The N300: An Index For Predictive Coding Of Complex Visual Objects and Scenes

[1]Manoj Kumar, [2]Kara D. Federmeier, [2]Diane M. Beck

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA.

[2]Department of Psychology, Program in Neuroscience, and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Corresponding Author: Manoj Kumar

Email: mk35@princeton.edu

Kumar ORCID 0000-0002-4911-2885
Federmeier ORCID 0000-0002-7815-1808
Beck ORCID 0000-0001-9802-5828

Classification

BIOLOGICAL SCIENCES

Psychological and Cognitive Sciences

Keywords

N300, predictive coding, visual perception, statistical regularities

Author Contributions

All authors worked together to design the study and analysis plan and to interpret the results and write the paper. Data collection and processing were done by M.K. with guidance from K.D.F.

This PDF file includes:

Main Text
Figures 1 to 3
Tables 1 to 2

Abstract

The bulk of support for predictive coding models has come from the models' ability to simulate known perceptual or neuronal phenomena, but there have been fewer attempts to identify a reliable neural signature of predictive coding. Here we propose that the N300 component of the event-related potential (ERP), occurring 250-350 ms post-stimulus-onset, may be such a signature of perceptual hypothesis testing operating at the scale of whole objects and scenes. We show that N300 amplitudes are smaller to representative ("good exemplars") compared to less representative ("bad exemplars") items from natural scene categories. Integrating these results with patterns observed for objects, we establish that, across a variety of visual stimuli, the N300 is responsive to statistical regularity, or the degree to which the input is "expected" (either explicitly or implicitly) by the system based on prior knowledge, with statistically regular images, which entail reduced prediction error, evoking a reduced response. Moreover, we show that the measure exhibits context-dependency; that is, we find the N300 sensitivity to category representativeness only when stimuli are congruent with and not when they are incongruent with a category pre-cue, suggesting that the component may reflect the ease with which an image matches the current hypothesis generated by the visual system. Thus, we argue that the N300 ERP component is the best candidate to date for an index of perceptual hypotheses testing, whereby incoming sensory information for complex visual objects and scenes is accessed against contextual predictions generated in mid-level visual areas.

Significance Statement

Predictive coding models postulate that our perception of visual sensory input is guided by prior knowledge and the situational context, such that it is facilitated when the input matches expectation and hence produces less prediction error. Here, we show that an electrophysiological measure, the N300, matches the features hypothesized for a measure of predictive coding: complex scenes (like objects) elicit less N300 activity when they are statistically regular (e.g., more representative of their categories), in a manner that itself is context dependent. We thus show that the N300 provides a window into the interaction of context, prediction, and visual perception.

Introduction

The stars in the night sky are not arranged in the shape of a great bear and there is no rabbit on the moon; it is our prior knowledge of these shapes that invokes such descriptions. Increasingly, it is clear that more generally our perception does not depend on the sensory stimulus alone but is also dynamically influenced by our prior knowledge (Smith & Loschky, 2019; Caddigan et al., 2017; Lupyan, 2017; Vo & Wolfe, 2013; Voss et al., 2012; Summerfield et al., 2006). Indeed, many models of perception include some form of perceptual hypothesis testing (PHT), in which perception, a hard inverse problem, is conceived of as a process of generating a hypothesis on the basis of both sensory input and prior knowledge and the current context (Clark, 2013; Gregory, 1980; Hochberg, 1981; Huang & Rao, 2011; Rock, 1983; Helmholtz, 1925). Recently, one class of PHT models has garnered increased interest: hierarchical predictive coding models (Friston, 2005; Rao & Ballard, 1999; Spratling, 2010), which posit that each area of, for example, visual cortex learns statistical regularities from the world that it then uses, jointly with the input from the preceding area, to make predictions about the stimulus. In particular, the prediction and incoming sensory signal are proposed to undergo an iterative matching process at each stage of the processing hierarchy, with the prediction feeding back on the preceding area. The mismatch ("prediction error"), if any, between the prediction and the incoming sensory signal is then propagated to higher layers in the processing hierarchy, revising the weights of the hypotheses, until the feedback matches the incoming signal and the error is zero (Friston, 2005; Lange et al., 2018; Rao & Ballard, 1999). These predictive coding models have risen to prominence in recent years, in part because they represent an efficient coding scheme for the complexity of the visual world and, perhaps more importantly, because they posit

a role for the abundant feedback connections known to exist between visual areas. The bulk of support for predictive coding models has come from the models' ability to simulate known perceptual or neuronal phenomena (reviewed (Spratling, 2016)), but there have been fewer attempts to identify a reliable neural signature of predictive coding. Here we propose that the N300 component of the event-related potential (ERP) may be just such a signature of predictive coding mechanisms in occipitotemporal visual cortex, operating at the scale of whole objects and scenes.

