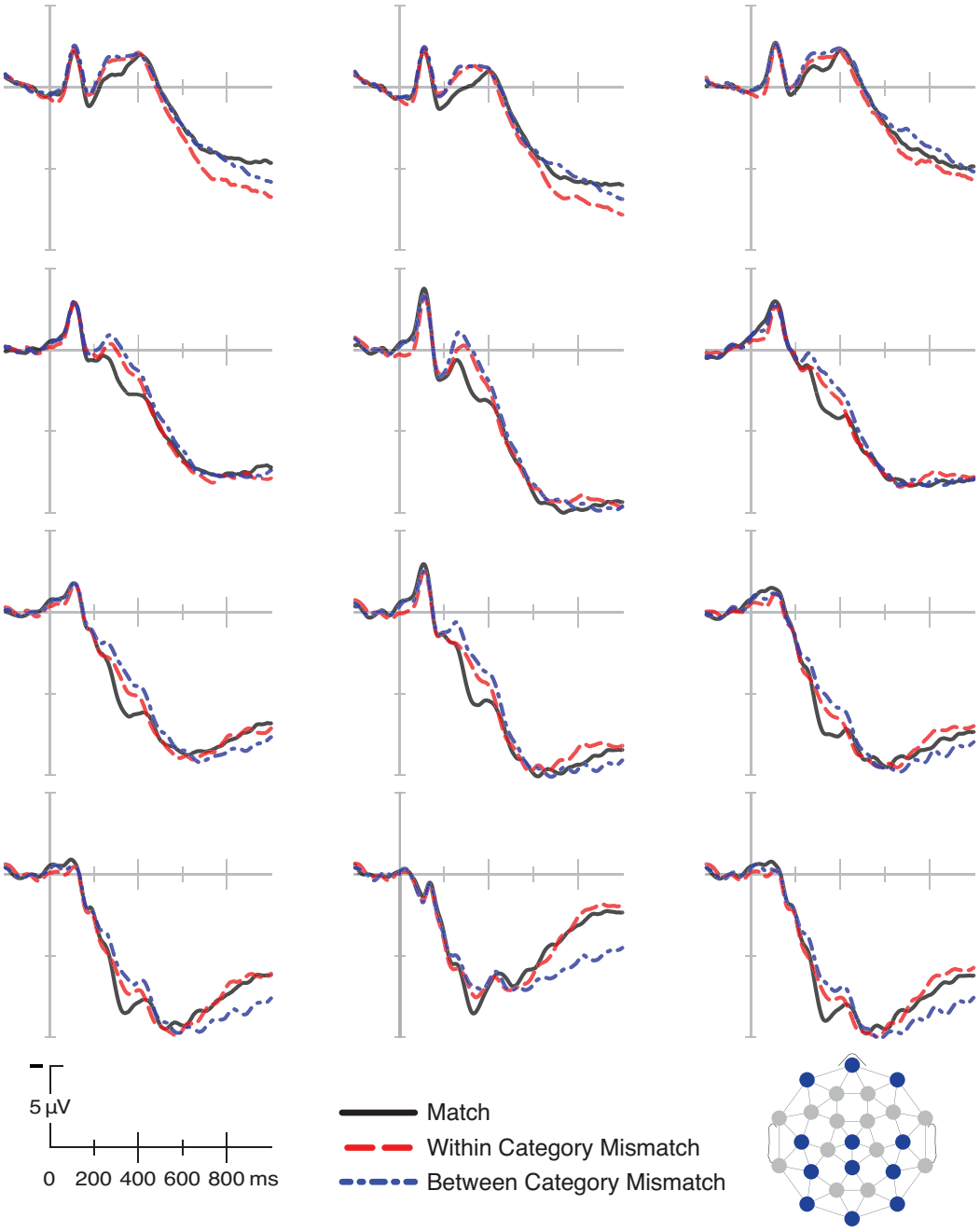


Figure 2.7. Distribution of Euclidean distance to the target object in the V1-like feature space across trials, plotted separately for each condition. Match condition is not shown because the visual distance was always exactly 0. Red lines indicate the demarcations used in Table 2.1.

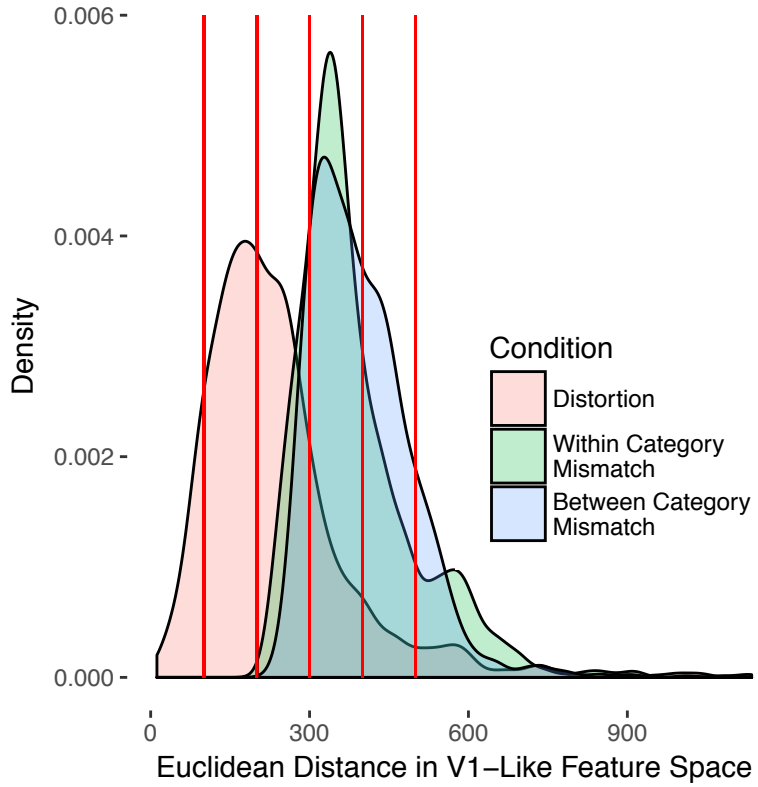


Figure 2.8. ERP averages by visual distance bin. Distance = 0 indicates the Exact Match condition, while higher distances indicate that the presented object was more visually distinct from the target object at test. Only behaviorally correct and artifact free trials included. An additional 5 Hz low pass filter was applied prior to plotting. Sites used to generate each plot are indicated at bottom left.

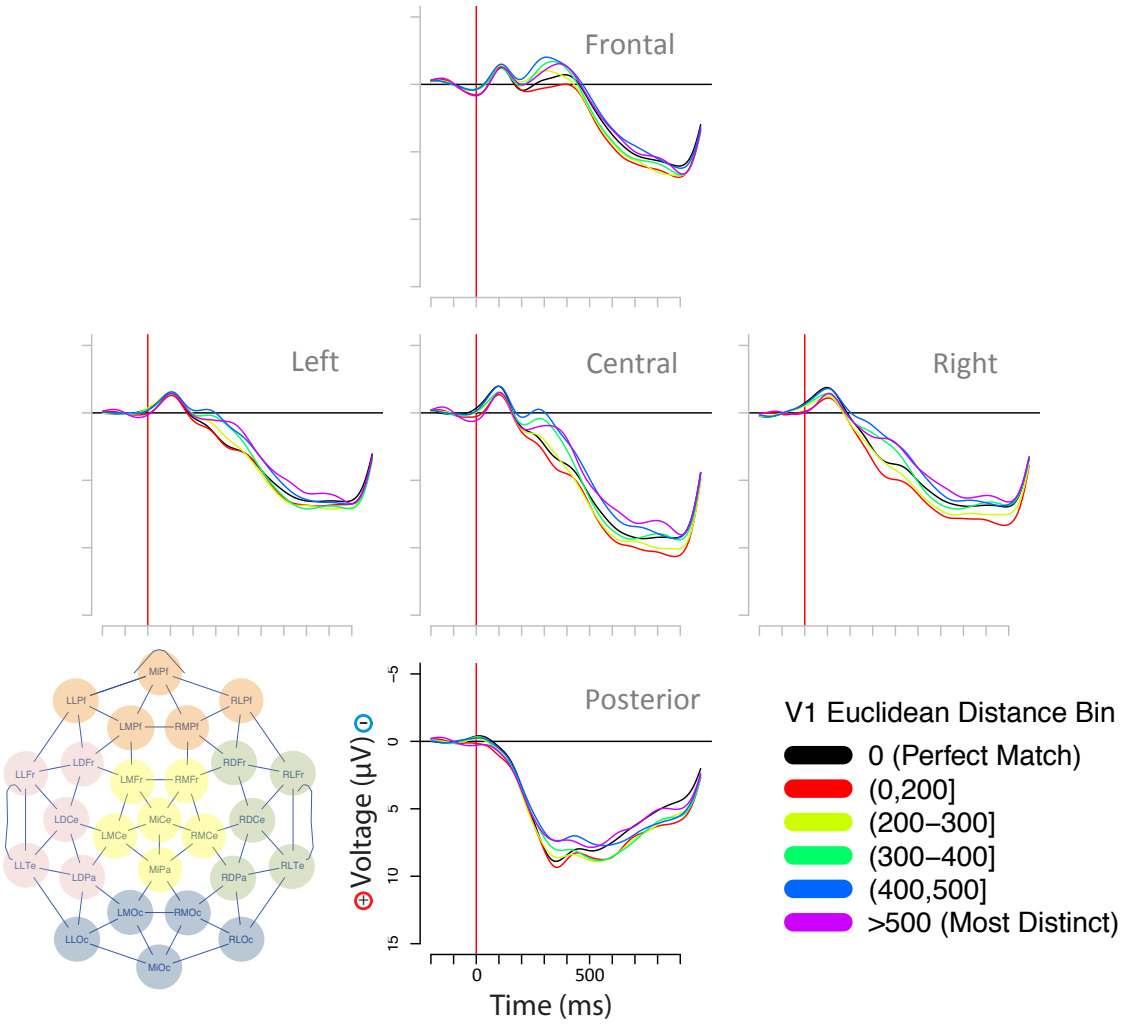
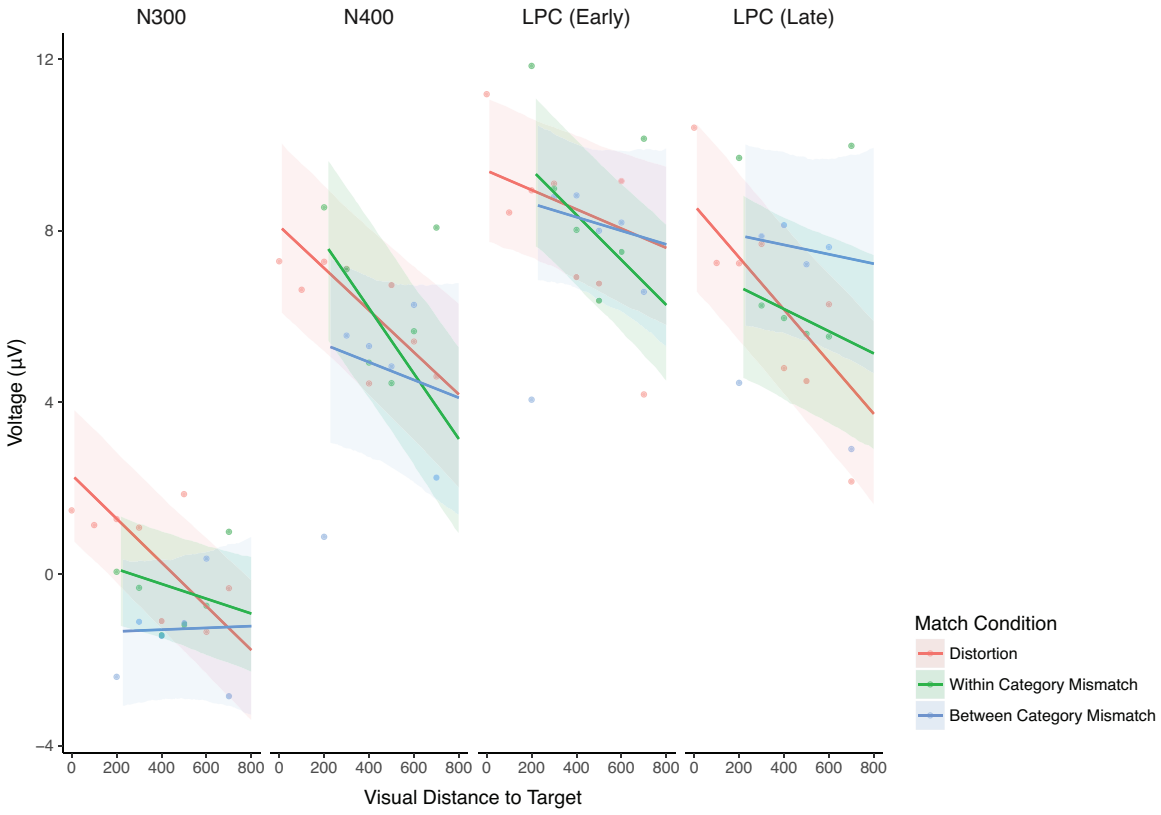


Figure 2.9. Estimated effect of visual similarity to target by component and match condition. Higher distances indicate that the presented object was more visually distinct from the target object at test. Visual distance generally impacted the ERP waveform similarly across the Distortion, Within-Category Mismatch, and Between-Category Mismatch conditions, in that the waveform was more positive the more visually similar the presented object was to the target. Only behaviorally correct and artifact free trials included; points aggregated across subjects by visual distance rounded to the nearest 100; points containing fewer than 10 trials were dropped. Linear trends were derived from a linear mixed effects model including the interaction between visual distance and condition. Highlighting indicates 95% bootstrap confidence intervals on linear trends (2000 iterations).



CHAPTER 3: PRIOR EXPOSURE TO SCENE CONTEXT FACILITATES HIGHER LEVEL VISUAL PROCESSING OF RECENTLY ASSOCIATED OBJECTS

Abstract

Visual object recognition and categorization is facilitated by contextual congruency with the surrounding scene. However, the precise mechanisms involved remain unclear. Building on prior work showing that previewing a scene can facilitate processing of associated objects even at the level of visual analysis (Chapter 2), we here asked to what extent such facilitation is dependent on viewing time. Across two ERP studies (N=60), we had participants study categorically organized novel object-scene pairs in an explicit paired associate learning task. At test, we varied contextual pre-exposure duration, both between (200 ms vs. 2500 ms) and within subjects (0-2500 ms). We found that amplitude reductions on the N300, an event-related potential component linked to the processing of object structure, are enhanced for scene-congruent objects with longer scene previews, up to approximately one second of preview time. Similar results were obtained in a separate component-neutral analysis of mismatch trials using visual similarity between the target (contextually congruent) and presented objects as a covariate. Results are consistent with a predictive pre-activation account of scene-object facilitation.

Key Words: visual object recognition; N300; contextual priming; paired associate learning; statistical learning

Introduction

Visual object processing is facilitated when objects are embedded in supportive (vs. incongruent) scene contexts. For example, in prior work (Chapter 2) we demonstrated that visual processing of even newly-associated objects is facilitated by the prior presentation of contextually congruent scenes, consistent with views suggesting an important role for predictive processing in the brain (Friston, 2005; Rao & Ballard, 1999; Mumford, 1992). However, it remains unclear the extent to which such contextual facilitation effects develop over time, and if so, what that time course is like. The current study addresses this line of inquiry by manipulating the amount of time that participants have to view a scene context, prior to presentation of a probe object. If the amount of

contextual pre-exposure time impacts later processing of the probe object, this suggests that participants are indeed doing something while they are looking only at the scene context that is facilitating subsequent processing of the object. Thus, rather than assuming that prediction is taking place, we are testing a necessary corollary of the predictive pre-activation account of contextual facilitation for visual object processing. Moreover, by measuring the brain's electrophysiological response to the object using event related potentials (ERPs), while manipulating the object's degree of match to the most statistically likely, contextually congruent object, we can form a clearer understanding of the types of information that are facilitated by contextual pre-exposure under various timing constraints. In particular, the N300 response can be used as an index of the extent to which the visual form of the object has been primed by the preceding scene context. Similarly, sensitivity of the N300 to timing constraints on prediction can be used to inform our understanding of how timing impacts the level of detail that context-informed predictions may assume in the service of visual object processing.

Congruent Contexts Facilitate Visual Object Understanding

Visual object recognition and categorization is facilitated by congruent scene contexts. More specifically, visual object categorization judgments for ambiguous or degraded stimuli are strongly influenced by the surrounding context (e.g., Palmer, 1975; Bar & Ullman, 1996; Freeman et al., 2015; Barenholtz, 2014; Davenport & Potter, 2004; Brandman & Peelen, 2017). Davenport and Potter (2004), for example, demonstrated that the identification of foreground objects in briefly presented (and backward masked) photographs was facilitated when they were thematically associated with the simultaneously presented background. For instance, a priest would be more likely to be correctly identified on a church background than on a football field background. Not only do identification judgments become more accurate (e.g., Palmer, 1975; Davenport & Potter, 2004; Davenport, 2007; Auckland, Cave, & Donnelly, 2007) but response times in categorization and identification tasks are generally reduced given a facilitating visual context, when compared with an incongruent, equally cluttered context (identification: e.g., Ganis & Kutas, 2003; Gronau, Neta, & Bar, 2008; categorization: e.g., Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008; Sun, Simon-Dack, Gordon, & Teder, 2011). Ganis and Kutas (2003), for example, used a two-stage response paradigm in which participants hit a button as soon as they

felt they could identify an object, then typed their identification response. Response times for the initial button press were faster for objects that were superimposed on congruent contexts after a short (300 ms) delay, when compared to incongruent contexts. Along a similar vein, Barenholtz (2014), using a paradigm in which objects embedded in scenes were presented at successively less severe levels of visual degradation until identification, found that objects could be identified at a higher level of visual degradation when they were associated with a congruent visual context. The question of which specific processing stages are facilitated by contextual congruency is still under debate, with some advancing the hypothesis that even early visual processing stages are contextually facilitated (e.g., Bar, 2003; Kok, Failing, & de Lange, 2014), and others suggesting that context effects may only come into play at a later deliberative stage that is not tightly yoked to perceptual processing (e.g., Henderson & Hollingworth, 1999; Firestone & Scholl, 2016).

Neurophysiological Indices of Visual Object Processing Shed Light on the Nature of Contextual Facilitation

Neural evidence suggests that scene-object priming is at least partially visuo-structural in nature. That is, it goes beyond amodal semantic priming, and includes priming of information, such as global shape, which is specific to the visual modality. One particularly relevant source of evidence for this comes from the event-related potential (ERP) component known as the N300, and related fMRI studies using manipulations known to induce N300 effects. The N300 (also known as the N350, Ncl or N3 complex) is a negative-going waveform peaking between 200-400 ms at fronto-central sites, which is sensitive to several manipulations relevant to visual object understanding and might be described at a cognitive level as reflecting a fairly high-level processing stage in visual object recognition and categorization. One prominent neurocognitive model (the multistate interactive or MUSI account[2]; Schendan & Ganis, 2012; 2015) has specifically linked the N300 response to the dynamic interaction between fronto-parietal control networks and more modally selective occipito-temporal networks involved in higher level visual processing. The sensitivity profile of the N300 supports a tight link to higher level visual

[2] An earlier form of this theory was known as the two-state interactive account of visual object cognition (Schendan & Kutas, 2007).

processing. For example, N300 amplitude is sensitive to the noise level at which a visually degraded object is recognized (Doniger et al., 2000; Schendan & Kutas, 2002), to noncanonical viewpoint accommodation using semantically rich stimuli (Schendan & Kutas, 2003; McPherson & Holcomb, 1999), and to the degree of rotational discrepancy between abstract object images in a mental rotation task (Schendan & Lucia, 2009). Objects that are more difficult to recognize tend to elicit larger amplitude N300's, as is evident in all of the tasks just mentioned, as well as in contrasts between pseudo-objects and real objects or scrambled versus intact objects (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999; Schendan & Lucia, 2010; Schendan & Ganis, 2015). The N300 also shows selective sensitivity to form-based visual priming, in attempts to differentiate it from multimodal semantic priming indexed by the later peaking and more centro-parietal (but often co-modulated) N400 response (Hamm, Johnson, & Kirk, 2002; Kovalenko, Chaumon, & Busch, 2012). Attempts at ERP source localization and parallel studies using fMRI have suggested that the N300 response is likely localized to extrastriate occipito-temporal cortex (possibly including the Lateral Occipital Complex, or LOC, which encodes higher level form information for visual objects), but reflects feedback connections from a fronto-parietal control network that only comes into play after roughly 200 ms post stimulus onset (reviewed in Schendan & Ganis, 2015).

Of particular relevance to the current study, N300 amplitude is also modulated by the congruity of a visual object with its context, and is reduced by the copresence of a related visual scene (e.g., object-object priming: Barrett & Rugg, 1990; McPherson & Holcomb, 1999; scene-object priming: Võ & Wolfe, 2013; Mudrik, Lamy, & Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014; Truman & Mudrik, 2018). This is important, because it suggests that contextual priming of objects by scenes goes beyond amodal semantic priming, and extends to modality specific priming of the visual form of the anticipated object. However, most studies of N300 priming have relied on preexisting scene-object associations present in the natural world. We previously tested whether visual form priming of the N300 extends to novel associations between objects and scenes, validating that the learning paradigm used in the current study, which has benefits for controlling lower-level visual properties of the stimuli, can reproduce N300 visual form priming of objects by scenes (Chapter 2). Participants studied multiple pairs of novel objects and natural scenes and, in a subsequent test period, were required to judge whether an object superimposed on a scene was the same object originally studied with that scene.

Importantly, because we used novel objects with no preexisting associations to the scenes, we were able to perfectly counterbalance the visual statistical properties of contextually associated and unassociated objects. We found that participants are able to rapidly associate novel object and scene categories. Moreover, when scene contexts are presented 2500 ms prior to the onset of a matching or mismatching novel object, N300 scene congruity effects can be elicited in response to the object – effects that are qualitatively similar to those found using natural, pre-existing statistical associations between objects and scenes. We further showed that effects beginning at approximately 200 ms showed a graded sensitivity to visual similarity between the presented and target (congruent) object (more similar objects eliciting more positive responses), which provides component-neutral evidence that, with an extended scene preview, the visual form of the target novel object is primed by the scene in our paradigm

We care not only what stages of processing can be aided by contextual facilitation, but also about the timing properties of these effects. Timing is of interest because it constrains hypotheses regarding the types of scene information used to inform object understanding, and also the degree to which contextual facilitation effects are attentionally-mediated and subject to strategic control. For example, contextual facilitation of visual form could be an automatic process relying solely on information that can be rapidly extracted from a scene. Alternatively, it could reflect a more temporally extended predictive process. A temporally extended process would open up the possibility that object processing is additionally influenced by scene properties that take longer to extract. It also support the hypothesis that features of the upcoming contextually congruent object are predictively preactivated in an attentionally-mediated fashion, which could in turn be more subject to strategic control. Some contextual facilitation effects do indeed require time to unfold. In the domain of language comprehension, for example, time must elapse following contextual exposure in order to observe certain kinds of contextual benefits. For instance, the order of words in a sentence can greatly change the meaning of the sentence (e.g., "the cop arrested the thief" vs. "the thief arrested the cop"), but it can take a second or longer to use word order information to inform predictions regarding upcoming words in a sentence (Chow, Lau, Wang, & Phillips, 2018). Similarly, the stimulus onset asynchrony (SOA) of words in sentences presented using rapid serial visual presentation (RSVP) has been shown to affect predictive processing, with shorter SOAs associated with diminished predictive effects (Wlotko & Federmeier, 2015). Specifically, priming of contextually incongruent words that were

semantically related to the expected word was diminished for 250 ms SOAs compared to 500 ms SOAs[3]. In the domain of scene-object priming, the extent to which pre-exposure time constrains the extent and nature of the contextual benefit for object processing remains an open area of inquiry, motivating the current study.

Mixed Interpretations of the Processing Stages Impacted by Scene-Object Priming

The ERP literature on scene-object priming has revealed that scenes prime objects to some extent with both successive/delayed and simultaneous presentation, although precise latency estimates and component-based interpretations (as reflecting an N300 or N400 effect, or some combination thereof, in addition to later LPC or slow-wave effects, which are often co-present) have varied across studies. Võ and Wolfe (2013) had participants judge whether scene-object pairs had been presented previously (which was true only of filler trials). They first presented the scene alone for 500 ms, then superimposed a cue to the target object location for approximately an additional 500 ms,. They then removed the location cue and in its place superimposed on the scene a semantically congruent object, a semantically incongruent object that was visually similar to the expected object and obeyed the laws of physics, or a semantically congruent or incongruent object in an unexpected or improbable location given the scene. Võ and Wolfe separately reported N300 and N400 effects of semantic congruency, time-locked to object onset. Specifically, they found a broadly distributed congruency effect between 250-600 ms, such that scene-congruent objects elicited a more positive waveform than incongruent objects. The effect appeared topographically similar across early and late time-windows of analysis (250-350 ms, and 350-600 ms). ,Ganis and Kutas (2003), however, used a similar paradigm which also employed a contextual preview prior to object onset, but interpreted their results as supporting scene-object N400 priming in the absence of N300 priming. Ganis and Kutas (2003) first cued participants as to where an object would appear, then displayed a context scene by itself for 300 ms, before superimposing congruent and incongruent objects at the cued location. For example, either a soccer ball or a toilet paper roll might appear on top of a soccer field. Participants were asked to identify the object that appeared. Ganis and Kutas found what

[3] Interestingly, effects of SOA interacted with block order, in that participants showed a more similar pattern across presentation rates if they initially viewed sentences at the slower rate, before speeding up, consistent with a practice effect that carried over to facilitate rapid RSVP sentence comprehension.

they interpreted as an anterior N400 congruency modulation between 250-500 ms, and further suggested that the absence of a clear N300 effect, as distinct from the N400, better supported the hypothesis that scenes prime objects at a semantic/conceptual level, rather than priming their visual form.

Another line of work has explored the role that advance pre-exposure to the context scene might play in driving scene-object congruency effects, using simultaneous presentation of scenes containing objects that semantically match or mismatch the scene (e.g., Demiral, Malcolm, & Henderson, 2012; Mudrik, Lamy, & Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014; Truman & Mudrik, 2018). Of these studies, only one has employed a between subjects manipulation of scene preview duration (Demiral et al., 2012), while the others made inferences based on the pattern of effects for simultaneous scene-object presentation alone. Unfortunately, Demiral et al. (2012) has some design features that make its results difficult to interpret. Specifically, it repeated the scene contexts within subjects, presenting them on one trial with a congruent and, on another trial, an incongruent object. This introduces a memory confound, if the incongruent object/scene pairs are more memorable than the congruent ones, leading to better memory for the scene context. Indeed, in their first experiment, where scenes were presented 300 ms prior to objects, they found uncharacteristically early effects of object-scene congruency (onsetting < 50 ms post object onset[4]) that could be explained by a reduced N400 (and possibly N300) to the scene, when the current scene was congruent with the object, meaning that in an earlier presentation it was incongruent with the object (and therefore, the scene is better remembered and the repetition effect is larger). In their second experiment, where scenes and objects were presented simultaneously, there does appear to be a delay in the onset of classical scene-object priming effects, consistent with an account in which visual form-based priming is more likely to appear when the scene precedes the object. However, if the N400 response (and potentially also the N300 response) to congruent scenes is again reduced due to enhanced repetition priming as discussed above, we would now have two competing effects: larger N300-N400's to congruent scenes, and smaller N300-N400's to congruent objects. For this reason, it is very difficult to interpret the pattern of congruency effects for simultaneous scene-object presentation in this study.

[4] It takes roughly 50 ms for visual information to make first contact with primary visual cortex in humans (Hillyard & Anllo-Vento, 1998), and the filter settings applied to EEG data in Demiral et al. are mild (.1-80 Hz half-amplitude cut-off) and are unlikely to explain this discrepancy.

Mudrik and colleagues examined scene-object congruency effects using simultaneous scene-object presentation, without the above repetition confound, in a series of experiments wherein similarly shaped but contextually incongruent objects were artificially inserted in place of congruent objects in natural photographs (Mudrik, Lamy, & Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014; Truman & Mudrik, 2018). Mudrik, Lamy, and Deouell (2010) had participants judge whether a cued object in each photograph was being manipulated with zero, one or both hands by people in the image. Critically, on half the trials, the original object in the cued position was digitally replaced with a semantically incongruent object. The incongruent object roughly matched the shape of the original object on some trials, but the degree of match varied across stimuli; for example, an arrow might be replaced by a tennis racket, a basketball with a watermelon, etc. An attempt was made to control for lower level visual properties of the congruent and incongruent images, including spatial frequency, contrast, and chromaticity. Similar to scene-object priming paradigms including a delay between scene and object onset, Mudrik et al. (2010) found that contextually incongruent objects elicited a more negative waveform over anterior sites between roughly 270-650 ms (although cluster analysis results suggest an even earlier onset at some electrode sites). This basic finding has since been replicated twice on different sets of participants using subsets of these stimuli (Mudrik et al., 2014; Truman & Mudrik, 2018).

Mudrik and colleagues interpret their results as suggesting that N300 congruency effects can be elicited even with simultaneous scene and object presentation. Importantly, however, only the incongruent images were photo-manipulated, and, by the nature of the design, the stimuli could not be counterbalanced across conditions. Truman and Mudrik (2018) compared the onset latency of two ERP contrast effects within subjects: scrambledversus intact objects and congruent versus incongruent objects. They found very similar onset latencies for these two effects (240 ms for scrambled versus intact, 260 ms for congruent versus incongruent). Notably, however, scrambled objects were presented within intact scenes and still shared the same rough shape as the original intact (congruent and incongruent) objects, making them somewhat identifiable from context. Overall, Mudrik and colleagues have demonstrated that at least some kinds of contextual congruity effects can include N300 modulations, even with simultaneous presentation of the object and scene. Quite possibly, though, these effects may rely on gradual statistical learning of object-context associations that may not extend to novel object stimuli.

Mudrik and colleagues' findings also do not rule out the possibility that additional, more temporally extended processing may further refine and enhance contextual priming effects, including priming of an object's visual form. In the current study, we will build on our prior work, extending the paradigm used in Experiment 2.1 to observe how visual form priming interacts with timing constraints on contextual pre-exposure.

Testing and Refining the Predictive Pre-activation Hypothesis of Context Effects in Visual Object Understanding

In the current study, we report two experiments designed to assess the extent to which the contextual facilitation effects we observed previously are contingent on having sufficient pre-exposure time to the context scene. Participants studied novel object-scene pairs in a paired-associate learning study. In Chapter 2 (Experiment 2.1), scenes had been presented 2500 ms prior to objects in the test phase, allowing ample time for temporally extended predictive processing. Here, in our first experiment (Experiment 3.1), we shortened contextual pre-exposure time from 2500 ms to 200 ms, leaving participants with enough time to rapidly extract the gist of the scene (reviewed in Larson, Freeman, Ringer, & Loschky, 2014) prior to object onset, but potentially restricting their ability to actively prepare for viewing the upcoming object. Active preparation might be limited by imposing time constraints either by limiting the time participants have to use (potentially any kind of) information about the scene to form a detailed prediction for the upcoming object, or by restricting the nature of the information participants are able to extract from the scene in advance (i.e. to a coarser gist, rather than a finer-grained scene representation incorporating information derived from higher spatial frequencies). To permit statistical inferences regarding the effect of prolonging contextual pre-exposure time on visual object processing, we directly compare the results of our first experiment to the dataset in Chapter 2. Specifically, we examine the size and timing properties of the scene-object contextual match/mismatch effect across the two different pre-exposure times in a between subjects comparison. If the N300 match/mismatch effect is larger given a longer pre-exposure time, or if shortening the pre-exposure time leads to qualitative and/or quantitative evidence for processing delays, this would suggest that participants are in fact using the time that they are observing the context in isolation (beyond the first 200 ms) to prepare for the most likely upcoming object at

various stages of processing. An interaction with N300 match/mismatch effect size, specifically, would support the idea that visuo-structural priming of objects by scenes is contingent, in at least some cases, on having sufficient time to process the scene in advance.

We also examine whether another contextual congruency effect that we observed and reported for the first time in Chapter 2 is contingent on having sufficient contextual pre-exposure time. Specifically, we previously observed that contextually incongruent objects that are visually similar to the target matching object elicit more positive ERP waveforms beginning at roughly 200 ms. This effect showed finer sensitivity to visual features than what would be expected from our broader contextual congruency manipulations alone and suggests that participants evoked a visual template of the target scene-congruent object and visually compared the presented object to this target. If target similarity effects are diminished or absent with a reduced contextual pre-exposure time, this would support the hypothesis that they are contingent on having sufficient time to process the scene in advance.

In our second experiment (Experiment 3.2), we examine this same question of how timing constraints affect scene-object priming using a within-subjects design to ensure that processing differences contingent on scene preview duration hold within individuals in a single experimental session. We also vary scene preview duration (i.e. contextual pre-exposure time) continuously between 0 and 2500 ms to better approximate how much contextual pre-exposure time is needed to observe visuo-structural priming on the N300 elicited by a contextually congruent/incongruent visual object. Unlike Chapter 2 Experiment 2.1 and Experiment 3.1 of this paper, we also modified the design to collect response time data in addition to observing the electrophysiological response. By observing how response time and ERP component amplitudes change as a function of contextual pre-exposure time, we can gain a finer grained understanding of how long it takes to extract relevant information from a context scene and apply that information towards preparing for the task of determining contextual congruency of a later presented object.

Thus, in Experiment 3.1, we use a high-powered design to identify context-based effects when scene preview time is limited, and directly compare our results to prior work to assess the impact of prolonged pre-exposure to a scene context on object understanding. In Experiment 3.2, we conduct a within subjects replication and extension of our first experiment that allows us to gain a finer grained understanding of how preparatory processing tuned to facilitate processing

of upcoming visual objects evolves over time. Across the two experiments, we are able to assess how people use contextual information to predict (or otherwise facilitate processing): N300 modulations and component-neutral visual similarity effects suggest visual form-based priming, N400 effects index semantic priming, and late positive complex (LPC) and response time (RT) data speak to the decision-making process involved in determining contextual congruity.

Experiment 3.1

In our first experiment, we replicate the design of Experiment 2.1 but with one critical adjustment: scene preview duration in the test phase has now been shortened to 200 ms. By directly comparing the current experiment with Experiment 2.1, we can examine whether and how participants use extended contextual preview time to prepare to see the target object image at test.

Methods

Participants

Data are reported from 24 participants (mean age 22, range 18-29; 9 males), all native English speaking University of Illinois undergraduates, who were compensated with payment. One additional participant was replaced due to excessive trial loss. All participants provided written informed consent, according to procedures established by the IRB at the University of Illinois. Handedness was assessed using the Edinburgh inventory (Oldfield, 1971). All participants were right-handed; mean score: .84, where 1 denotes strongly right-handed and -1, strongly left-handed. 8 reported having left-handed family members. No participants had major exposure to languages other than English prior to the age of 5, and none had a current diagnosis of any neurological or psychiatric disorder or brain damage or was using neuroactive drugs. All had normal or corrected-to-normal vision for the distances used in the experiment. All participants also passed a behavioral criterion for inclusion: they showed significant sensitivity in their response distribution to the match/distortion vs. mismatch conditions (Pearson's Chi Squared Statistic, all individual participant p's < .001). Participants were randomly assigned to one of 24 experimental lists.

Materials

Overview

Materials and counterbalancing are described in detail in Chapter 2; however, key points are reproduced here. Novel objects resembling either biological organisms ("germs") or mechanical devices ("machines") were paired with natural scenes. Objects were organized hierarchically such that germs and machines could be subdivided into major categories, then subcategories, then exemplars (i.e., distortions). At study, each major object category was consistently paired with a scene type (e.g. germ category 1 with beaches, germ category 2 with forests) for any given participant. In the test phase, participants viewed each scene from the study phase, followed by an object that exactly matched what they had studied with that scene, a distortion of that object, an object from a different subcategory (that had been associated with that scene type, but not that specific scene, at study), or an object from a different major category (that would never have appeared on that scene type at study). Across the experiment, all object types were paired with all scene types, and objects and scenes were never repeated in the study phase.

Scenes

Scenes depicted one of 6 categories: beaches, city streets, mountains, forests, highways, and offices. Scenes were drawn from a pool of 288 images, 48 per category, that were previously normed as being highly representative of their respective scene types and which had been rescaled to 800 x 600 pixels (see Torralbo et al., 2013 for norming details).

Objects

Line drawings of novel object prototypes for biological organisms ("germs") or mechanical devices ("machines") were created by an artist with the aid of Adobe Photoshop to maintain a consistent set of visual textures. Within the two classes of germs and machines, drawings were further organized into 6 major categories, each with 3 subcategories. Thus, there were 18 total subcategories of germs and machines, respectively, each with a single representative prototype image. Major categories shared aspects of their visual structure and texture, as well as homologous parts. From each subcategory prototype image, 24 exemplar images were derived by changing the relative positions, proportions and orientations of the object parts. Prototype images

were never shown to participants. For details on how exemplar images were created, see Chapter 2. All novel object line drawings were resized to 274 x 274 pixels.

Counterbalancing and Experimental Conditions

24 experimental lists of stimuli were used, identical to those in Chapter 2, which contains more details on counterbalancing. For any given list (and participant), each germ or machine type would only ever appear on a single scene category (e.g. "beaches") at study, and this relationship would hold for all 3 subtypes of the object major category. The mapping between object and scene categories was then systematically rotated across the 6 sets, so that over the entire experiment every object category was associated with every scene category at study. Each list consisted of 288 study and 288 test pairs of novel object and scene stimuli. Stimuli within each list were organized into 18 blocks of 16 study pairs followed by 16 test pairs. Blocks alternated between all germs and all machines; the first block for each list was always germs. The same set of 288 unique scenes was used across all lists; scenes were presented exactly once at study and once in the corresponding test phase in each list. Test objects were never repeated within a list. Within each block, an approximately equal number of pairs involving each scene category and object major category were presented at both study and test. All 16 scenes presented at study within a block were presented in a different pseudorandom order in the corresponding, and immediately following, test phase. In the test phase, 4 of the 16 test trials were assigned to each of 4 experimental conditions (See Figure 3.1):

5. **Exact match**: test object exactly matched the object paired with the scene at study
6. **Distortion**: test object was a different exemplar of the same subtype of object presented with the scene at study
7. **Within-category mismatch**: test object was a different subtype of object within the same major category as the object presented with the scene at study
8. **Between-category mismatch**: test object was a different major category of object from that presented with the scene at study (and thus, belonged to a category of object that would never have been studied with that scene category previously; e.g., an object category that had only ever appeared on offices at study, now appearing on a beach at test).

Half of the between category mismatches in the test phase were created by swapping two object subcategories from different major categories that had been presented in the preceding study phase. The other half were created by introducing an object category that had not been presented in the preceding study phase. Within category mismatches were always created by swapping object subcategories within a major category that had been presented in the preceding study phase. Other than in the exact match condition, the exact object exemplar presented at test never matched that presented in the preceding study phase (i.e. a distortion derived from the same object prototype was used).

Each list contained 72 trials per condition. Across lists, each test object and each scene appeared in each of the four experimental conditions an equal number of times. Trial order within each block was pseudo-randomized, such that no more than two trials corresponding to each condition were presented in a row, and no more than 3 trials mapping on to a 'same' response (that is, trials in the exact match or distortion conditions) occurred in a row at test. Response hand was counterbalanced across lists.

Procedure

Participants passively studied the paired scenes and novel objects and then were tested by being asked to indicate, for each in a new set of pairs, whether the object matched the presented scene. In each study trial, the scene alone was first presented centrally (2500 ms) on a black background, followed by the object superimposed on the scene (2500 ms). In the test phase, the scene was again presented alone, this time for a brief interval (200 ms), followed by the object superimposed on the scene (2500 ms), after which a blank screen with a question mark appeared, prompting the participant's response. Participants had 3 response options: the object on this scene is (1) the same object that I studied with this scene, (2) not the same object, but "could have gone with" this scene, (3) not the same object and could not have gone with this scene. Participants were told that if the object was only slightly visually distinct from what they remembered studying (e.g. had a different body position or proportions), they should still respond (1). They were also told that an object and scene 'could go together' if they believed that pair looked similar to other study items and could hypothetically be presented in an upcoming trial of the experiment, even if they knew that they hadn't studied it. Participants were never told

that there was a structured relationship among the object and scene categories. Participants were explicitly instructed that each test phase only covered materials studied in the immediately preceding study phase and that testing was non-cumulative across blocks. Prior to the main experiment, participants were given a practice block of 4 study and 4 test trials.