What would characterize a good measure of predictive coding in visual processing? First, it should be observed to a variety of types of visual stimuli (objects, faces, natural scenes) across most or all viewing conditions. More importantly, it should be responsive to statistical regularity, or the degree to which the input is "expected" (either explicitly or implicitly) by the system based on prior knowledge. In particular, the measured response should increase with increasing irregularity, in keeping with the increased iterations, or inference-based error, proposed to occur when an item does not match the prediction. Finally, the measure should show context-dependency, as statistical regularities need to be sensitive to the immediate context in order to be of use to the system.

Here we show that the N300 ERP component meets all three criteria and thus may serve as a signature of hierarchical inference/predictive coding mechanisms for complex visual stimuli in occipitotemporal cortex. The N300 is a negative going component with a frontal scalp distribution that peaks around 300 ms after the onset of a visual stimulus. It is sensitive to global perceptual properties of visual input (Mcpherson & Holcomb, 1999;

Schendan & Kutas, 2002, 2003) but not to manipulations limited to low level visual features (e.g., color, or small-scale line segments; (Schendan & Kutas, 2007) that are known to be processed in early visual cortex. Thus, the N300 is a “late” visual component, with likely generators in occipitotemporal cortex (Schendan, 2019; Sehatpour et al., 2006). It immediately precedes access to multimodal semantic memory (reflected in the N400, which is observed later in time than the N300 when both are present; (Kutas & Federmeier, 2011)), and is therefore well positioned to capture the iterative, knowledge- and context-sensitive process of visual processing of the global features of the stimuli as proposed by predictive coding models.

Importantly, as hypothesized by predictive coding models, the amplitude of the N300 increases for less “expected” (i.e., less statistically regular) stimuli. The N300 is larger to pictorial stimuli that lack a global structure as compared to when the global structure of the object is clearly discernible (Schendan & Kutas, 2003). The N300 is also sensitive to repetition, with a reduced amplitude for repeated presentations; importantly, however, N300 repetition effects (but not those on earlier components) depend on knowledge, as they are larger when the visual stimulus is meaningful (Schendan & Maher, 2009; Voss et al., 2010; Voss & Paller, 2007). Similarly, and critically, N300 amplitudes are sensitive to a variety of factors that reflect the degree to which an object fits with prior experience. For example, N300 amplitudes are sensitive to the canonical view of an object; an open umbrella oriented horizontally (non-canonical) elicits a larger N300 amplitude than an open umbrella oriented vertically (Schendan & Kutas, 2003; Vo & Wolfe, 2013). Amplitude modulations are also linked to factors such as object category membership, presence of category-diagnostic object features, and (rated) match to object knowledge

(Gratton et al., 2009; Schendan, 2019; Schendan & Maher, 2009). This pattern of data suggests that the N300 may be a good marker for not only the global structure of an object but the degree to which the input matches learned statistical regularities more generally, with larger N300 amplitudes for stimuli that do not match predictions based on learned regularities and hence require further processing. Indeed, Schendan (2019) has come to a similar conclusion, arguing that the N300 reflects “object model selection,” in which the input is matched to possible known objects.

Although the N300 conforms to a number of the criteria outlined above, there are a few open questions that need to be addressed in order to better situate the component as a potential neural correlate of hierarchical inference or predictive coding. Thus far, the N300 has primarily been elicited in response to objects, sometimes in the context of a scene (Lauer et al., 2020; Mudrik et al., 2010; Vo & Wolfe, 2013), but still ostensibly to an object. If the N300 reflects a general signature of hierarchical inference or predictive coding, then it should be elicited by other meaningful visual stimuli, such as natural scenes. In fact, scrambled scenes (created by recombining parts of the scene image into a random jigsaw) elicit larger N300 amplitudes than do intact and identified scenes (Pietrowsky et al., 1996). Because the scrambled scenes were degraded, however, it is not clear whether these effects simply reflect the disruption to global structure or a deviation from statistical regularity more generally. Here we use intact scenes that are either highly representative of their category (e.g., good exemplars of that category) or less representative of their category (bad exemplars). Importantly, all the images are good photographs of real world scenes (i.e., they are not degraded); they are statistically regular or irregular by virtue of how representative they are of their category. A highly

representative exemplar of its category, by definition, contains more information about its category and thus serves as a better initial prediction (i.e., has high statistical regularity). We ask whether such statistically regular and irregular stimuli elicit differential N300s, as would be hypothesized if this component is a measure of hierarchical inference or predictive coding.