During recording, participants were seated 100 cm away from the computer in a comfortable chair. The visual angle of the scenes was 13.6° by 9.3° and that of the object images was 4.6° by 4.2°. The recording session lasted approximately 90 minutes. Afterwards, for a subset of the object images (one from each of the 6 major categories of germs and machines), participants indicated which scene categories the object had been associated with during the experiment by circling one or more of 6 scene category labels.

For further details on the procedure, see Chapter 2. The experimental procedure was the same as in Chapter 2 Experiment 2.1, except that during the test phase, scenes were presented in isolation for 200 ms (not 2500 ms) prior to super-imposing the object on top of the scene. This enables a between-subjects comparison across the experiments.

EEG Data Acquisition and Preprocessing

The electroencephalogram (EEG) was recorded from 26 silver/silver-chloride electrodes evenly spaced over the scalp. The sites are midline prefrontal (MiPf), left and right medial prefrontal (LMPf and RMPf), left and right lateral prefrontal (LLPf and RLPf), left and right medial frontal (LMFr and RMFr), left and right mediolateral frontal (LDFr and RDFr), left and right lateral frontal (LLFr and RLFr), midline central (MiCe), left and right medial central (LMCe and RMCe), left and right mediolateral central (LDCe and RDCe), midline parietal (MiPa), left and right mediolateral parietal (LDPa and RDPa), left and right lateral temporal (LLTe and RLTe), midline occipital (MiOc), left and right medial occipital (LMOc and RMOc), and left and right lateral occipital (LLOc and RLOc). The midline central (MiCe) electrode was placed where the "Cz" electrode would appear using the international 10-20 system. Eye movements were monitored via a bipolar montage of electrodes on the outer canthus of each eye. Blinks were detected by an electrode below the left eye. Impedances were kept below 5 KΩ. Signals were amplified with a .02–250 Hz bandpass using a BrainVision amplifier and digitized at 1000 Hz. Data were referenced online to the left mastoid and rereferenced offline to the average of the left

and right mastoids. Each trial consisted of a 1000 ms epoch preceded by a 200 ms prestimulus baseline. Trials contaminated by eye movements, blinks or other recording artifacts were rejected offline. Artifact rejection procedures using subject-specific threshold parameters resulted in average trial loss of 22.2% for the exact match condition, 21.9% for the distortion condition, 23.1% for the within-category mismatch condition, and 21.1% for the between category mismatch condition. A digital lowpass Butterworth IIR filter with a 30 Hz half-amplitude cut-off and 12 dB/octave roll-off was applied prior to statistical analysis. Prior to permutation-based cluster analysis, data were further down-sampled to 100 Hz.

Analysis

Behavioral analyses of response distributions were conducted using logistic regression modeling in R. Statistical analyses of individual trial EEG data were conducted using mixed effects models built with the lme4 package in R. Models initially included crossed random effects of subject and item, and by-subjects random slopes of each predictor of interest, but the random effects structure was sometimes scaled back to address convergence issues. EEG dependent measures consisted of mean amplitudes over time windows and electrode sites selected to capture particular components of interest, and are identical to Chapter 2 to maximize comparability. Dependent measures of interest are mean amplitudes at: 250-349 ms over fronto-central sites to capture the N300, 350-499 ms over centro-parietal sites for the N400, and 500-699 ms and 700-899 ms over posterior sites to capture early and late time windows of the Late Positive Complex (LPC). Key within-subjects predictors of EEG amplitude included match condition and visual similarity to the target object. In a between-subjects analysis combining the current data set with that in Chapter 2, we assessed interactions between scene preview time at test (200 ms vs. 2500 ms) and the effects of match condition and visual similarity to target. We also conducted an exploratory time-domain cluster analysis examining effects of match condition in the current experiment, results of which are presented in Appendix E.

Results

Behavioral – Online Accuracy

As in Chapter 2 Experiment 2.1, participants discriminated well among test objects that matched and mismatched the presented scene. Accuracy was computed after collapsing the exact match

and distortion conditions (which were treated as a 'match') and the two mismatch conditions along with the two different mismatch responses ('possible' and 'impossible'). Mean accuracy was 80.3%, range 59.7-93.4%. Comparing the current dataset to the dataset in Chapter 2, accuracy did not differ from when scene preview was 2500 ms (Welch two-sample t-test, t = .45, df = 45.6, p = .67). Participants were also sensitive to the type of mismatch. A logistic regression model predicting the probability of response 'impossible mismatch' with fixed effect of condition (exact match or distortion vs. within category mismatch vs. between category mismatch), using subject random intercepts, by-subjects random effects of match condition, and nested model comparisons, confirmed sensitivity to mismatch type: intercept = -0.938, β = 2.57, SE = .324, z = 7.934, χ_1^2 = 30.92, p < .001. Figure 3.4a shows the mean response distribution across subjects for each condition. Distribution of responses was comparable to that seen when scene preview was 2500 ms (Chapter 2).

Participants were numerically more likely to respond to between category mismatches as being impossible, compared to Experiment 2.1 (due to convergence issues, a simplified random effects structure was used that removed the by subjects random slope for match/distortion vs. within/between category mismatch; β = -.76, SE = .405, z = -1.87, χ_1^2 = 3.40, p < .1). There was no difference in probability of responding 'match' to the exact match and distortion conditions across the current experiment and Experiment 2.1 (due to convergence issues, a simplified random effects structure was used that removed the by subjects random slope for between category mismatch vs. other match conditions; χ_1^2 < 1).

Behavioral – Posttest Categorization

On the posttest, participants demonstrated explicit knowledge of the categorical mapping among objects and scenes. Participants were more likely to indicate that objects were associated with a scene category when the two had been paired at study. Figure 3.4b shows the normalized confusion matrix indicating the probability that a scene category, if circled, belonged to the correct scene category for the depicted object. This was assessed with a logistic regression model predicting the probability of circling a scene category, with a fixed effect of match, crossed random intercepts for subject, object (response item) and scene (response choice), a by-subjects random effect of match, and nested model comparisons: intercept = -4.79, β = 8.16, SE = .869, z = 9.388, χ_1^2 = 43.87, p < .001. When results were compared with Experiment 2.1, there were no

significant interactions with experiment (probability of circling a scene type was similar for both matching and mismatching scene types across experiments, |z|'s < 1).

ERP Analysis: Match Condition

We followed the component-based analysis approach in Chapter 2, important aspects of which are reproduced below. We also extended the analysis to directly compare the current dataset with that in Chapter 2 (see Figure 3.6 for a visual comparison of mismatch - match ERP difference waves across these two studies). Only behaviorally correct (response option 1 = 'match' for match and distortion and response options 2 = 'possible mismatch' or 3 = 'impossible mismatch' for within and between category mismatches) and artifact free trials were included. Linear mixed effects models were fit to the individual trial data, including fixed effects of:

(4) condition (contrasting match, distortion, within, and between category conditions)
(5) response type (for within and between category mismatch conditions only, since only one response type was considered correct for match and distortion conditions); condition (within vs. between) was moderately associated with response ('possible' vs. 'impossible' mismatch) in the current experiment; Cramér's V = .43.
(6) the interaction between condition and response

Models also included crossed random intercepts of subject, item (object+scene), and channel, and by subjects random slopes of condition, response and their interaction. We also compared this model to one with an additional fixed effect, which contrasted between category mismatch swap trials (generated by swapping the condition of object images that had been presented in the immediately preceding study phase) and between category mismatch new trials (generated by presenting an object image that had not been presented in the immediately preceding study phase).

All models were fit using maximum likelihood estimation. Fixed effects were initially tested using likelihood ratio tests with nested model comparisons. Follow up comparisons of condition means were conducted using the contest function in the lmerTest package in R, with family-wise error rate corrected p-values and the Satterthwaite approximation for the degrees of

freedom. Except where otherwise indicated, condition contrasts collapse across response type using linear combinations of beta weights but were conducted on a model that included a fixed effect of response type. The distortion condition was included in all models, but, as in Experiment 2.1, never differed reliably from the match condition. For ease of reporting, we thus describe only contrasts among the exact match and mismatch conditions.

N300

To examine the N300, voltages were separately averaged across time for each trial from 250-349 ms at each of 16 frontal and central sites. Due to convergence issues, the by-subjects random effect of distortion condition was removed. There was a main effect of condition (χ_3^2 = 12.95, p < .01), which remained when recency was accounted for by including between category mismatch swap versus new as a fixed effect (χ_3^2 = 9.33, p < .05, random effect of response to mismatch trials removed to achieve convergence; accounting for recency improved model fit, χ_1^2 = 9.45, p < .01). Follow-up comparisons revealed that between category mismatches were more negative than matches (diff = -1.20 μV, $F_{(1,28.7)}$ = 9.2, p < .01). Within category mismatches were also numerically more negative than matches (diff = -.93 μV, $F_{(1,23.0)}$ = 3.1, p < .1). Between and within category mismatches did not differ from each other (F < 1). Among between category mismatches, new trials were more negative than swap trials (diff = -1.03 μV, $F_{(1,3240)}$ = 9.2, p < .01). When examined separately, between category mismatch new trials were significantly more negative than matches (diff = -1.71 μV, $F_{(1,41.1)}$ =16.0, p < .001), but swap trials were not ($F_{(1,39.9)}$ = 2.5, p > .1). There was no main effect of response nor interaction between match condition and response among mismatch trials (response: χ_1^2 < 1; response x match condition: χ_1^2 = 2.53, p > .1).

Thus, in the N300 time window there was a more negative response to new items (which were only used in the between category mismatch condition) compared to repeated items. There was also a tendency for mismatches to be more negative than matches, as in Experiment 2.1, but this effect was small and not reliable in the pairwise comparisons.

N400

To examine the N400, voltages were separately averaged across time for each trial from 350-499 ms at 8 central and parietal sites. There was a main effect of condition (χ^2_3 =14.15, p < .01), that persisted even when recency was accounted for by including between category mismatch swap versus new as a fixed effect (χ^2_3 = 9.11, p < .05; accounting for recency improved model fit, χ^2_1 = 6.36, p < .05). Follow-up comparisons revealed that both between and within category mismatches were more negative than the match condition (between: diff = -1.66 μV, $F_{(1,26.7)}$ = 12.0, p < .01; within: diff = -1.03 μV, $F_{(1,17.6)}$ = 4.9, p < .05). Between and within category mismatches did not significantly differ from each other ($F_{(1,23.5)}$ = 1.6, p > .1). Among between category mismatches, new trials were more negative than swap trials (diff = -.99 μV, $F_{(1,3126)}$ = 6.7, p < .01). Both new and swap between category mismatch trials were more negative than matches (new: diff = -2.16 μV, $F_{(1,36.2)}$ = 17.5, p < .001; swap: diff = -1.17 μV, $F_{(1,35.2)}$ = 5.1, p < .05). There was no main effect of response nor interaction between match condition and response among mismatch trials (response: χ^2_1 < 1; response x match condition: χ^2_1 = 1.09, p > .1).

In the N400 time window, there was thus a robust effect of match versus mismatch, as well as an effect of recency, replicating the pattern seen in Experiment 2.1. However, there was no significant effect of mismatch type.

LPC

To examine the LPC, voltages were separately averaged across time for each trial from 500-699 ms (early LPC) and 700-899 ms (late LPC) at each of 8 posterior sites. In the early time window there was a main effect of condition (χ^2_3 =11.99, p < .01), which persisted even when recency was accounted for by including between category mismatch swap versus new as a fixed effect (χ^2_3 = 10.56, p < .05; accounting for recency did not significantly improve model fit, χ^2_1 < 1). Both between and within category mismatches were more negative than matches (between: diff = -1.46 μV, $F_{(1,25.9)}$ = 7.9, p < .01; within: diff = -1.59 μV, $F_{(1,19.5)}$ = 11.5, p < .01). Between and within category mismatches did not significantly differ from each other (F < 1). There was no main effect of response nor interaction between condition and response among mismatch trials (χ^2_1's < 1). This pattern differs substantively from that seen in Experiment 2.1, which found only an effect of response type and no effects of match condition.

In the late time window, there again was a main effect of match condition (χ_3^2 =10.15, p < .05), that persisted even when recency was accounted for by including between category mismatch swap versus new as a fixed effect (χ_3^2 =9.96, p < .05; accounting for recency did not significantly improve model fit, χ_1^2 < 1). Within category mismatches were more negative than matches and between category mismatches (within – match, diff = -1.73 μV, $F_{(1,16.7)}$ = 15.4, p < .01; between – within, diff = 2.21 μV, $F_{(1,23.8)}$ = 20.1, p < .001). Between category mismatches did not significantly differ from matches, and were numerically more positive than matches (diff = .48 μV, $F_{(1,24.4)}$ =1.0, p > .1). Experiment 2.1 also found more positive responses to between compared with within category mismatches in this time window, although in that case matches patterned more like within than like between category mismatches.

There was no main effect of response nor interaction between condition and response among mismatch trials (response: χ_1^2 = 1.52; response x match condition: χ_1^2 = 1.92; p's > .1). Again this differs from Experiment 2.1, which showed a continued effect of response type in the late (as in the early) LPC window.

Between-Subjects Comparison: 200 ms vs. 2500 ms Scene Preview

To examine interactions with scene preview time, we directly compared the current experiment (200 ms scene preview) with the dataset in Chapter 2 (2500 ms scene preview).

For the N300, there was a significant interaction between match condition and scene preview time (χ_3^2 = 13.98, p < .01). The between category mismatch – exact match contrast was larger given a longer scene preview (β = -1.12, SE = .55, χ_1^2 =3.94, p < .05). Among mismatch trials, there was no significant interaction of scene preview time with response, or response x match condition (scene preview x response: χ_1^2 < 1; scene preview x response x match condition: χ_1^2 = 2.04, p > .1).

For the N400, there were no significant interactions with scene preview time (scene preview x match condition: χ_3^2 = 2.54; scene preview x response: χ_1^2 = 1.90; scene preview x response x match condition: χ_1^2 < 1; p's > .1).

In the early LPC window, there was a significant interaction between match condition and scene preview time (χ_3^2 = 11.93, p < .01). Given a long preview time, there was no reliable effect of match condition, with the within and between category mismatch conditions

numerically more positive than the exact match condition. With a short preview, the within and between category mismatches were significantly more negative than the exact match condition. Thus, the [within category mismatch – exact match] and [between category mismatch – exact match] contrasts were significantly more positive given a long scene preview, (scene preview x (within – exact): β = 3.27, SE = 1.03, χ_1^2 = 9.04, p < .01; scene preview x (between – exact): β = 2.29, SE = .76, χ_1^2 = 8.24, p < .01). There was also a significant interaction between scene preview time and response, reflecting that 'possible' mismatch responses elicited a more negative waveform than 'impossible' responses with long scene previews, but not short scene previews (β = -2.84, SE = 1.34, χ_1^2 = 4.34, p < .05).

In the late LPC time window, there was no significant interaction between scene preview time and match condition (χ_3^2 = 5.54, p > .1). There was, however, a significant interaction between scene preview time and response, again reflecting that 'possible' mismatch responses elicited a more negative waveform than 'impossible' responses with long scene previews, but not short scene previews (β =-3.47, SE = 1.54, χ_1^2 = 4.88, p < .05).

ERP Analysis: Target Similarity

Following Chapter 2, we also conducted a separate set of analyses in order to assess whether low-level visual features of the scene-congruent object were accessed in memory even when it was not displayed. We added an additional fixed effect to our match condition analysis models: visual distance from the target object. We then tested whether this predictor improved model fit using nested model comparisons. Models did not distinguish among the two types of between category violations (new vs. swap). The same random effects structure was used as for modeling the effects of match condition. Visual distance was computed as Euclidean distance between the presented and scene-congruent object in a V1-like feature space derived from the model in Pinto, Cox, and DiCarlo (2008). Chapter 2 provides more details. Beta values are reported as 1000 times the original estimates.

There was a significant interaction between experiment (2500 ms vs. 200 ms scene preview time) and the effect of visual similarity to target, for all components of interest (N300: β = -2.255, SE = .426, χ_1^2 =27.96, p < .001; N400: β = -4.124, SE = .602, χ_1^2 = 46.83, p < .001; early LPC: β = -3.926, SE = .599, χ_1^2 =42.94, p < .001; late LPC: β = -3.265, SE = .627, χ_1^2

=27.06, p < .001). Given a 2500 ms scene preview, the waveform was more positive for objects more similar to the target across all four components, but given a 200 ms scene preview (the current experiment), the effect was smaller, absent, or, for the early LPC, there was a significant effect that tended in the opposite direction (N300, β = -1.509, SE = .485, χ_1^2 = 9.67, p < .01; N400 and late LPC: χ_1^2's < 1; early LPC: β = 1.399, SE = .631, χ_1^2 = 4.90, p < .05).

Discussion

With a short preview time, match effects were weak in the N300 window, emerging robustly in the N400 time window (with more negative responses to mismatch than to match trials) and continuing into the early LPC time window. On the late LPC, within category mismatches were more negative than both matches and between category mismatches. There were no interactions with response on any of the four components.

When directly compared with Experiment 2.1, 2500 ms scene previews were found to elicit larger N300 mismatch effects at the test object than 200 ms scene previews. N400 match effect size did not differ across short and long preview times. Having a short (200 ms) scene preview also led to the emergence of match effects (of a similar form to those seen for the N400) in the early LPC time window. The combined pattern of smaller mismatch effects at an early point in time, and larger ones at a later point in time, suggests a delay in the timing with which the brain appreciates the match/mismatch distinction when the participant is given less time to process the scene context prior to viewing the test object. In addition, the early and late LPC were sensitive to the participant's response choice independent of condition only when a long preview time was given. This may partially reflect the tighter correlation between mismatch type and response type in the current experiment, compared with Experiment 2.1 (Cramér's V = .43 vs. .27). Alternatively, it is consistent with a delay in response-related processing.

The results of the target similarity analysis corroborate the patterns in the match analysis in suggesting that short preview times provide less evidence for preparatory processing. With a short preview time, effects of target similarity were smaller or absent, suggesting participants were less likely to have activated a representation of the target object in response to the scene.

Experiment 3.2

In Experiment 3.2, we used a within-subjects parametric design to replicate and extend the findings of Experiment 3.1: that N300 match-mismatch effects are enhanced by giving participants an extended contextual preview prior to viewing the object at test. By treating scene preview duration as a continuous parameter (0-2500 ms), we can 1) estimate how much scene preview time is enough to maximize the contextual benefit and 2) explore whether our findings extend to a case where there is temporal uncertainty in when the object image will appear following the context scene.

Methods

Participants

Data are reported from 36 participants (mean age 20, range 18-25; 10 males), all native English speaking University of Illinois undergraduates, who were compensated with payment. 3 additional participants were replaced due to excessive trial loss (1) or poor behavioral performance on the online task (2). The criterion for poor behavioral performance was showing no significant sensitivity to match condition (Pearson's Chi-squared test). All participants provided written informed consent, according to procedures established by the IRB at the University of Illinois. Handedness was assessed using the Edinburgh inventory (Oldfield, 1971). All participants were right-handed; mean score: .78, where 1 denotes strongly right-handed and -1, strongly left-handed. 15 reported having left-handed family members. No participants had major exposure to languages other than English prior to the age of 5, and none had a current diagnosis of any neurological or psychiatric disorder or brain damage or was using neuroactive drugs. All reported normal or corrected-to-normal vision for the distances used in the experiment.

Participants were randomly assigned to one of 36 experimental lists.

Materials

Similar materials were used as in Experiment 3.1, with the following changes. 36 lists of experimental stimuli (drawing from the same set as in Experiment 3.1) were generated such that each list consisted of 16 blocks of study/test trials. Each study and test phase contained 18 object-scene pairs. Test phase trials were divided evenly into three experimental match

conditions: exact match, within category mismatch, and between category mismatch. Unlike in Experiment 3.1, all mismatch trials were created by recombining objects and scenes that had been presented in the preceding study phase. The same exemplar object images presented in the study phase were presented at test. Blocks alternated between germ and machine objects. Whether the first block of trials consisted of germs or machines was counterbalanced across participants. Test trials were divided evenly into three of the four object match conditions from Experiment 3.1: **exact match**, **within category mismatch**, and **between category mismatch**. In addition, the amount of time scenes were presented prior to object onset at test (scene preview duration) was varied continuously between 0 and 2500 ms, orthogonal to match condition. Test trials were pseudorandomized such that no match condition was ever presented more than 3 times in a row. Furthermore, no object major category or scene type was ever presented more than twice in a row at study or test.

Procedure

The task and procedure were similar to Experiment 3.1, with the following changes. As in Experiment 3.1, participants were instructed to respond 'match' if they felt that the presented object image in the test phase was a distorted version of the image they had studied with that scene in the study phase. And, as in Experiment 3.1, a distortion trial was included in the practice. However, no distortion trials were presented in the main experiment. Unlike in Experiment 3.1, participants were instructed to respond as quickly and accurately as possible as soon as the object appeared in each test phase trial. On the object-scene category matching post test, all 36 object prototypes were shown and asked to be matched with one or more of 6 scene type labels (whereas, in Experiment 3.1 participants responded to only a subset of the objects).

EEG Data Acquisition and Preprocessing

EEG data acquisition and preprocessing was the same for Experiment 3.1. Artifact rejection procedures using subject-specific threshold parameters resulted in average trial loss of 27.0% for the exact match condition, 22.0% for the within-category mismatch condition, and 22.4% for the between category mismatch condition.

Analysis

Effects of match condition and preview time on the brain response and behavior were assessed using mixed effects modeling, similar to Experiment 3.1. However, scene preview time at test was now treated as a continuous within-subjects predictor. Also, because participants were asked to respond immediately in this experiment, the impact of match condition and preview time on response time was also assessed.

Results

Behavioral – Online Accuracy

Initial behavioral analyses were conducted to screen participants for poor behavioral performance. All but two participants included in the study showed significant sensitivity to the match vs. mismatch contrast, collapsing across mismatch type and 'possible' vs. 'impossible' mismatch response (Pearson's Chi-squared test for 2 x 2 contingency table, p's < .01). The remaining two participants were sensitive to between category mismatches vs. other trials, but responded similarly to matches and within category mismatches, again collapsing across 'possible' and 'impossible' mismatch responses (Pearson's Chi-squared test for 3 x 2 contingency table, p's < .01).

For subsequent online behavioral analyses, trials were excluded for which no response was registered within the first 5000 ms of object onset (37 total trials across the experiment, 0.36% of the data). Mean accuracy was 76.1%, range 58.5-96.9%. Participants as a group were sensitive to the type of mismatch and were more likely to respond that between category mismatches were 'impossible' than that within category mismatches were 'impossible' (intercept = -1.742, β = 1.492, SE = .197, z = 7.587, χ_1^2 = 34.12, p < .001). Figure 3.8a shows the mean response distribution across subjects for each condition. The probability of responding correctly to match trials increased with increasing scene preview duration (β =.161, SE = .069, z = 2.337, χ_1^2 =5.44, p < .05). However, sensitivity to mismatch type did not improve with increasing scene preview (based on two models respectively predicting probability of a 'possible' or 'impossible' mismatch response, examining interactions between scene preview duration and each mismatch condition ID variable, |z|'s < 1).

Behavioral – Response Time

Linear mixed effects models were fit to test whether log response time was predicted as a function of match condition, response, mean centered scene preview duration (in seconds), and their interactions. Following the ERP analyses for Experiments 3.1 and 3.2, for all models random intercepts of item (scene + object) and subject were included, as well as by-subjects random slopes of condition, response, and their interaction. Nested model comparisons were used to test individual fixed effects, in a forward model selection procedure. Only behaviorally correct trials were examined.

Within category and between category mismatches were both responded to more slowly than exact matches (within – exact: β = .250, SE = .019, χ_1^2 = 62.07, p < .001; between – exact: β = .197, SE = .017, χ_1^2 = 54.41, p < .001). 'Possible' responses did not differ reliably from 'impossible' responses (χ_1^2 = 1.62, p > .1). Within category mismatches were responded to more slowly than between category mismatches (β = .053, $F_{(1,89.9)}$ = 16.56, χ_1^2 = 11.54, p < .001).

The more time participants had to preview the context scene prior to test object onset, the faster their response times (β = -.096, SE = .006, χ_1^2 = 256.0, p < .001). This was modulated by match condition (χ_2^2 = 14.51, p < .001). Longer preview times sped up the confirmation of match trials more than the rejection of mismatch trials (effect of preview time on match trials: β = -.129, SE = .011; decrease in preview facilitation for mismatches relative to matches: within category, β = .041, SE = .014; between category, β = .051, SE = .014). Within mismatches, there was no significant further modulation of the preview facilitation effect by whether a 'possible' or 'impossible' response was given (χ_1^2 = 2.00, p > .1). Figure 3.8b plots response time as a function of scene preview duration and match condition.

Behavioral – Posttest Categorization

As in Experiment 3.1, participants showed sensitivity to the association between object and scene types at study. Figure 3.8c shows the normalized confusion matrix indicating the probability that a scene category, if circled, belonged to the correct scene category for the depicted object. This was again assessed with a logistic regression model predicting the probability of circling a scene category, with a fixed effect of match, crossed random intercepts for subject, object (response item) and scene (response choice), a by-subjects random effect of match, and nested model comparisons. As in Experiment 3.1, there was a significant effect of

matching the correct scene type on the probability of circling a response option (intercept = -3.45, β = 5.85, SE = .451, z = 12.99, χ_1^2 = 61.55, p < .001).

ERP Analysis: Match by Preparation Time

We assessed whether the match/mismatch effect would be enhanced by giving the participant more preparation time (i.e., a longer scene preview) at test. In contrast with Experiment 3.1, we now used a within-subjects design and a continuous measure of scene preview duration. Linear mixed effects models predicting mean EEG amplitude were fit to the individual trial data for each of the four components of interest used in Experiment 3.1 (targeting the N300, N400, early LPC, and late LPC). Models included the following fixed effects:

(1) match condition (contrasting match, within, and between category conditions)
(2) response type (for within and between category mismatch conditions only); condition (within vs. between) was moderately associated with response ('possible' vs. 'impossible' mismatch); Cramér's V = .22.
(3) the interaction between match condition and response
(4) mean centered scene preview duration as a continuous linear predictor, in seconds
(5) the interaction between match condition and scene preview duration
(6) the 3-way interaction between match condition, response, and scene preview duration

The same random effects structure was used as for Experiment 3.1, with random intercepts of subject, item (scene+object), and channel, and by-subjects random slopes of match condition, response, and match condition x response. Our attempt to include a by-subjects random slope of scene preview duration led to convergence issues.

All four component models revealed a numeric or significant three-way interaction between match condition, response, and scene preview duration (N300: χ_2^2 = 5.57, p < .1; N400: χ_2^2 = 89.77, p < .001; early LPC: χ_2^2 = 54.11, p < .001; late LPC: χ_2^2 = 39.08, p < .001). Therefore, the interaction between scene preview duration and the mismatch – match effect is reported separately by mismatch and response type (within vs. between category mismatch type x 'possible' vs. 'impossible' response). To aid in interpretation, corresponding mismatch – match

effect estimates at the mean scene preview duration are also reported. Because changing scene preview duration led to more or less overlap between the scene and object elicited ERPs, we do not attempt to interpret the overall effect of scene preview duration.

N300

At the mean scene preview duration (1264 ms), between category mismatches were more negative than exact matches, across response type ('possible': diff = -1.21 μV, $F_{(1,38.8)}$ =4.98, p < .05; 'impossible': diff = -1.33 μV, $F_{(1,38.2)}$ = 4.62, p < .05). Within category mismatches did not differ significantly from between category mismatches (F's < 1) and were numerically more negative than matches for both response types ('possible': diff = -.83 μV, $F_{(1,32.2)}$ =2.58; 'impossible': diff = -.94 μV, $F_{(1,30.0)}$ = 1.90; p's > .1).

The mismatch – match effect was larger (more negative) given a longer preview duration, for both mismatch types and both response types ('possible' within category mismatch: β = -.908 μV/s, $F_{(1,94771)}$ = 69.9, p < .001; 'impossible' within category mismatch: β = -.781 μV/s, $F_{(1,90640)}$ = 24.1, p < .001; 'possible' between category mismatch: β = -1.061 μV/s, $F_{(1,5704)}$ =11.9, p < .001; 'impossible' between category mismatch: β = -1.416 μV/s, $F_{(1,5747)}$ = 21.0, p < .001).

The interaction of N300 effect size and scene preview duration is of particular interest in the current study. We took advantage of our treatment of scene preview duration as a continuous parameter spanning 0-2500 ms, in order to get an approximate estimate of the amount of contextual pre-exposure time needed to show a reliable mismatch – match effect on the N300, given our design. Thus, we fit a separate mixed effects model predicting N300 amplitude, this time treating preview time as an unordered binned predictor with the following factor levels (in ms): [0, 500], (500, 1000], (1000, 1500], (1500, 2000], (2000, 2500]. For each preview time bin, we then computed the unweighted average of four mismatch – match effects: within category 'possible' response, within category 'impossible' response, between category 'possible' response, between category 'impossible' response. By taking the unweighted average, we compensated for any fluctuations in the proportions of mismatch type and response type across preview time bins. This gross measure of the N300 mismatch – match effect became significant and remained so for scene preview durations greater than approximately 1000 ms ([0,500] ms: diff = -.56 μV, F < 1; (500,1000] ms: diff = -2.50 μV, $F_{(1,73.3)}$ = 1.87, p > .1; (1000,1500] ms: diff = -4.06 μV, $F_{(1,71.7)}$ = 4.97, p < .05; (1500,2000] ms: diff = -8.25 μV, $F_{(1,69.6)}$ = 20.9, p <

.001; (2000,2500] ms: diff = -6.01 μV, $F_{(1,XX.X)}$ = 10.8, p < .01). See Figure 3.10b for a graphical illustration.

N400

At the mean scene preview duration, mismatches were estimated to be more negative than exact matches, regardless of mismatch type or response type ('possible' within: diff = -1.08 μV, $F_{(1,32.8)}$ = 5.38, p < .05; 'impossible' within: diff = -2.51 μV, $F_{(1,26.7)}$ =11.5, p < .05; 'possible' between: diff = -1.85 μV, $F_{(1,34.3)}$ = 6.77, p < .05; 'impossible' between: diff = -1.06 μV, $F_{(1,38.9)}$ =3.29, p < .1). Within and between category mismatches did not differ significantly for either response type ('possible': $F_{(1,34.3)}$ = 1.52; 'impossible': $F_{(1,30.9)}$ = 2.80; p's > .1).

As was true for the N300, for the N400 the mismatch – match effect was larger (more negative) given a longer preview duration for both mismatch types and both response types ('possible' within category mismatch: β = -.344 μV/s, $F_{(1,46172)}$ = 5.23, p < .05; 'impossible' within category mismatch: β = -2.585 μV/s, $F_{(1,44455)}$ =137.8, p < .001; 'possible' between category mismatch: β = -.938 μV/s, $F_{(1,5220)}$ = 7.07, p < .01; 'impossible' between category mismatch: β = -1.140 μV/s, $F_{(1,5252)}$ =10.4, p < .01).

Early LPC

At the mean scene preview duration, mismatches were generally more negative than matches ('possible' within: diff = -1.49 μV, $F_{(1,32.3)}$ = 15.7, p < .001; 'possible' between: diff = -1.62 μV, $F_{(1,39.4)}$ = 6.64, p < .05); 'impossible' within: diff = -1.26 μV, $F_{(1,29.1)}$ = 3.52, p < .1). However, 'impossible' between category mismatches were numerically more positive than exact matches (diff = .34 μV, F < 1), and significantly more positive than 'impossible' within category mismatches (diff = 1.60, $F_{(1,36.7)}$ =4.86, p < .05).

Effects did not vary by preview time for items that were given a 'possible' response ('possible' within: $F_{(1,45455)}$ = 1.10; 'possible' between: F < 1; p's > .1). For 'impossible' within category mismatches, the mismatch – match effect was more negative given longer scene preview (β = -1.692 μV/s, $F_{(1,43048)}$ =59.2, p < .001) and the same pattern was seen numerically for 'impossible' between category mismatches (β = -.559 μV/s, $F_{(1,5174)}$ =2.83, p < .1). The effect of scene preview on 'impossible' within category mismatches was larger than on 'impossible' between category mismatches (β = 1.133 μV/s, $F_{(1,7811)}$ = 9.21, p < .01).

Late LPC

At the mean scene preview duration, the only significant match effect was that 'possible' within category mismatches were estimated to be more negative than exact matches (diff = -1.24 μV, $F_{(1,33.3)}$ =10.9, $p < .01$) and numerically also more negative than 'possible' between category mismatches (diff = 1.04 μV, $F_{(1,33.7)}$ =3.52, $p < .1$). No other condition differed from exact matches ('possible' between: diff = -.20 μV, $F < 1$; 'impossible' within: diff = -.01 μV, $F < 1$; 'impossible' between: diff = 1.36 μV, $F_{(1,38.8)}$ = 3.50, $p < .1$).

'Impossible' mismatch – match contrasts became more negative with increasing scene preview time, but not 'possible' mismatch – match contrasts ('impossible' within: β = -1.009 μV/s, $F_{(1,45005)} = 20.4$, $p < .001$; 'impossible' between: β = -.845 μV/s, $F_{(1,5187)} = 6.27$, $p < .05$; 'possible' within: β = -.090 μV/s, $F < 1$; 'possible' between: β = .230 μV/s, $F < 1$). Effects of preview time were not modulated by mismatch type, holding response type constant (F's < 1).

Discussion

Earlier ERP measures (targeting the N300 and N400) were robustly sensitive to category-level mismatches between the presented and target object at test. Specifically, mismatching objects elicited a more negative waveform than matching objects at these latencies. Moreover, this sensitivity increased when participants were given more time to process the context scene prior to the onset of the test object, replicating the pattern shown between subjects in our analysis of Experiment 3.1. We further estimated that N300 match effects emerged and stabilized after roughly 1000 ms of contextual pre-exposure time given the current design and sample size.

Later ERP measures (targeting the early and late LPC) showed a more complex pattern of sensitivity, reflecting the interaction of stimulus and response related processing. When participants ultimately decided a mismatch was 'impossible' given the context scene, to the extent that they were given more time to process the scene in advance, the mismatch – match effect became more negative in this interval. However, when they ultimately decided the mismatch was 'possible,' the duration of contextual preview no longer affected amplitude of the mismatch – match effect.

General Discussion

In two separate experiments, we tested a necessary corollary of the hypothesis that contextual facilitation effects on object recognition are partially mediated by predictive pre-activation of object features. Specifically, we tested whether (1) contextual priming of the visual features of an object associated with a scene is reduced when participants have little time to preview the scene before object presentation and (2) whether the amount of priming from a scene onto its associated object increases with duration of scene preview. We found that viewing the context scene in advance does in fact result in enhanced visuo-structural priming of the associated object, as indexed by N300 facilitation. Moreover, graded effects of visual similarity between the expected object and the actually presented one are linked to a longer contextual pre-exposure time. In our between subjects analysis (200 vs. 2500 ms scene preview), we found evidence for delays in processing the degree of mismatch between the presented object and the contextually congruent target, given a shorter contextual pre-exposure time. ERP match/mismatch effects were more prevalent later in the waveform given a shorter scene preview duration, and response-related processing was also attenuated on later components, consistent with a delay in shifting from the task of matching the object to the scene to the task of selecting a response. In our within-subjects manipulation (Experiment 3.2), we again found that longer contextual pre-exposure times were associated with greater N300 and N400 facilitation for contextually congruent objects. We also found converging evidence from response times and ERP match/mismatch effects, suggesting that participants require approximately 1000-1500 ms of contextual pre-exposure time to show early effects of contextual benefit (i.e., N300 facilitation for contextually congruent objects). Overall, our findings suggest that participants use additional time when viewing a scene context alone to prepare to visually process and categorize an upcoming visual object. Below, we discuss the impact of scene-object congruency on different aspects of object processing, as assessed through behavioral measures and ERPs. We then detail the impact of preview time on each and discuss the implications of those patterns for predictive accounts of contextual facilitation on visual object processing. Finally, we consider factors relevant to assessing the likelihood that our findings will generalize to more naturalistic real-life situations and task demands.

First we review evidence for successful associative learning, and for an overall contextual congruency benefit, which can be further broken down by dependent measure to form stronger ties to specific aspects of cognition. In both of the current experiments, as in Experiment 2.1, participants demonstrated that they were able to rapidly associate novel objects and scenes. Not only were participants successful in their online memory task, but they also demonstrated explicit knowledge of which object types were associated with which scene types on the offline post-test.. The early and late LPC reflect increasingly response and task-oriented activity (Mostert, Kok, & De Lange, 2015). As we found previously, within category mismatches were more negative than between category mismatches on the late portion of the LPC in both experiments, reflecting greater difficulty in determining the appropriate response for possible but unstudied scene-object combinations (e.g., Finnigan, Humphreys, Dennis, & Geffen, 2002). This interpretation is also supported by response time data in Experiment 3.2, which showed slower overall responding to within than between category mismatches. Earlier in the waveform, matches tended to be more positive than mismatches in both the N400 and N300 time windows. While the N400 is associated with multimodal semantic priming (reviewed in Kutas & Federmeier, 2011), the N300 is associated with an earlier stage of visuo-structural priming (e.g., Schendan & Kutas, 2002, 2003, 2007; Hamm, Johnson, & Kirk, 2002; Kovalenko, Chaumon, & Busch, 2012). Match facilitation on earlier portions of the ERP waveform (in particular, the N300 component) is consistent with the idea that contextual congruency facilitates not only response-related processing and categorical judgments, but also higher-level visual analysis of the (contextually associated or unassociated) object itself, including visuo-structural priming. The next question, then, is to what extent this priming is contingent on an extended scene preview, which would be consistent with a predictive pre-activation account of context effects on object recognition.

Across the two experiments, shorter scene preview times were associated with diminished or delayed contextual congruency effects. Specifically, shorter scene previews were associated with smaller match/mismatch effects on earlier components: the N300 and N400 (although the N400 only showed this pattern in Experiment 3.2). As is nicely captured in Figure 3.6, in our between subjects comparison (Experiment 3.1 vs. Experiment 2.1), shorter scene preview times were also associated with larger match/mismatch effects in the early LPC window (500-699 ms), consistent with a global latency shift in the peak match/mismatch response. In

Experiment 3.2, which used a parametric, within-subjects design, we also observed an interaction between scene preview duration and decision related processing on the early and late LPC, although it differed qualitatively from Experiment 3.1 and was sensitive to response, as discussed further below. Increased scene preview duration was also associated with benefits in accuracy and speed of responding in Experiment 3.2. Below we further discuss the time-dependent enhancement of contextual benefit effects on behavior, decision-related brain responses (the LPC), and semantic and visual form priming, each in turn.

We observed behavioral effects of contextual pre-exposure time, when immediate responding was required (Experiment 3.2). With delayed responding, the distribution of responses was similar across preview times (Experiment 3.1 vs. Experiment 2.1). With immediate responding (Experiment 3.2), however, correct responses were faster overall given longer scene previews, and particularly so for matching objects. Moreover, the probability of correctly identifying a match also increased with increasing contextual pre-exposure time (a similar increase in accuracy was not observed for mismatch trials). This combined increase in accuracy and decrease in response time for match trials cannot be explained by a simple speed-accuracy trade-off. The overall pattern of shorter and more accurate responding with extended scene preview duration is consistent with facilitated object processing given more time to process the scene, and thus, is consistent with a predictive pre-activation account. The interaction between preview benefit and match condition is more challenging to interpret but still of some interest, because it provides further indirect evidence supporting the hypothesis that the representation of the most likely object is selectively enhanced by pre-exposure to the scene. Broader findings from the visual attention literature have indeed suggested that people are able to hold the attentional template of a particular target in mind (in this case, the contextually congruent object), but are less successful at ruling out mismatches based on a 'negative' template (e.g., Robinson, Clevenger, & Irwin, 2018). More quantitative modeling inspired by the visual attention literature, and more closely integrating decision theory, with separate estimates for the amount and nature of information required to reach a 3-way decision bound for matches, close and distant mismatches, may be required to fully understand what implications this pattern of responding has for processing in the brain. Now that the behavioral data has confirmed that object processing is indeed facilitated by longer contextual previews, we turn to the specific

pattern of ERP facilitation effects, which helps answer the question of which specific stages of processing are facilitated by longer contextual pre-exposure times.