We have previously found that good scene exemplars are more readily detected than bad exemplars (Caddigan et al., 2010, 2017) that is, participants are better at discriminating briefly presented and masked intact photographs from fully phase-scrambled versions when those images are good exemplars of their category (i.e., beaches, forests, mountains, city streets, highways, and offices). Importantly, again, there were no artificially introduced objects in any of the bad exemplars nor were they impoverished or degraded in any way. Instead their good and bad status derived entirely from how representative they were of the category being depicted (established in a separate Amazon Mechanical Turk rating task). Although category was relevant to the choice of stimuli and whether they were designated good or bad, it was completely irrelevant to the judgement being made (was the stimuli an intact photo or noise?). Nonetheless, participants had significantly higher sensitivity (d') for good than bad exemplars (Caddigan et al., 2010, 2017), suggesting that good exemplars perceptually cohere sooner than bad exemplars. Relatedly, the categories of those same good exemplars are better decoded, using fMRI multi-voxel pattern analysis, than are the categories of the bad exemplars in a number of visual areas, including V1 and the parahippocampal place area (PPA; (Torralbo et al., 2013)). Interestingly, the BOLD signal for those same bad exemplars is larger than that for good exemplars in the PPA

186 (Torralbo et al., 2013), in keeping with predictions from hierarchical predictive coding
187 (i.e., increased activity for the less statistically regular images). The poorer detection with
188 brief presentations, weaker representations in the brain, and greater activity evoked by
189 bad than good scene exemplars make these stimuli good candidates for eliciting a
190 neural signature of hierarchical predictive coding.

191 In **Experiment 1**, we recorded scalp EEG while participants viewed good and bad scene
192 exemplars and made a good/bad judgment. If the N300 serves as an index of matching
193 incoming stimuli to learned statistical regularities, then N300 amplitude should be
194 smaller for good exemplars of natural scenes than the bad exemplars. In **Experiment 1**
195 participants viewed the stimuli without any forewarning of what to expect (category and
196 good/bad status were fully randomized; see **Figure 1A**). In **Experiment 2**, however, we
197 then manipulated the expectations of the participants at the beginning of each trial by
198 presenting a word cue (e.g., 'Beach') that either matched the upcoming scene's category
199 (on 75% of trials) or mismatched the upcoming image category (e.g., preceding a forest
200 with the 'Beach' cue; see **Figure 1B**). If the N300 reflects the process of matching the
201 input to statistical regularities, it should also be sensitive to the particular template (i.e.,
202 statistical regularity) activated on that trial, as predicted by the cue.