The LPC was sensitive to preview time across both experiments, but with some qualitative differences. As previously mentioned, the late LPC was sensitive to match condition, such that within category mismatches were more negative than between category mismatches. Within category mismatches were more likely to elicit a 'possible' response than between category mismatches in both experiments. Possibly partially due to this, in Experiment 3.1, there was no interaction with response (as distinct from mismatch type) on the late LPC. However, in Experiment 3.2, at mean scene preview duration, it was found that only within category mismatches that elicited a 'possible' response were more negative than the other match conditions, while within category mismatches that were responded to as being 'impossible' had similar late LPC amplitudes to exact matches and were only numerically more negative than 'impossible' between category mismatches. When scene preview duration was lengthened in our between subjects comparison (Experiment 3.1 vs. Experiment 2.1), longer scene preview was associated with a stronger effect of response on the late LPC, such that 'possible' responses were associated with a more negative waveform in Experiment 2.1 only. There was no interaction between scene preview duration and match condition. In contrast, for Experiment 3.2, there was a three-way interaction between match condition, response, and scene preview duration on the late LPC: match/mismatch effects grew larger with increased scene preview, only on trials responded to as 'impossible.' Notably, the correlation between mismatch type and response type was somewhat stronger for Experiment 3.1 (Cramér's V = .43) than either Experiment 3.2 (Cramér's V = .22), or Experiment 2.1 (Cramér's V = .27). Thus, it is possible that it is simply more difficult to tease apart effects of response vs. mismatch type in Experiment 3.1 than the other two experiments due to higher multicollinearity between these predictors. However, another more interesting possibility is that response-related processing itself may be delayed under conditions where there is both delayed responding and an inadequate scene preview time, as in Experiment 3.1.

On the early LPC, with speeded responding (Experiment 3.2), we again saw an interaction between match and response at mean scene preview duration: 'possible' mismatches were reliably more negative than the match condition, but not 'impossible' mismatches. The three-way interaction with scene preview duration was similar to the late LPC. With delayed

responding, short scene previews (Experiment 3.1) were associated with an effect of match but not response (both mismatch conditions were about equally more negative than the match condition). Long scene previews (Experiment 2.1) were associated with an effect of response type but not match: 'possible' mismatch responses were more negative than 'impossible' mismatch responses. In contrast with Experiment 3.1, both mismatch types were numerically slightly more positive than matches overall, and this crossover interaction makes it unlikely that the different patterns observed across scene preview durations could be explained by higher multicollinearity between condition and response in Experiment 3.1. Direct between-subjects statistical comparisons confirmed the observed differences across scene preview durations. Thus, as with the late LPC, results from the early LPC window are consistent with a delay in response selection when responses are not immediate and insufficient scene preview time (200 ms) is given. However, in the current study we are even more interested in the effect that extended contextual pre-exposure may have on earlier components: the N400 and N300.

On the N400, an earlier index of semantic processing, we observed contextual facilitation effects that were robust even at short preview times. Mismatches were consistently more negative than matches on the N400, across all experiments. For Experiment 3.2 only, with speeded responding, we also observed an interaction between match condition and response in the N400 time window, again consistent with speeded responding leading to earlier response selection. In our between subjects comparison, there was no interaction of either match condition or response with scene preview duration. In Experiment 3.2, match-mismatch effects did increase in size given longer scene previews. Overall, the relative robustness of the N400 match effect suggests consistent contextually-sensitive semantic processing of the object images, with or without an extended contextual preview. The Experiment 3.2 results further suggest that under at least some conditions, semantic processing of object images may be further enhanced by contextual pre-exposure.

Of particular importance to the current study was the question of whether N300 match/mismatch effects, an index of visuo-structural priming, would be modulated by contextual pre-exposure time. This is because scene priming of object-specific features (in contrast with contextual bias effects on categorization and decision processes) is a tighter, more targeted corollary of the hypothesis that people predictively pre-activate visual features of upcoming objects based on scene information. Although we did observe an N300 effect of match condition

given only a 200 ms scene preview, this was driven primarily by recency. That is, while there were numeric trends for other kinds of match/mismatch effects in the expected direction, the only statistically reliable effect of match condition was that images of objects that were not displayed at all in the preceding study phase elicited larger N300 amplitudes than contextually matching objects. In our between subjects comparison (combining Experiment 3.1 with the dataset in Chapter 2 Experiment 2.1), we confirmed that the size of N300 match/mismatch effects was modulated by contextual pre-exposure time. Specifically, between-category mismatches elicited larger amplitude N300s than matches, but particularly so given a longer (2500 ms) scene preview duration. In Experiment 3.2, we replicated this pattern: N300 match/mismatch effects were larger given a longer scene preview. Moreover, we found that N300 match/mismatch effects only became significant using the current design and sample size given at least 1000-1500 ms of scene preview. We thus have strong support for the idea that scene-object priming on the N300 is at least partially dependent on contextual pre-exposure time.

There were a few discrepancies between the results of our between-subjects (Experiment 3.1) and within-subjects (Experiment 3.2) analyses of the N300. For example, there was still some N300 priming by congruent scene contexts given a consistent 200 ms scene preview in Experiment 3.1, but N300 priming was only apparent with longer (1000 ms +) scene preview times in Experiment 3.2. This could partly be explained by the fact that N300 effects in Experiment 3.1 were driven by target recency (i.e., 'new' between category mismatches), while the design of Experiment 3.2 removed the recency confound. (Notably, recency effects alone could not explain N300 priming given a constant 2500 ms scene preview in Experiment 2.1, either). In addition, the power to detect N300 effects at short scene previews may have been enhanced in a design with a consistent temporal relationship between scene and object onset. The fact that we detected item recency effects on the N300 underscores the benefits of avoiding or controlling for item repetition and its potential interactions with the conditions of interest, which, as mentioned in the introduction, may have clouded interpretability of the latency manipulation in Demiral, Malcolm, & Henderson (2012).

We also confirmed that visuo-structural priming is contingent on adequate contextual pre-exposure time in a novel component-neutral analysis. In Chapter 2, we showed for the first time that higher visual similarity between the presented object and the most contextually expected object based on the scene leads to a more positive waveform beginning at roughly 200

ms, consistent with a visuo-structural priming benefit. The predictor of ERP amplitude in this analysis is tightly yoked to visual properties of the target, contextually congruent object, and is derived from V1-like gabor filter features (Pinto et al., 2008). Thus, this analysis allows us to side-step difficulties in disentangling N300 and N400 effects, which partially overlap in time and space. Previous studies examining the question of how contextual pre-exposure benefits visuo-structural priming have generally used component-based analyses where component overlap is likely a partial confound (e.g., Mudrik, Lamy, & Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014; Truman & Mudrik, 2018). In our between-subjects analysis, we found that this index of visuo-structural priming is enhanced given a longer contextual pre-exposure time. In fact, with only 200 ms of scene preview, we only found evidence of a small effect in the expected direction on the N300, despite detecting robust effects across the N300, N400 and LPC in our earlier Experiment 2.1 using a 2500 ms scene preview. This component neutral analysis provides further support for our claim that visuo-structural priming of a contextually congruent object is dependent on adequate contextual pre-exposure time, which is consistent with a predictive pre-activation account.

One important critique of the current study is that an enhancement of contextual facilitation effects given more time to process the context may simply reflect the formation of a more detailed or sophisticated representation of the context itself. This can be differentiated from a predictive pre-activation account, which specifies that features of the associated object representation are brought online prior to object onset. We cannot rule out this alternative hypothesis at present, pending using machine-learning based approaches to better target pre-activation of object-based features during the scene preview period (similar to the approach adopted in Kok, Mostert, & De Lange, 2017). However, our current results have already provided evidence against any theory in which assessing contextual congruency between an object and scene is independent of the duration of prior exposure to the scene. Also, our N300 and visual similarity analyses have more tightly yoked facilitation effects to visual features of the anticipated object, which is more complicated to explain under a non-predictive account than generic facilitation at a semantic level.

Another important critique of the current work is that it is as yet unclear to what extent conscious and explicit memorization of scene-object pairs in a paired associate learning paradigm would generalize to cognitive processes employed during object recognition in daily

life. One argument in favor of generalizability is the fact that global semantic properties of visual contexts remain relatively stable over time, from the perspective of a human observer carrying out day-to-day activities. This temporal stability (on the order of minutes to hours) makes it more likely that not only rapid gist extraction (which takes place during the first 200 ms of scene exposure, recently reviewed in Larson, Freeman, Ringer, & Loschky, 2014), but also slower processes that relate contexts to specific objects over the course of 1 second or more, might contribute to visual object recognition in the real world. The broader literature on statistical learning (and its pervasiveness even at early stages of cognitive development; e.g. Saffran, Aslin, & Newport, 1996; Kirkham, Slemmer, & Johnson, 2002) is also key to addressing the generalizability question (some recent studies applying implicit statistical learning paradigms to the study of visual processing in adults include: Turk-Browne, Scholl, Johnson, & Chun, 2010; Kok, Failing, & de Lange, 2014). We do not yet have strong enough links assessing the precise relationship between explicit paired associate learning paradigms and implicit statistical learning paradigms; however, some preliminary evidence suggests some degree of overlap in the processes used to achieve learning of statistical associations in these two types of tasks (Pearson & Westbrook, 2015). Specifically, clinical populations that show deficits in implicit statistical learning often also show deficits in explicit paired associate learning (Pearson & Westbrook, 2015). The visual search literature also provides some insight into the question of generalizability. It's long been known that statistical regularities between contextual factors and target locations facilitate visual search for more abstract stimuli (reviewed in Jiang & Chun, 2003) as well as objects embedded in more naturalistic scenes (reviewed in Wolfe, Võ, Evans, & Greene, 2011), and that semantically incongruous objects in a scene attract more and longer duration fixations (e.g., Loftus & Mackworth, 1978; Võ & Henderson, 2009). Recently, it was also shown that contextually incongruent objects in a scene attract longer total viewing times than consistent objects, even when their presence is irrelevant to the task at hand (i.e., finding the letter T, which has been artificially overlaid on top of the image; Cornelissen & Võ, 2017). Similarly, Munneke, Brentari, and Peelen (2013) found that scene consistency effects on object recognition persist regardless of whether the object location has been cued in advance. Together, these studies suggest that scene information may facilitate object recognition in a fairly automatic fashion, independent of specific task demands (although object-object spatial congruency effects may be more easily modulated by attentional cuing, see Gronau & Shachar,

2014). Future work could combine more naturalistic stimuli and task demands with neuroimaging techniques such as EEG/MEG that can provide biomarkers specific to visual form-based priming, as differentiated from amodal categorization and decision-making, to further address the question of generalizability.

In summary, consistent with prior work, scenes were found to facilitate processing for contextually congruent objects at a variety of processing stages, even when the association between object and scene was formed recently. We additionally extended prior work by showing that many scene-object facilitation effects, indexing dissociable cognitive processes, interact with scene preview duration. Of particular interest, our results suggest that visual contexts can be processed in advance of viewing an object to facilitate visual form-based processing of the subsequently presented object, consistent with a predictive preactivation account. The current study thus contributes to a larger body of work testing the boundaries and levels of abstraction used for predictions, as well as their sensitivities to time constraints and task-demands, within a predictive coding account of the brain (e.g., Demiral et al., 2012; Wlotko & Federmeier, 2015; Chow, Lau, Wang, & Phillips, 2018). Through the use of ERPs, we have shown that temporally-contingent context effects do not simply affect semantic processing, or later stage decision-making, but, as demonstrated in our N300 and visual similarity-based analyses, also affect predictive processing within high-level vision itself.

Figures

Figure 3.1. Experimental conditions.

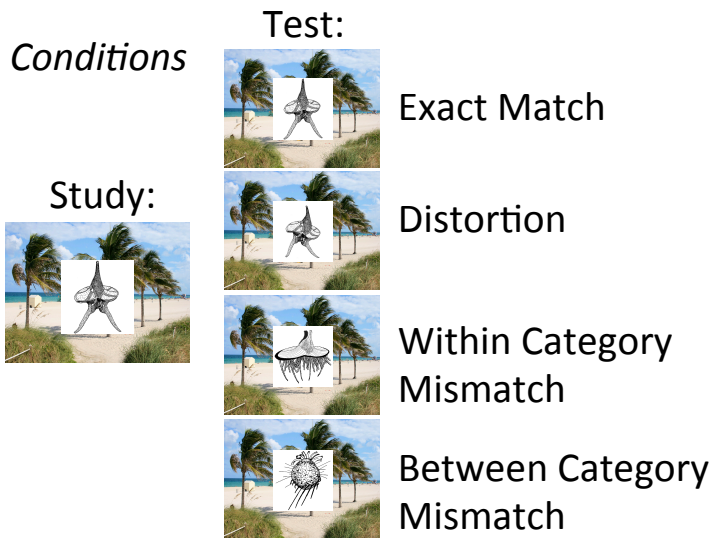


Figure 3.2. Experiment 3.1 Procedure

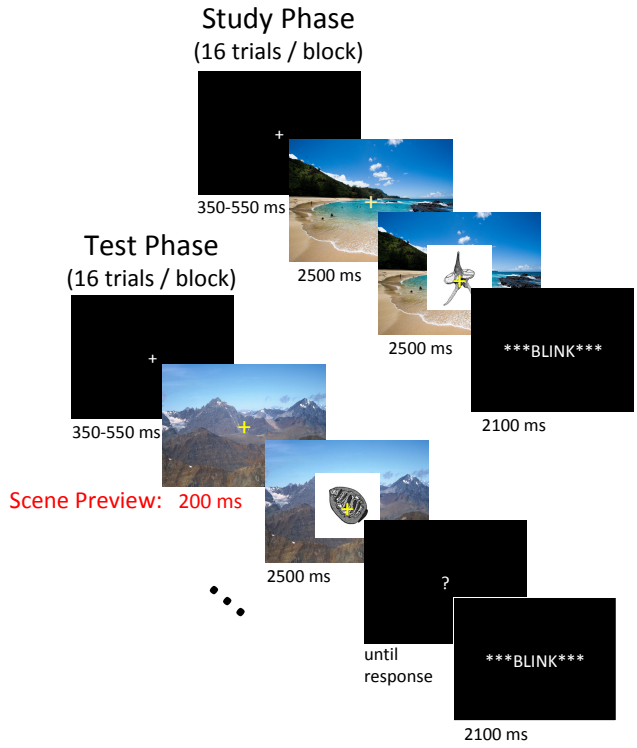


Figure 3.3. Experiment 3.2 Design and Conditions.

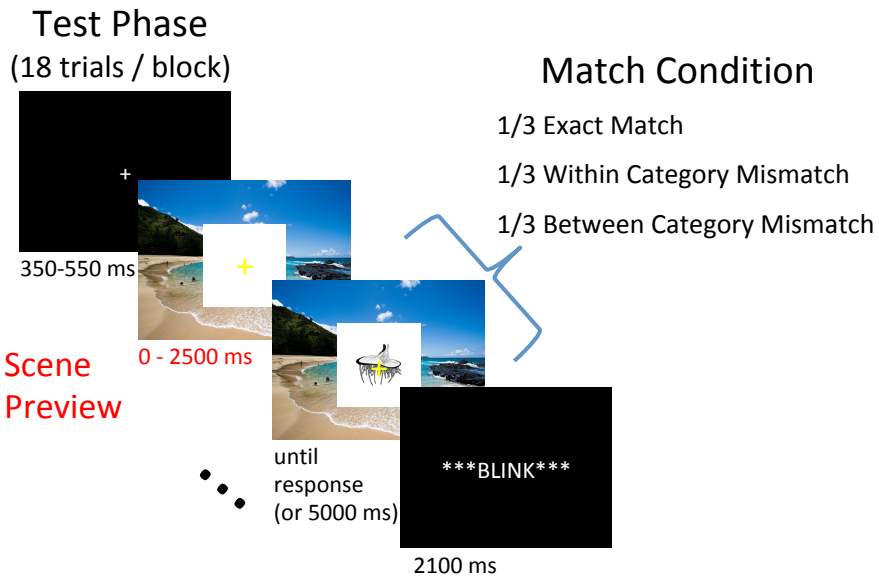
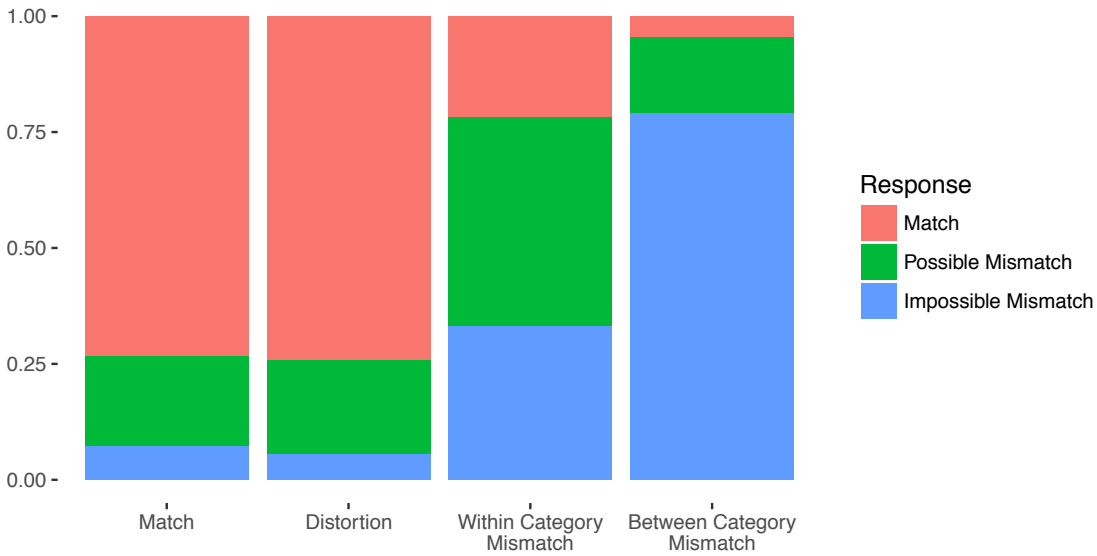


Figure 3.4. Experiment 3.1 behavioral results. a) Proportion of responses by condition in the online memory task. Participants were reliably sensitive to condition, and responded to distortions similarly to the exact match condition, as instructed. b) Confusion matrix of scene-object category associations indicated at post-test. Scenes circled by participants were the associated scene category for the displayed object 82-92% of the time, demonstrating explicit knowledge of the scene – object category mapping.

a)



b)

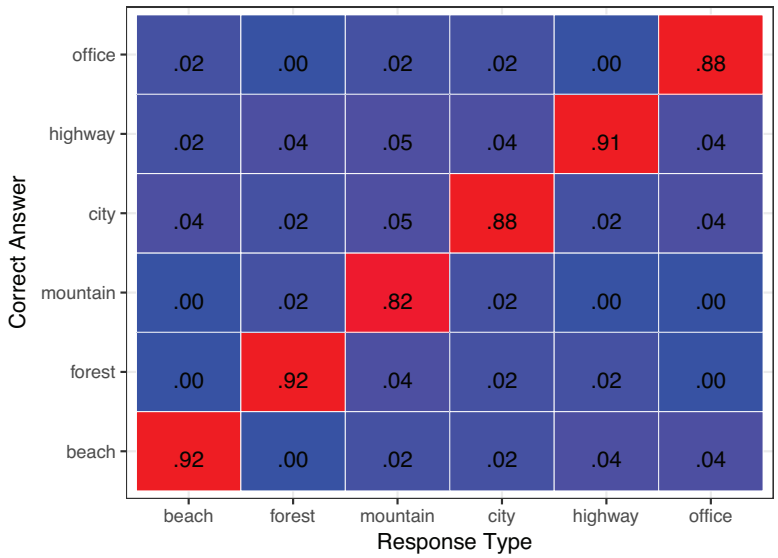


Figure 3.5. Experiment 3.1 match vs. mismatch conditions at 12 representative sites (scalp locations indicated at bottom right). An additional 15 Hz low pass filter was applied after averaging for display purposes.

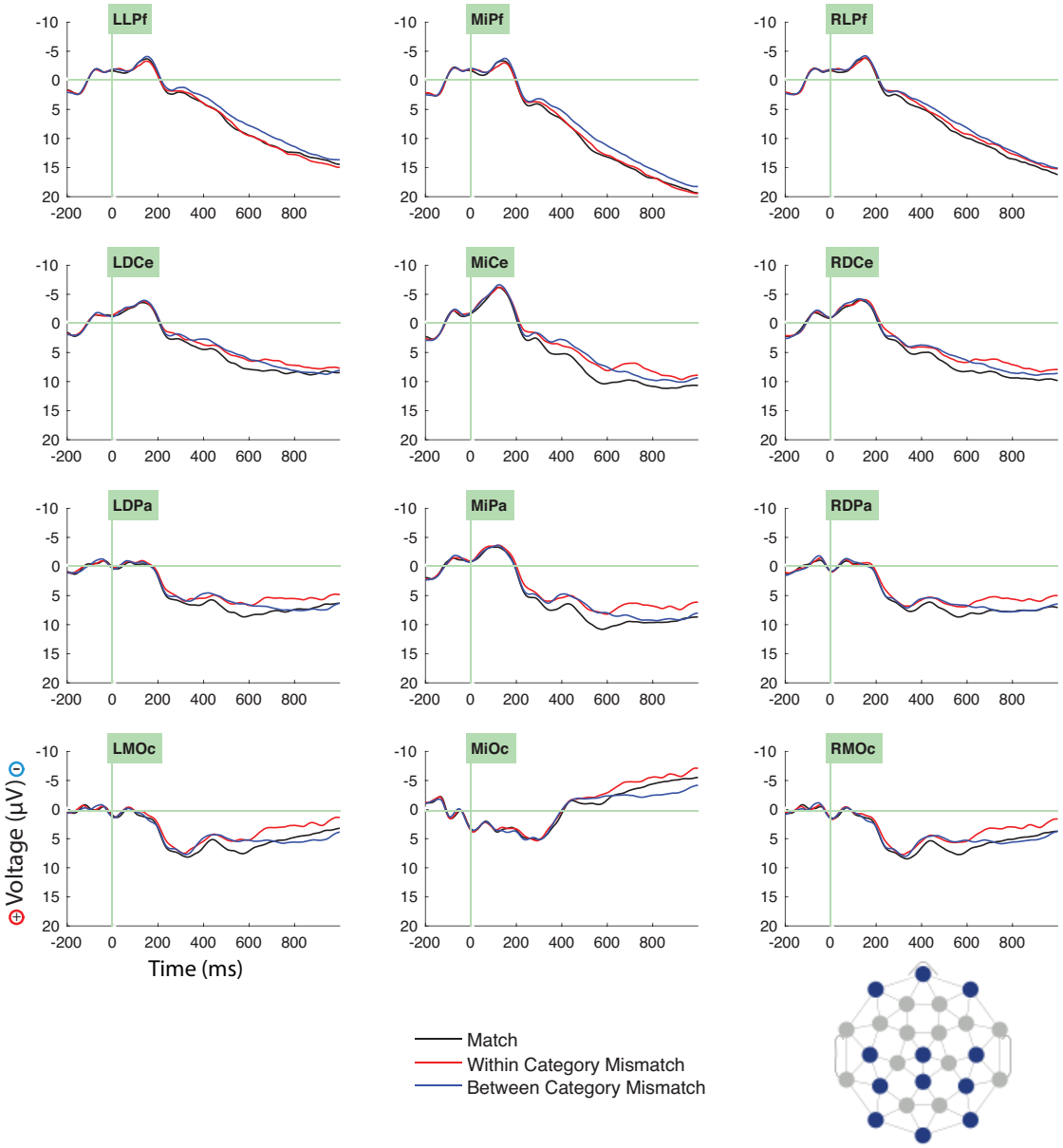


Figure 3.6. Comparing effects of match condition across scene preview durations: Experiment 3.1 vs. Experiment 2.1. Difference waves of within category mismatch – match, and between category mismatch – match, plotted separately by scene preview duration (200 ms vs. 2500 ms). Additional 5 Hz low pass filter applied following averaging for display purposes.

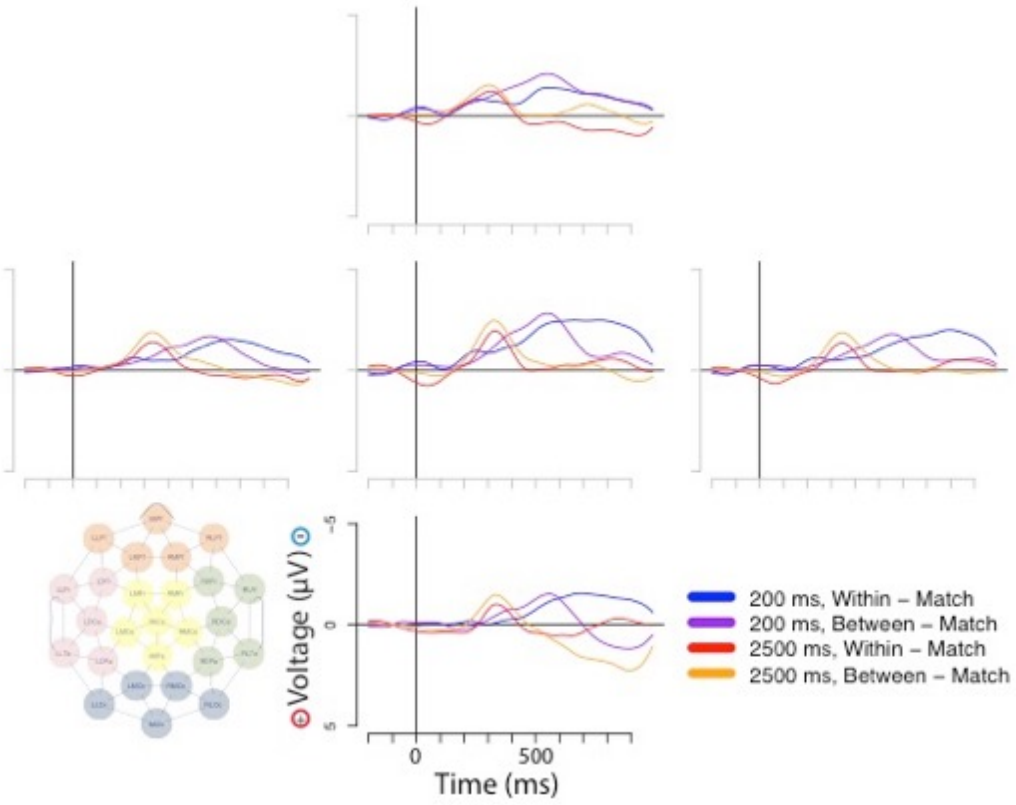


Figure 3.7. Experiment 3.1 ERP averages by visual distance bin. Distance = 0 indicates the Exact Match condition, while higher distances indicate that the presented object was more visually distinct from the target object at test. Only behaviorally correct and artifact free trials included. An additional 5 Hz low pass filter was applied prior to plotting. Sites used to generate each plot are indicated at bottom left.

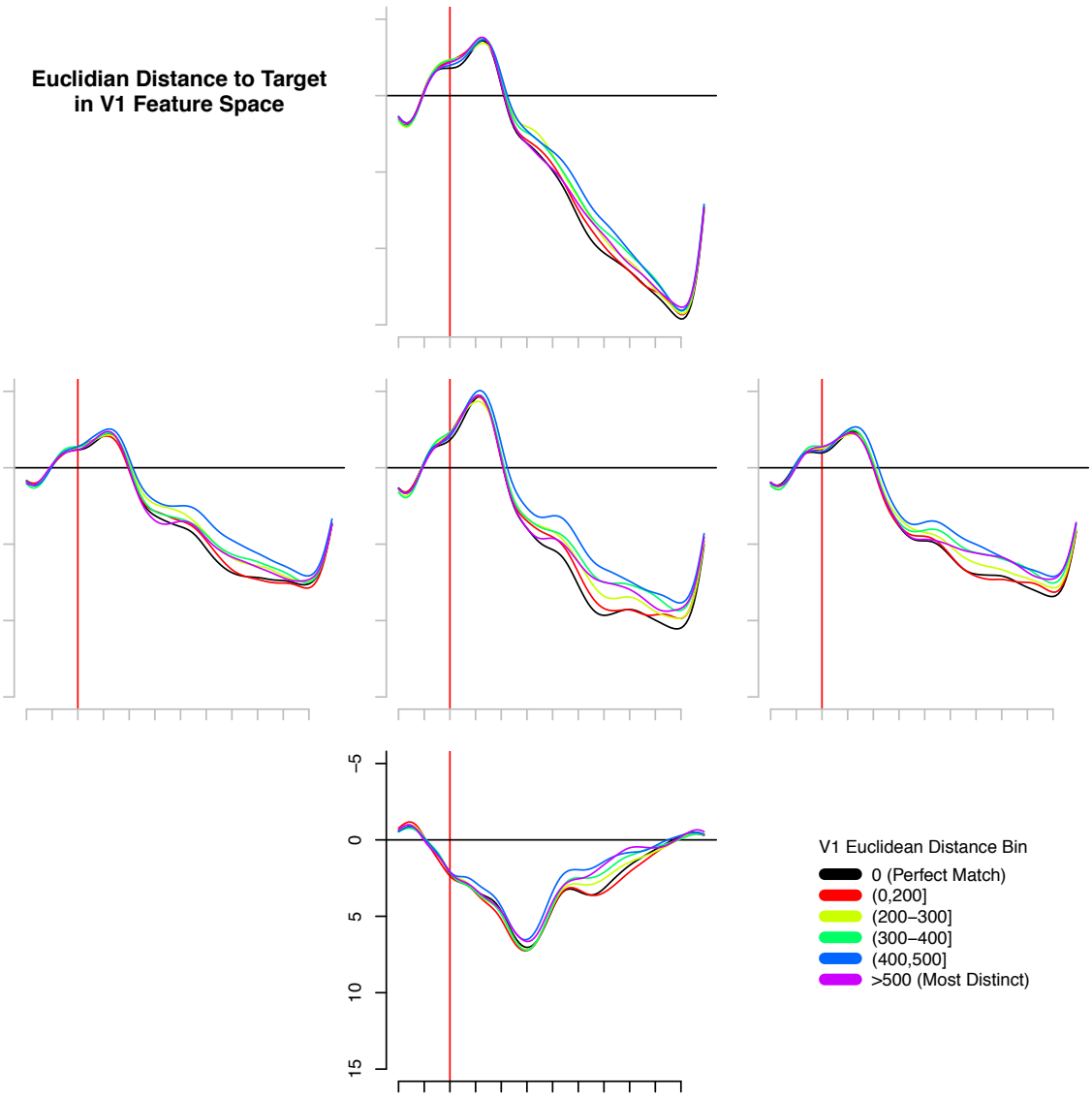
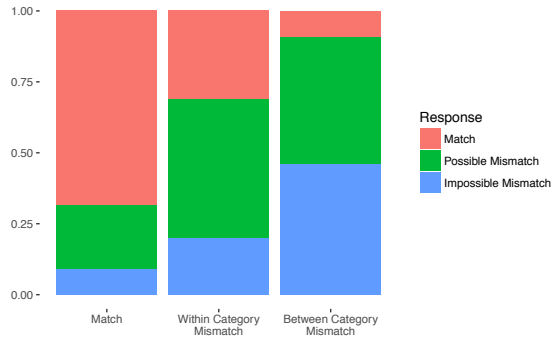
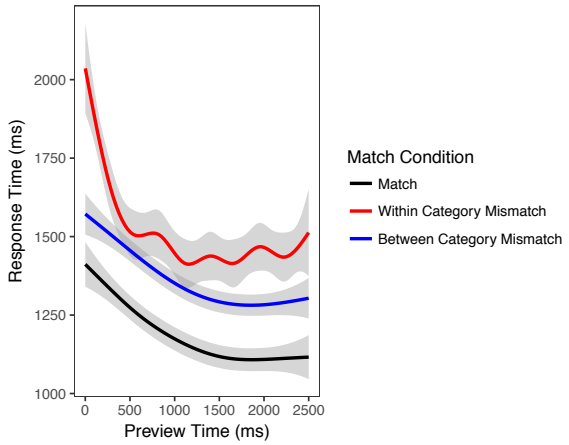


Figure 3.8. Experiment 3.2 behavioral results. a) Proportion of responses by condition in the online memory task. b) Response time is sensitive to match condition and scene preview duration. c) Confusion matrix of scene-object category associations indicated at post-test.

a)



b)



c)

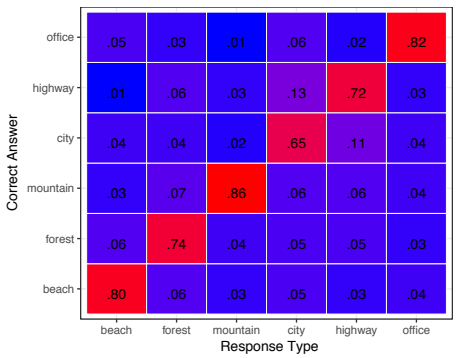


Figure 3.9. Experiment 3.2 test object match vs. mismatch condition ERPs at 12 representative sites (scalp locations indicated at bottom right). An additional 15 Hz low pass filter was applied after averaging for display purposes.

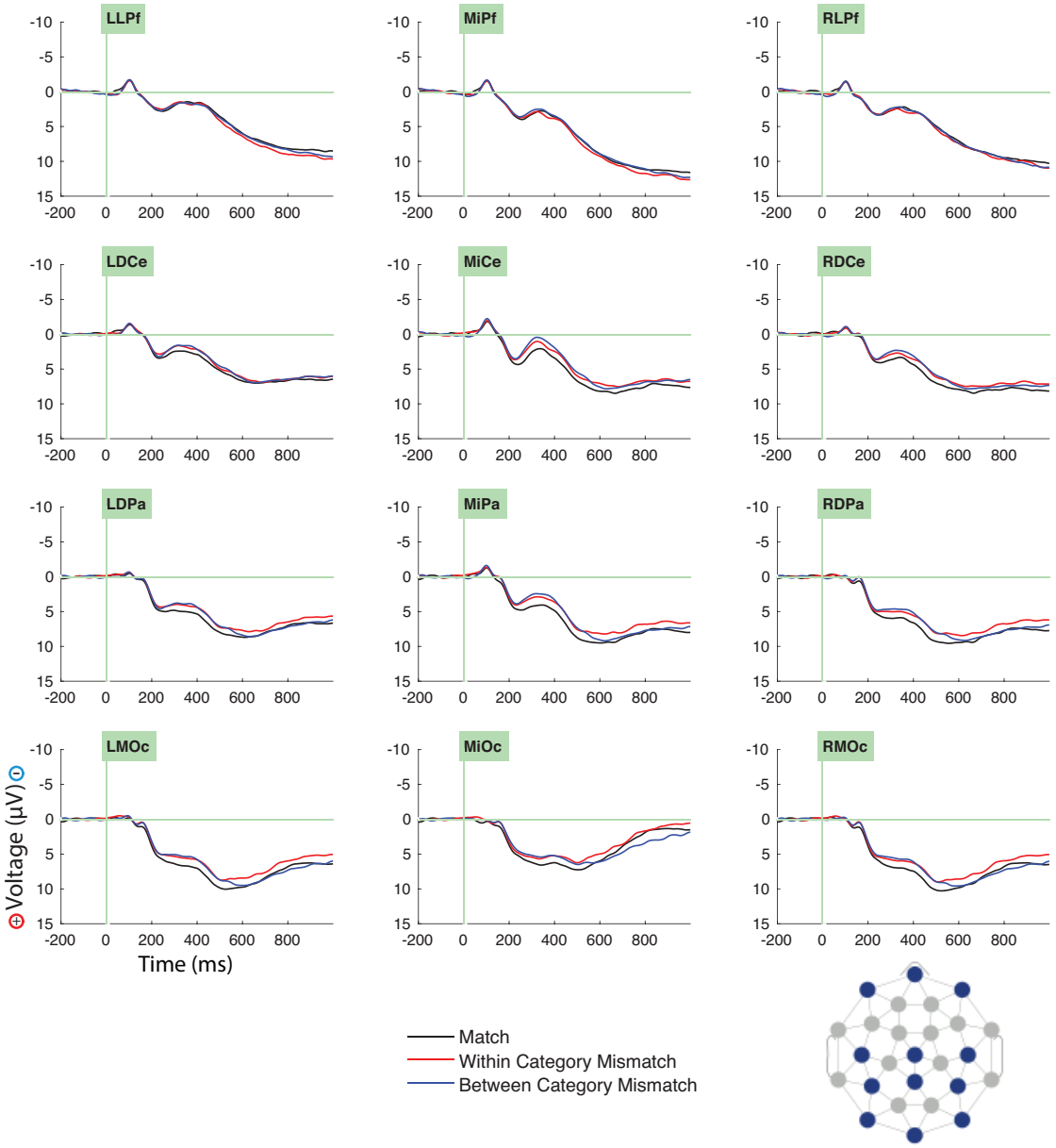
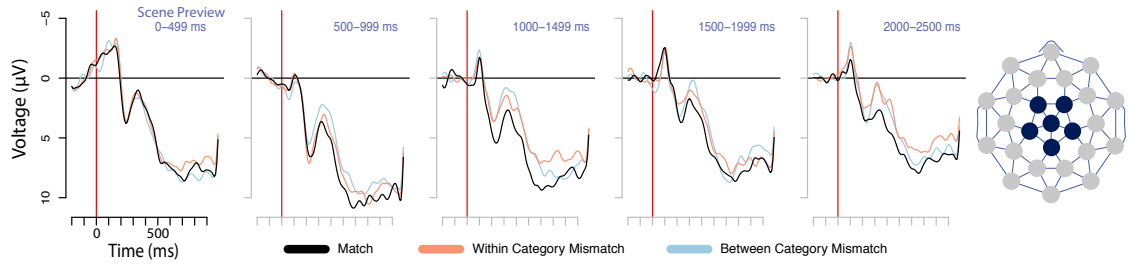


Figure 3.10. Experiment 3.2 ERP waveforms time-locked to test object onset, plotted separately by match condition and scene preview duration (500 ms bins). a) ERP waveforms aggregated across 5 central electrode sites, indicated at right. Additional 15 Hz low pass filter applied after averaging for display purposes. b) Estimated effect size of N300 mismatch – match effect by scene preview duration bin. Estimates computed for each preview time bin as unweighted average of four separate mismatch – match effects: within vs. between category mismatch x 'impossible' vs. 'possible' response. Modeled over 16 frontal channels in N300 time window using linear mixed effects model including 5 x 500 ms scene preview duration bins. Error bars show 2 x approximate standard error.

a)



b)

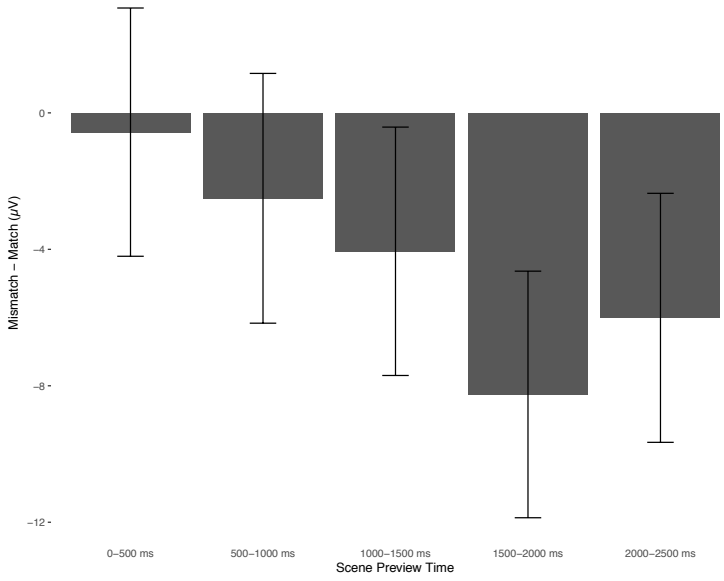


Figure 3.11. Experiment 3.2 within category mismatch – exact match difference waves, plotted separately by scene preview duration (500 ms bins), aggregated across frontal, central, posterior, left lateral, and right lateral electrode sites. Darker shades indicate shorter preview times. Channel locations at bottom left. Additional 5 Hz low pass filter applied after averaging for display purposes.

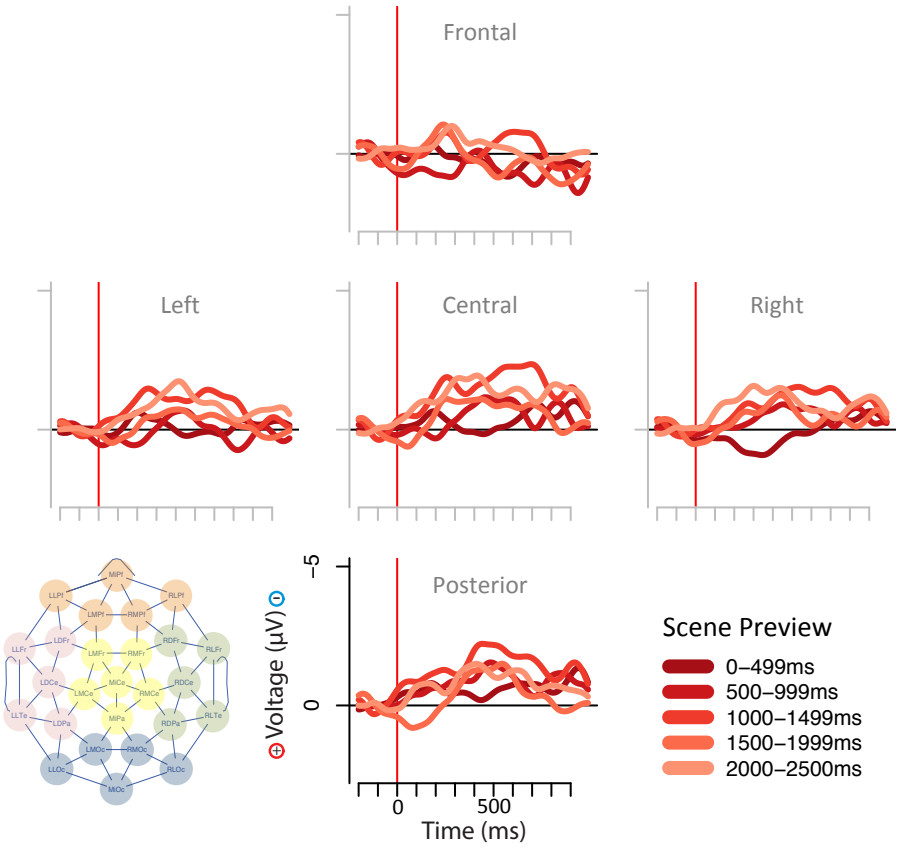
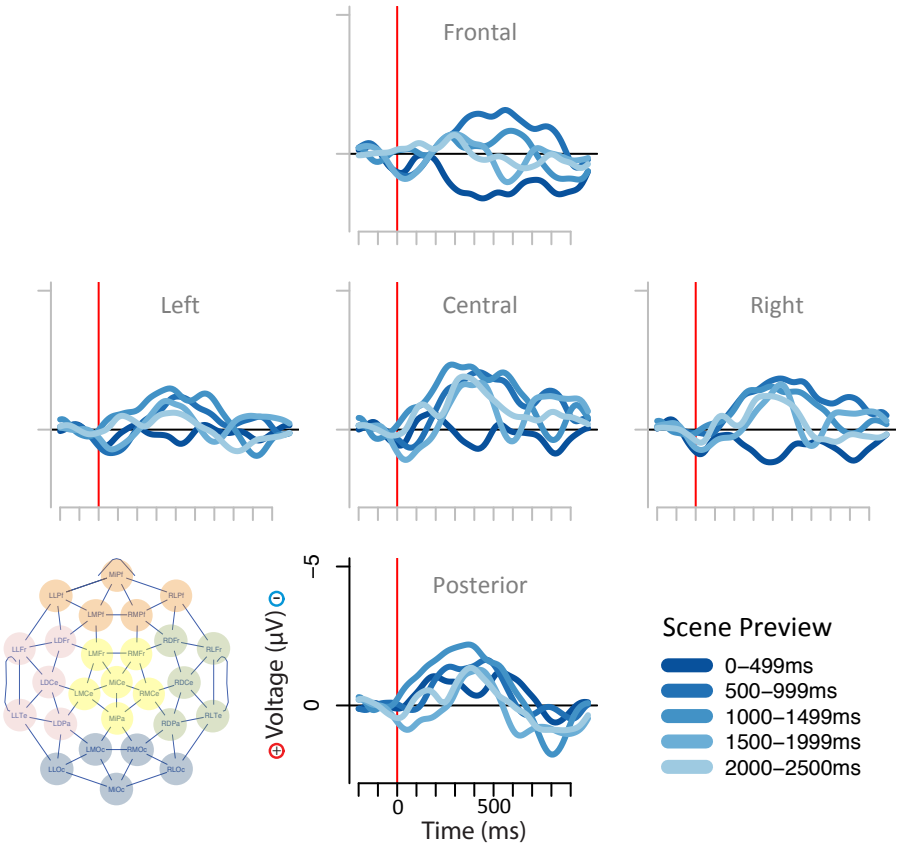


Figure 3.12. Experiment 3.2 between category mismatch – exact match difference waves, plotted separately by scene preview duration (500 ms bins), aggregated across frontal, central, posterior, left lateral, and right lateral electrode sites. Darker shades indicate shorter preview times. Channel locations at bottom left. Additional 5 Hz low pass filter applied after averaging for display purposes.



CHAPTER 4: DISCUSSION

Over three ERP experiments, we explored the nature of the contextual benefit to visual object processing from statistically associated scenes. In Chapter 2 Experiment 2.1, we demonstrated that even novel associations between objects and scenes can result in scene-object priming after a brief training session, including scene-based facilitation for visual form processing of the object. This was shown not only through examination of the N300 component, which is linked to visual form processing, but also through a separate set of component-neutral analyses on mismatch trials using visual similarity effects of the target, expected object vs. the actually presented object to draw inferences about whether the visual form of the target object is selectively activated by the scene. In Chapter 3 Experiments 3.1 and 3.2, we further tested a corollary of predictive processing accounts of context effects on visual object recognition: that longer duration contextual pre-exposure times should enhance contextual facilitation effects. Indeed, that's what we found: across two sets of analyses, longer contextual pre-exposure times were associated with stronger effects of context at an intermediate stage of object processing associated with high-level visual form template matching, indexed by the N300 response. A between-subjects visual similarity analysis confirmed this finding in a component-neutral way. In the following section, we review future directions for research that could further refine our understanding of how context in general, and scene information in particular, facilitates object recognition and categorization.

Future Directions

Using Machine-Learning to Differentiate Predictive vs. Integrative Accounts of Contextual Facilitation Effects

Machine-learning based approaches are increasingly popular in the domain of cognitive neuroscience, and when combined with EEG/MEG, can allow us to make fine-grained inferences when specific representations are brought online (for a recent review and tutorial focused on EEG/MEG applications, see King & Dehaene, 2014). To further differentiate between predictive and integrative interpretations of contextual facilitation effects and their enhancement with

increased exposure to the priming context, future work could apply a machine-learning based approach. Specifically, a classifier predicting the identity of a contextually associated target object could be fit to EEG data following onset of the scene prime, but preceding onset of the object itself. If an appropriate performance baseline were specified, or careful counterbalancing was used to ensure that there were no residual statistical associations between, e.g., physical properties of the scene context and the following object's identity, then above-chance classifier accuracy would be strong evidence for a predictive pre-activation account. A similar approach has been used by Kok, Failing, and de Lange (2014) with fMRI to make the inference that (100% valid) predictive auditory cuing can induce the representation of an expected visual stimulus (a gabor patch in a particular orientation) to be brought online in V1. The data presented in the current thesis could be used as a pilot exploratory dataset for a first-pass machine-learning based search for evidence of predictive pre-activation of visual object features during the scene preview phase. If positive evidence was found, an additional EEG dataset would then be collected that would confirm the result while maximizing power for the design.

Utilizing Different Measures of High and Low Level Visual Similarity As a Component-Neutral Approach to Identifying the Level(s) of Visual Processing Facilitated by Context

In the current thesis, we only presented visual similarity analyses using a single set of visual features derived from the V1-like representation of Pinto, Cox, & DiCarlo (2008) as applied to our stimulus set. In another set of analyses that we are actively working on, we have explored alternative measures of visual similarity derived from a convolutional neural network (Alexnet; Krizhevsky, Sutskever, & Hinton, 2012), that might tap into more or less orthogonal measures of the extent to which high and low-level visual representations are brought online by a given context. More specifically, the first convolutional layer of Alexnet also generates a V1-like feature space but may have somewhat distinct properties from the Pinto et al. (2008) features used in the current thesis; the Pinto et al. features in fact correlate better with slightly later layers within Alexnet. The fully connected layers of Alexnet, then, can be used to derive a higher-level visual representation that captures more categorical information about the objects. For example, using multidimensional scaling, it can be seen that for the higher-level representational similarity space, objects are clustered with other objects of the same category, and germs and machines are

cleanly separated. In the lower-level feature representational similarity space, which is more sensitive to local visual features, the objects are quite mixed up across lower and higher level category boundaries in terms of which specific exemplar is considered close in space to another exemplar. In future work, not only could these Alexnet derived measures be further explored, but additional measures, like the shape-based matching algorithm used in Kovalenko, Chaumon, & Busch (2012), could be examined and compared in a more extensive methods paper. By identifying an optimal set of visual similarity metrics and addressing any multicollinearity issues that might arise, we could potentially much more strongly link EEG and other neurophysiological data on anticipatory processing / template matching to specific stages of visual processing with fewer assumptions based on timing or spatial localization.

The Role of Temporal Expectancy in Optimal Use of Context to Facilitate Visual Object Processing

In our discussion of Chapter 3 Experiments 3.1 and 3.2, we noted some discrepancies across the between-subjects and within-subjects analyses, and suggested they may be explainable in terms of our shift from delayed responding to immediate responding. Another possibility is that temporal expectancy is higher for Chapter 2 Experiment 2.1 and Chapter 3 Experiment 3.1, because in these two experiments the object always had a consistent temporal relationship with the scene (either fixed at a 200 ms or 2500 ms SOA). In Chapter 3 Experiment 3.2, when we continuously varied preview time, we also broke down this temporal structure and presumably made it more difficult to rhythmically synchronize processing of the scene context with that of the following object. Temporal expectancy could also play a role in explaining differences between the two fixed timing experiments, in that temporal expectations are likely to be more precise at shorter delays (200 ms) than substantially longer ones (2500 ms).

One goal of future work branching off from the current project will be to more systematically assess and quantify how the temporal relationship between a scene prime and object target modulates pre-activation of target visual features. Manipulating not only the overall SOA, but also the degree to which SOA is consistent across trials and predictable across the course of the experiment, may also impact the extent to which participants use prime information to facilitate processing of the upcoming target.

Temporal expectations refer to predictions about the perceptual timing of an upcoming stimulus formed implicitly in response to the temporal structure of a task. They are behaviorally detectible when used to enhance performance on a motor or perceptual task. Response latencies across a variety of perceptual, memory, and motoric tasks decrease as stimulus presentation times become more predictable. Saccade latencies to targets also decrease, and smooth pursuit behavior is modulated by increasing temporal regularity of a moving stimulus, including an increase in the occurrence of anticipatory eye gazes (see Nobre, Correa, & Coull, 2007).

Temporal expectation effects are also evident in patterns of brain activity, and are intimately linked to the concept of neural entrainment. One of the earliest ERPs to be characterized, the contingent negative variation (CNV), increases in amplitude during the delay period between a cue and target stimulus, and its amplitude is correlated with time perception and timing performance (reviewed in Martin, Houck, Kičić, & Tesche, 2008). More recently, intracranial recordings in monkey cortex have revealed that neuronal synchrony and spike rates in task-relevant areas are modulated by the so-called "hazard function": the conditional probability of a task-relevant stimulus appearing, given that it has not yet appeared (e.g., Riehle, Grün, Diesmann, & Aertsen, 1997; Janssen & Shadlen, 2005; reviewed in Nobre et al., 2007). Also, the entrainment of neural oscillations to quasi-periodic signals, such as the syllable onsets of continuous speech, are a plausible mechanism of temporal expectancy effects in domains such as speech recognition (see discussion in Arnal & Giraud, 2012; Peelle & Davis, 2012). More generally, converging behavioral and neural evidence suggest that knowing *when* a stimulus will appear in time may help people to anticipate *what* will appear, by more effectively coordinating a neural response.

We are currently working on an additional EEG experiment, in which participants memorize novel-object scene pairs in the paradigm used in the current thesis, but trials are split into blocks with a 200 ms scene-object SOA at test, blocks with a 2500 ms SOA, and blocks with a random 50/50 mixture of the two SOAs. This future work will give us a preliminary avenue for investigating the role that temporal expectation may play in driving the results of the experiments presented in the current thesis.

Hemispheric Differences in Predictive Processing

Another future direction will speak to hemispheric differences in how scene contexts are used to facilitate visual object processing. One of the clearest instances of hemispheric specialization is the strong left lateralization of language production in most people. Early evidence for this came from patients with brain damage in a single hemisphere, and later, the Wada test, in which a barbiturate is injected into the left or right carotid artery to selectively anesthetize the ipsilateral hemisphere. For most people, globally inhibiting the left hemisphere, as in the Wada test, hinders speech production, while right hemisphere inhibition leaves speech production intact. Subsequent research on stroke victims and split-brain patients, as well as studies using visual half-field presentation in neurally healthy individuals, have suggested that while the left hemisphere may drive speech production, both hemispheres contribute to language comprehension, albeit in different ways. For example, while left hemisphere damage is more strongly associated with a variety of comprehension deficits at the level of word meaning and syntax, right hemisphere damage has been linked to impairment in drawing elaborative inferences from verbal materials, as well as joke comprehension (e.g., Beeman, 1993; Brownell, Michel, Powelson, & Gardner, 1983). Moreover, the left and right hemispheres show different patterns of semantic and orthographic activation during word recognition and sentence reading tasks. As a general trend, the left hemisphere shows more signs of top-down modulation of semantic activation, on the basis of syntactic and other higher-level linguistic considerations, than the right (reviewed in Federmeier, 2007).

One method for exploring hemispheric processing differences that will be used in the current proposal is visual half-field presentation. By briefly presenting visual stimuli to the left or right of central fixation (> .5°), it is possible to bias processing to the contralateral hemisphere of the brain. This is driven by the neuroanatomical structure of the visual system – almost all visual input in a given hemifield is sent selectively to primary visual cortex in the contralateral hemisphere. While in principle inter-hemispheric communication may begin shortly thereafter via the corpus callosum, the number of connections between the two hemispheres is orders of magnitude smaller than the number of connections within each hemisphere, suggesting substantial information loss. Moreover, longer-latency topographic differences in the ERP brain response to left vs. right lateralized stimuli attest to the fact that even after hundreds of milliseconds, different neural generators are brought online in response to right vs. left hemifield stimulus presentation (e.g., Huang, Lee, & Federmeier, 2010). Though this speaks less strongly

to the efficacy of the technique with regard to lateralization per se, well-documented behavioral effects of visual hemifield further attest to the lasting effects of lateralized presentation on cognition (reviewed in Federmeier, Wlotko, & Meyer, 2008).

Visual half-field presentation has been used in combination with behavioral methods and ERPs to reveal differences in predictive processing across the two hemispheres in the domain of language comprehension. For example, Deacon and colleagues (2004) found N400 semantic priming for purely associatively related word pairs (e.g., 'honey'/'bee') was selective to rvf/LH (right visual field / Left Hemisphere) presentation, but N400 priming for purely categorically related word pairs (e.g., 'broccoli'/'tree') was selective to lvf/RH (left visual field / Right Hemisphere) presentation. This is consistent with the idea that the left hemisphere is anticipating upcoming linguistic input on the basis of current input, while the right hemisphere is reactively integrating semantic content across words in a more bottom-up fashion (see Federmeier, Wlotko, & Meyer, 2008 for discussion). Federmeier and Kutas (1999b) further demonstrated that for young adults, words that are incongruent with a particular sentential context but nonetheless semantically related to a congruent and predicted word show signs of partial activation in the left, but not right, hemisphere. Specifically, N400 amplitude is reduced for semantically related but incongruent words relative to semantically unrelated and incongruent words when the critical words are presented in the rvf/LH, but not the lvf/RH. Thus, only the left hemisphere demonstrated sensitivity to expected but never presented verbal input in a top-down fashion.

Noting that language production and many controlled aspects of predictive processing in comprehension seem to both be left-lateralized, Federmeier (2007) proposed the "Production Affects Reception in Left Only" (PARLO) framework for understanding hemispheric differences in language comprehension. This framework posits that anticipatory processing is more constrained by top-down linguistic knowledge in the left hemisphere precisely because the left hemisphere encodes information relevant to language production that is less accessible to the right hemisphere.

However, domain general accounts of hemispheric differences may also partially explain the more top-down, predictive nature of processing in the left hemisphere. Hemispheric asymmetries found in the temporal expectation literature, for example, also endorse a predictive bias for the left hemisphere and a reactive bias for the right hemisphere (see Coull & Nobre,

2008). Also, linguistic contexts have been shown to facilitate picture processing on the N400 with similar hemispheric asymmetries to words (Federmeier & Kutas, 2002).

In another experiment we are currently working on, we lateralize presentation of target stimuli to establish whether the left and right hemispheres show similar predictive vs. reactive processing biases in visual object recognition, when non-verbal contextual materials are used as primes. We particularly pay attention to how 'possible' mismatches between the target object and scene prime are processed by each hemisphere relative to the other match/mismatch conditions. Partial facilitation for 'possible' mismatches would be analogous to the 'incongruent-related' condition used in Federmeier and Kutas (1999a,b; 2002), and would similarly be suggestive of a top-down, predictive mechanism. If predictive processing, operationalized in this way, is shown to be lateralized to the left hemisphere in a similar manner to language comprehension, this might suggest a domain general mechanism or a common set of predictive processing principles with respect to hemispheric asymmetries. In addition, we will extend the visual similarity analysis approach applied in Chapters 2 and 3 of the current thesis, to examine component-neutral signatures of processing differences across the hemispheres.

REFERENCES

Anand, K. S., & Dhikav, V. (2012). Hippocampus in health and disease: An overview. *Annals of Indian Academy of Neurology*, *15*(4), 239.

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398.

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review*, *14*(2), 332–337.

Baars, B. J., & Gage, N. M. (2010). *Cognition, brain, and consciousness: Introduction to cognitive neuroscience*. Academic Press.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

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

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

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Rosen, B. R. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454.

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

Barenholtz, E. (2013). Quantifying the role of context in visual object recognition, *22*(1), 30–56.

Barenholtz, E. (2014). Quantifying the role of context in visual object recognition. *Visual Cognition*, *22*(1), 30–56.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, *14*(2), 201–212.

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain and Language*, *44*(1), 80–120.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177.

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.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556.

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, 0582–17.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216.

Brownell, H. H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients. *Brain and Language*, *18*(1), 20–27.

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage*, *49*(4), 3198–3209.

Chauncey, K., Holcomb, P. J., & Grainger, J. (2009). Primed picture naming within and across languages: an ERP investigation. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(3), 286–303.

Chen, J., Cook, P. A., & Wagner, A. D. (2015). Prediction strength modulates responses in human area CA1 to sequence violations. *Journal of Neurophysiology*, *114*(2), 1227–1238.

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and CA1 mismatch detection. *Learning & Memory*, *18*(8), 523–528.

Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience*, 1–26.

Cornelissen, T. H., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. *Attention, Perception, & Psychophysics*, *79*(1), 154–168.

Coull, J. T., & Nobre, A. C. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Current Opinion in Neurobiology*, *18*(2), 137–144.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393–401.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.

Deacon, D., Grose-Fifer, J., Yang, C.-M., Stanick, V., Hewitt, S., & Dynowska, A. (2004). Evidence for a new conceptualization of semantic representation in the left and right cerebral hemispheres. *Cortex*, *40*(3), 467–478.

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.

Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia*, *50*(7), 1271–1285.

Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., & Javitt, D. C. (2000). Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *Journal of Cognitive Neuroscience*, *12*(4), 615–621.

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, *22*(3), 389–398.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655.

Fabiani, M. (2012). It was the best of times, it was the worst of times: A psychophysiologist's view of cognitive aging. *Psychophysiology*, *49*(3), 283–304.

Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1994). Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(2), 148–160.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Federmeier, K. D., & Kutas, M. (1999a). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Federmeier, K. D., & Kutas, M. (1999b). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*(3), 373–392.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 202.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*(7), 730–747.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation*, *51*, 1–44.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146.

Federmeier, K. D., Wlotko, E. W., & Meyer, A. M. (2008). What's 'Right'in Language Comprehension: Event-Related Potentials Reveal Right Hemisphere Language Capabilities. *Language and Linguistics Compass*, *2*(1), 1–17.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP 'old/new'effects: memory strength and decisional factor (s). *Neuropsychologia*, *40*(13), 2288–2304.

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for" top-down" effects. *Behavioral and Brain Sciences*, *39*.

Freeman, J. B., Ma, Y., Barth, M., Young, S. G., Han, S., & Ambady, N. (2015). The neural basis of contextual influences on face categorization. *Cerebral Cortex*, *25*(2), 415–422.

Freeman, J. B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology*, *49*(2), 206–210.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 815–836.

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

Ganis, G., Schendan, H. E., & Kosslyn, S. M. (2007). Neuroimaging evidence for object model verification theory: role of prefrontal control in visual object categorization. *Neuroimage*, *34*(1), 384–398.

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, *3*(1), 128–156.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, *6*(4), 316–328.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, *20*(3), 371–388.

Gronau, N., & Shachar, M. (2014). Contextual integration of visual objects necessitates attention. *Attention, Perception, & Psychophysics*, *76*(3), 695–714.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*(8), 1339–1350.

Hannula, D. E., Federmeier, K. D., & Cohen, N. J. (2006). Event-related potential signatures of relational memory. *Journal of Cognitive Neuroscience*, *18*(11), 1863–1876.

Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *Journal of Cognitive Neuroscience*, *19*(10), 1690–1705.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain: A Journal of Neurology*, *121*(6), 1133–1142.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*(1), 243–271.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*(1), 62.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*(2), 259–276.

Huang, H.-W., Lee, C.-L., & Federmeier, K. D. (2010). Imagine that! ERPs provide evidence for distinct hemispheric contributions to the processing of concrete and abstract concepts. *NeuroImage*, *49*(1), 1116–1123.

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature Neuroscience*, *8*(2), 234–241.

Jiang, Y., & Chun, M. M. (2003). Contextual cueing. *Attention and Implicit Learning*, *48*, 277.

Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 11–11.

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Kleiman, G. M. (1980). Sentence frame contexts and lexical decisions: Sentence-acceptability and word-relatedness effects. *Memory & Cognition*, *8*(4), 336–344.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554.

Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*, 201705652.

Kok, P., & Turk-Browne, N. B. (2018). Associative prediction of visual shape in the hippocampus. *Journal of Neuroscience*, 0163–18.

Kovalenko, L. Y., Chaumon, M., & Busch, N. A. (2012). A pool of pairs of related objects (POPORO) for investigating visual semantic integration: behavioral and electrophysiological validation. *Brain Topography*, *25*(3), 272–284.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Kutas, M., & Federmeier, K. D. (2011a). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621.

Kutas, M., & Federmeier, K. D. (2011b). 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. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 471.

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.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565.

Martin, T., Houck, J. M., Kičić, D., & Tesche, C. D. (2008). Interval timers and coupled oscillators both mediate the effect of temporally structured cueing. *Neuroimage*, *40*(4), 1798–1806.

McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(3), 634.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14*(6), 648–652.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65.

Mostert, P., Kok, P., & De Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, *5*, 18253.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, *48*(2), 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.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251.

Munneke, J., Brentari, V., & Peelen, M. (2013). The influence of scene context on object recognition is independent of attentional focus. *Frontiers in Psychology*, *4*, 552.

Nieuwland, M., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Ito, A. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. *BioRxiv*, 111807.

Nobre, A. C., Correa, A., & Coull, J. T. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527.

Palmer, tephen E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526.

Pearson, J., & Westbrook, F. (2015). Phantom perception: voluntary and involuntary nonretinal vision. *Trends in Cognitive Sciences*, *19*(5), 278–284.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*.

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology*, *4*(1), e27.

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713.

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.

Riehle, A., Grün, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, *278*(5345), 1950–1953.

Robinson, M. M., Clevenger, J., & Irwin, D. E. (2018). The action is in the task set, not in the action. *Cognitive Psychology*, *100*, 17–42.

Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience*, *32*(5), 576–589.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Schendan, H. E., & Ganis, G. (2012). Electrophysiological potentials reveal cortical mechanisms for mental imagery, mental simulation, and grounded (embodied) cognition. *Frontiers in Psychology*, *3*.

Schendan, H. E., & Ganis, G. (2015). Top-down modulation of visual processing and knowledge after 250 ms supports object constancy of category decisions. *Frontiers in Psychology*, *6*, 1289.

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931–945.

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience*, *15*(1), 111–135.

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, *19*(5), 734–749.

Schendan, H. E., & Lucia, L. C. (2009). Visual object cognition precedes but also temporally overlaps mental rotation. *Brain Research*, *1294*, 91–105.

Schendan, H. E., & Lucia, L. C. (2010). Object-sensitive activity reflects earlier perceptual and later cognitive processing of visual objects between 95 and 500 ms. *Brain Research*, *1329*, 124–141.

Schendan, H. E., & Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage*, *44*(4), 1423–1438.

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.

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, *314*(5803), 1311–1314.

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, *59*(2), 336–347.

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.

Tootell, R. B., Tsao, D., & Vanduffel, W. (2003). Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. *Journal of Neuroscience*, *23*(10), 3981–3989.

Torralbo, A., Walther, D. B., Chai, B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2013). Good exemplars of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. *PloS One*, *8*(3), e58594.

Truman, A., & Mudrik, L. (2018). Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia*.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*(33), 11177–11187.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*(3), 24–24.

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

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, *39*(3), 483–508.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356–366.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32.

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84.

Wu, Y. C., & Coulson, S. (2011). Are depictive gestures like pictures? Commonalities and differences in semantic processing. *Brain and Language*, *119*(3), 184–195.

Yan, S., Kuperberg, G. R., & Jaeger, T. F. (2017). Prediction (Or Not) During Language Processing. A Commentary On Nieuwland et al.(2017) And Delong et al.(2005). *BioRxiv*, 143750.

APPENDIX A: MATCH CONDITION CLUSTER ANALYSIS RESULTS

The following condition contrasts were assessed using a by-subjects dependent samples T-test on the subject-level averaged waveforms (down-sampled to 100 Hz) from 10 to 990 ms in 10 ms (1 sample) increments:

Match vs. Distortion
Match vs. Within Category Mismatch
Match vs. Between Category Mismatch
Distortion vs. Within Category Mismatch
Within Category Mismatch vs. Between Category Mismatch
Between Category Mismatch – Swapped vs. Between Category Mismatch – New

Positive and negative clusters were separately assessed. Individual channel-time-points were considered for cluster inclusion at $\alpha = .05$, and were required to have at least two neighboring channels also included in the cluster. Cluster significance was computed by comparing the sum of the T values within each cluster to the distribution of the maximum sum of T values cluster score over a random permutation baseline, $\alpha = .025$, $n = 2000$ repetitions. See Figures A1 and A2 for the distributions of significant clusters over time and space for each condition contrast.

Distortion - Exact Match
No significant clusters were found.

Within Category Mismatch - Exact Match
A negative cluster was found from 250-400ms, which began at frontal sites, was broadly distributed across the head from ~290-350 ms, and ended at central sites (sumT = -818, p = .0095).

Between Category Mismatch - Exact Match
A negative cluster was found from 230-430 ms, which began at frontal sites, was broadly distributed at ~280-370 ms, and ended at centro-parietal sites (sumT = -1418, p =.0030).

Within Category Mismatch - Distortion

Two negative clusters were found. The first negative cluster was from 170-430 ms, again starting at frontal sites, being broadly distributed at ~290-390 ms, and ending at central sites (Tsum = -1575, p = .0025). The second negative cluster was from 760-990 ms at central and posterior sites (Tsum = -695, p = .0160).

Between Category Mismatch - Within Category Mismatch

A positive cluster was found from 760-990 ms at posterior sites, possibly reflecting response differences among correct trials for the between vs. within category mismatch conditions (Tsum = 738, p = .0070).

Between Category Mismatch Swap vs. New Trials

No significant clusters were found.

Figure A1. Timing and scalp topography of early match condition clusters, in 50 ms increments. Significant cluster sites indicated with an asterisk.

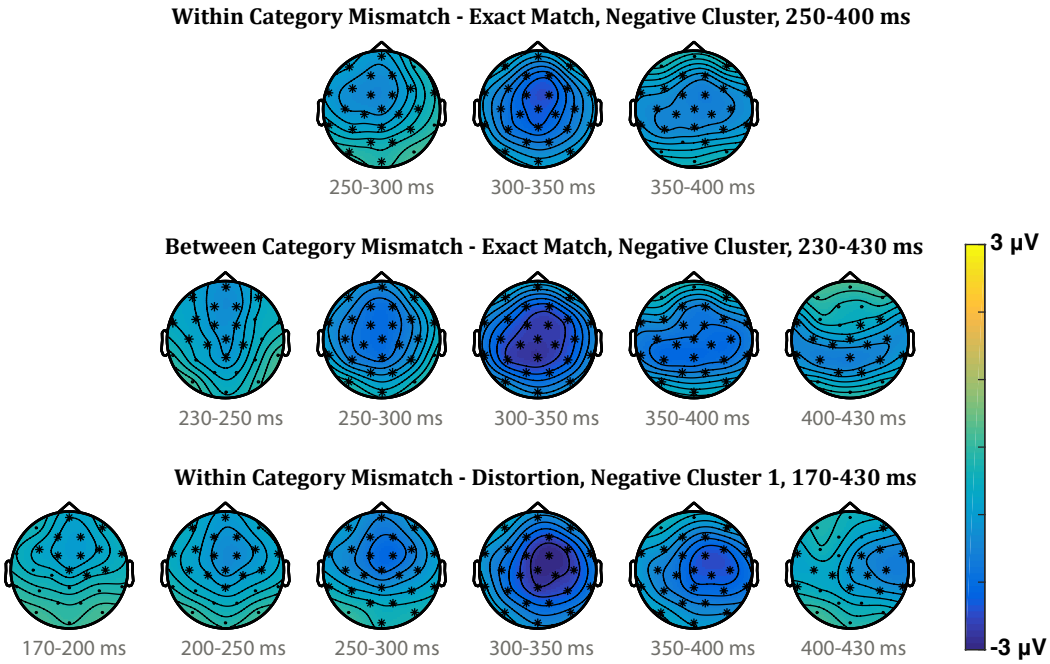
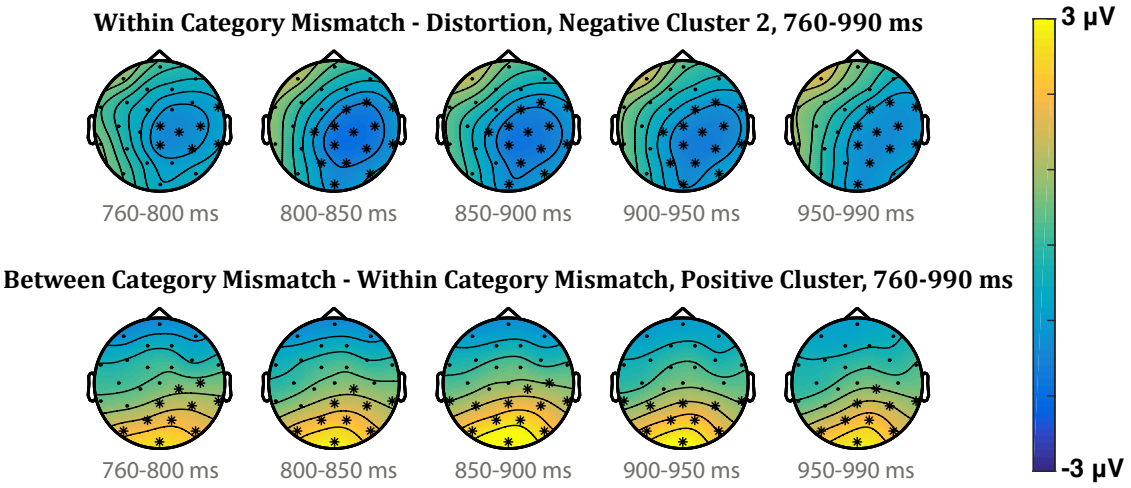
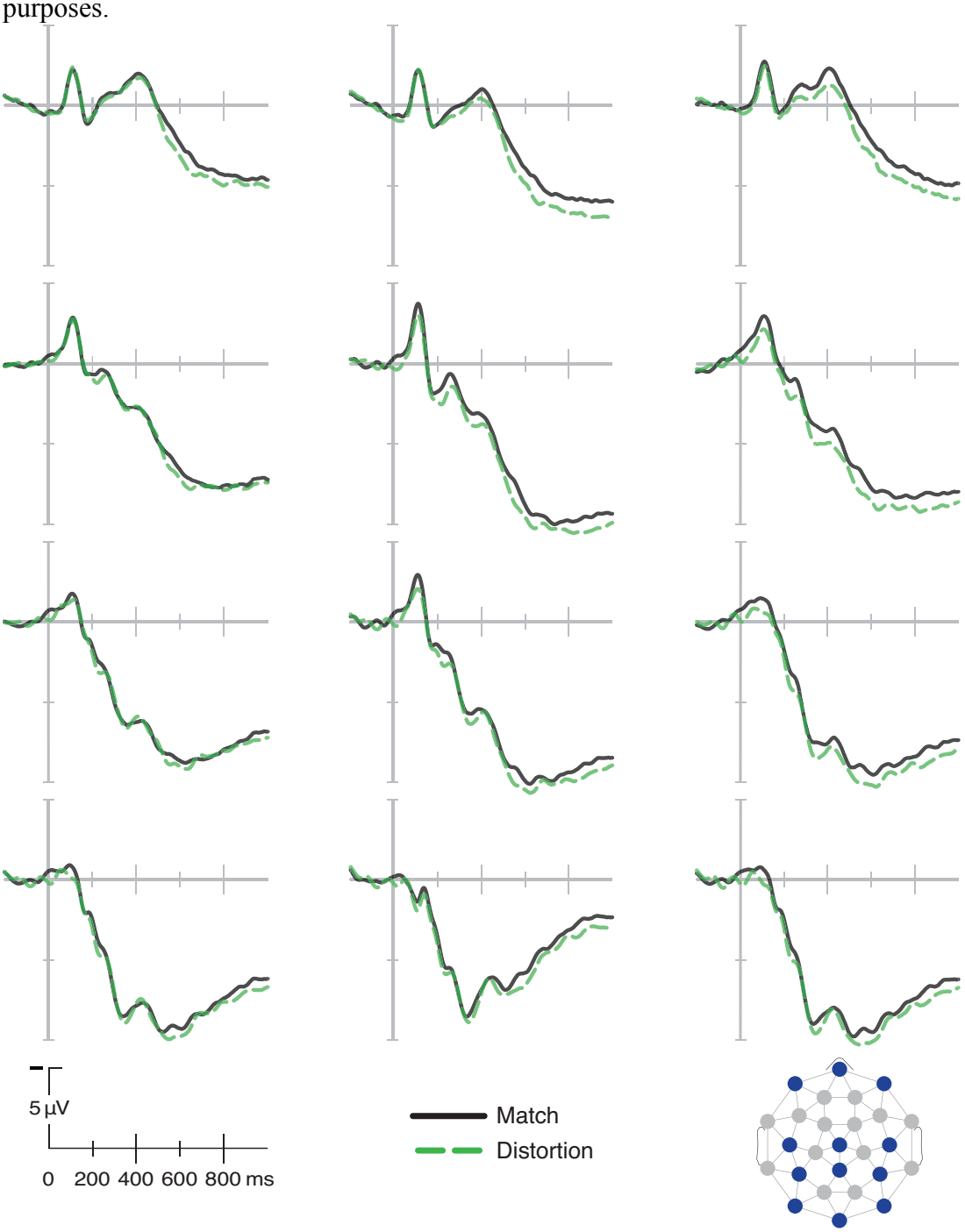


Figure A2. Timing and scalp topography of late match condition clusters, in 50 ms increments. Significant cluster sites indicated with an asterisk.



APPENDIX B: COMPARISON OF ERP RESPONSE TO MATCH VS. DISTORTION CONDITIONS

Figure B1. Match vs. distortion condition at 12 representative sites (scalp locations indicated at bottom right). An additional 15 Hz low pass filter was applied after averaging for display purposes.



APPENDIX C: VISUAL DISTANCE TO TARGET VS. PRESENTED OBJECT PROTOTYPICALITY

We tested that visual distance to target effects were not driven by prototypicality of the presented object (relative to its own category), which would be independent of the presented scene context. Visual distance to the 36 prototype images (18 germs, 18 machines) from which the exemplar object images were derived was computed for each object image presented in the test phase. The same feature space and procedure was used to compute visual distance as when computing visual distance between the presented and target object images. The same random effects structure was maintained as in the match condition and visual distance to target analyses. Effects of visual distance to prototype, and the additive benefit of including visual distance to target as an additional predictor, were assessed using nested model comparisons. To test for effects of prototypicality, models containing fixed effects of match condition, response, and their interaction, as well as (grand mean centered) visual distance to prototype, were compared to a null model excluding the effect of prototype. All beta values and standard errors (in parentheses) are reported as 1000 times the original estimates. Effects of distance to prototype were significant or numerically trended in the same direction across all four components (N300: β = -2.606 (1.466), χ_1^2 = 3.15, p < .1; N400: β = -5.094 (1.718), χ_1^2 = 8.77, p < .01; early LPC: β = -5.122 (1.609), χ_1^2 = 10.10, p < .01; late LPC: β = -3.074 (1.670), χ_1^2 = 3.38, p < .1). Next, (grand mean centered) visual distance to target was included as an additional predictor to distance to prototype, to see if it explained substantially more variance than a null model containing only distance to prototype. Including visual distance to target improved model fit for all four components (N300: χ_1^2 = 64.39, p < .001; N400: χ_1^2 = 82.12, p < .001; early LPC: χ_1^2 = 27.00, p < .001; late LPC: χ_1^2 = 53.77, p < .001). The converse was less true: adding distance to prototype as an additional predictor to a model that already included distance to target improved model fit more modestly or not at all (N300: χ_1^2 < 1; N400: χ_1^2 = 2.61, p = .106; early LPC: χ_1^2 = 5.63, p < .05; late LPC: χ_1^2 < 1). Within models that included both distance to prototype and distance to target as fixed effects, effect size estimates tended to be larger and standard errors tended to be smaller for distance to target effects (N300: distance to target β = -3.573 (.445), distance to prototype β = -1.111 (1.482); N400: distance to target β = -5.470 (.603), distance to prototype β = -2.810 (1.739); early LPC: distance to target β = -3.008 (.579), distance to prototype β = -3.867 (1.627); late LPC: distance to target β = -4.399 (.599), distance to prototype β = -1.237 (1.690)).

Taken together, visual distance to the specific target object image associated with each scene appears to be a more important explanatory variable than distance to the category prototype of the presented object.

APPENDIX D: VISUAL DISTANCE TO TARGET EXEMPLAR VS. TARGET PROTOTYPE

We assessed whether participants compared the current object image in the test phase to the specific target exemplar object image paired with the scene, or to the prototype target object image. The same random effects structure was maintained as in the match condition and visual distance to target analyses. Effects of visual distance to target prototype, and the additive benefit of including visual distance to target exemplar as an additional predictor, were assessed using nested model comparisons. Models containing fixed effects of match condition, response, and their interaction, as well as visual distance to target prototype and visual distance to target exemplar, were compared to a null model excluding the effect of visual distance to target exemplar. Continuous predictors were grand mean centered. Including visual distance to target exemplar improved model fit for all four component time-windows (N300: χ_1^2 = 57.05, p < .001; N400: χ_1^2 = 62.92, p < .001; early LPC: χ_1^2 = 17.55, p < .001; late LPC: χ_1^2 = 34.07, p < .001).

APPENDIX E: EXPERIMENT 3.1 MATCH CONDITION PERMUTATION-BASED CLUSTER ANALYSIS

Experiment 3.1 Match Condition Permutation-Based Cluster Analysis

The following condition contrasts were assessed using a by-subjects dependent samples T-test on the subject-level averaged waveforms (down-sampled to 100 Hz) from 10 to 990 ms in 10 ms (1 sample) increments:

Match vs. Distortion
Match vs. Within Category Mismatch
Match vs. Between Category Mismatch
Distortion vs. Within Category Mismatch
Within Category Mismatch vs. Between Category Mismatch
Between Category Mismatch – Swapped vs. Between Category Mismatch – New

Positive and negative clusters were separately assessed. Individual channel-time-points were considered for cluster inclusion at $\alpha = .05$, and were required to have at least two neighboring channels also included in the cluster. Cluster significance was computed by comparing the sum of the T values within each cluster to the distribution of the maximum sum of T values cluster score over a random permutation baseline, $\alpha = .025$, $n = 2000$ repetitions. See Figure A1 for the distributions of significant clusters over time and space for each condition contrast.

Distortion - Exact Match
No significant clusters were found.

Within Category Mismatch - Exact Match
A broadly distributed negative cluster was found from 340-990 ms, beginning at fronto-central sites and later extending to posterior sites (sumT = -3597, p = .0005).

Between Category Mismatch - Exact Match
A broadly distributed negative cluster was found from 300-710 ms (sumT = -2539, p = .0005).

Within Category Mismatch - Distortion

A broadly distributed negative cluster was found from 510-930 ms (sumT = -1980, p = .0010).

Between Category Mismatch - Within Category Mismatch

A positive cluster was found from 660-990 ms, focused on posterior sites (sumT = 1226, p = .0030).

Between Category Mismatch Swap vs. New Trials

No significant clusters were found.

Figure E1. Timing and scalp topography of experiment 3.1 match condition clusters. Significant cluster sites indicated with an asterisk.