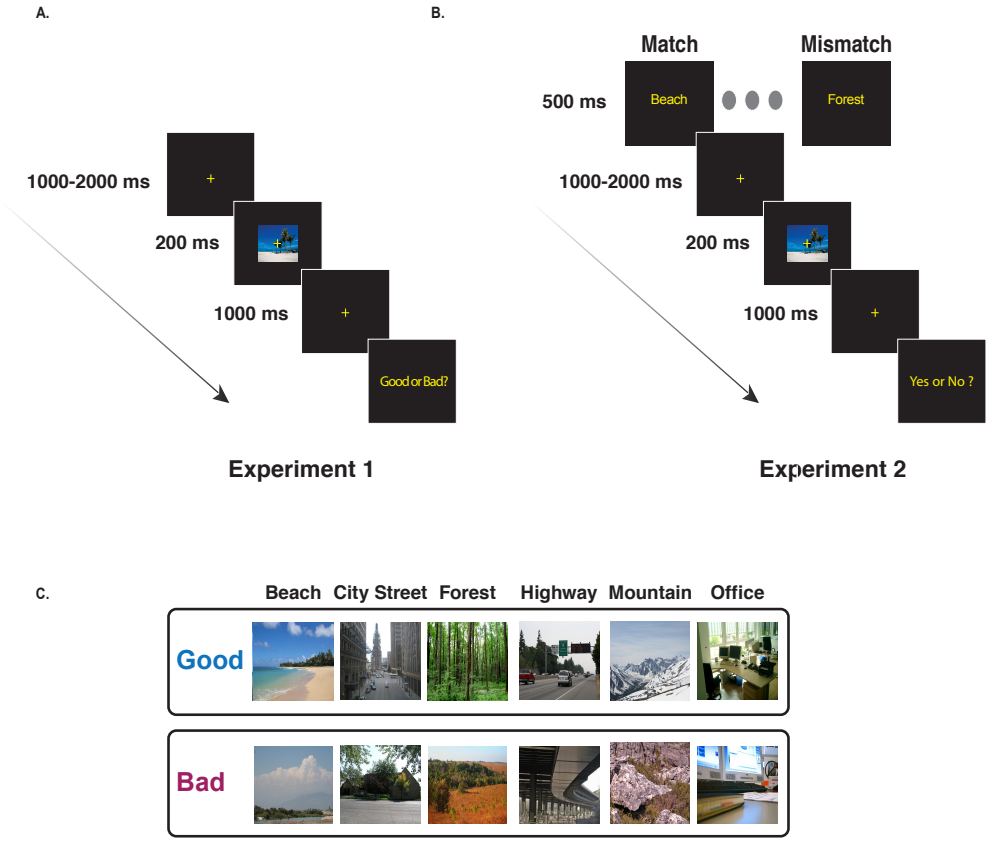

Figure 1. Schematic of one trial in each of the experiments. **A.** In **Experiment 1**, a fixation cross was shown in the center of screen for a randomly chosen interval between 1000-2000 ms. A good or bad exemplar image from one of the six categories was then presented for 200ms, followed by a fixation cross. After a delay of 1000ms, the subjects respond to the question "Good or Bad?" with a button press and the next trial begins. **B.** In **Experiment 2,** the trial sequence is similar to **Experiment 1** with the following differences. At the start of each trial a word cue (e.g., "Beach") from one of six categories (beaches, city streets, forests, highways, mountains, and offices) is shown. At the end of the trial the subjects make a delayed response, with a button press, to the question "Yes or No?" ("Yes" if the image matches the cue and "No" otherwise) and the next trial begins. On 25% of the trials, there is a mismatch between the word cue and the image category. **C.** A sample of good and bad exemplars from each category used in our study.

Results

Experiment 1

Behavior

To motivate participants to attend to the scenes, we asked participants to make a delayed response on each trial, judging whether the exemplar was a good or bad exemplar of the scene category to which it was presumed to belong. Participants labeled most good exemplars as “good” (mean = 86.2%, std. dev = 13.9%) and labeled bad exemplars as “bad” about half the time (mean = 56.2%, std. dev = 15.6%). All trials were used for the ERP analyses, irrespective of the choice of the participants.

ERPs

Grand-averaged ERPs at eight representative sites are plotted in **Figure 2**. Responses to good and bad exemplars can be seen to diverge beginning around 250 ms after stimulus onset, with greater negativity for bad exemplars than for good exemplars. The polarity, timing, and frontal scalp distribution of this initial effect is consistent with prior work describing the N300 (Mcpherson & Holcomb, 1999; Schendan & Kutas, 2002, 2003, 2007); see **Supplementary Information** for a formal distributional analysis.

N300

To characterize the good/bad effect on the N300, mean amplitudes were measured from all 11 frontal electrode sites between 250 and 350 ms (see **Methods** for more details). Bad exemplars elicited significantly larger (more negative) N300 responses (mean = -6.4 μV) than did good exemplars (mean = -5.3 μV); $t(19)=-5.4$ and Bayes factor = 737.7

(**Table 1**). In other words, we see the predicted differential response to statistically irregular exemplars (bad exemplars) as compared to the statistically regular exemplars (good exemplars). The larger amplitude for the bad exemplars, as compared to the good exemplars aligns with PHT predictions that would posit greater inference error, and, hence, greater iterative processing for the bad exemplars as compared to the good exemplars. These results also confirm that the N300 indexes a match to statistical regularities of natural scenes and thus extend the validity of the N300 to not only objects, or objects in scene contexts, but more broadly to complex natural scenes.

Post N300 Components

Although the N300 was the component of primary interest, to more completely characterize the brain's response to the scenes, we also examined good/bad differences in later time windows encompassing the N400 (350-500 ms) and Late Positive Complex (LPC) (500-800 ms). The details of the analyses and results are provided in the **Supplementary Information** and summarized here. N400 responses, which index multimodal semantic processing, were larger for bad (-3.3 μV) than for good exemplars (-2.2 μV), suggesting that items that better fit their category allow facilitated semantic access. We note however, that given the similar scalp distribution of the N300 and the N400 to picture stimuli (Ganis et al., 1996), it is difficult to tell where the boundary of the two components might be and thus how much the N400 pattern might be influenced by the preceding N300. LPC responses were larger -- more positive – to good (4.5 μV) than to bad (3.3 μV) exemplars. The LPC amplitude is known to positively correlate with confidence in decision making (Finnigan et al., 2002). Larger LPC responses to good

items, therefore, is consistent with the behavioral pattern in which good exemplars were classified more consistently than bad exemplars.

Figure 2.

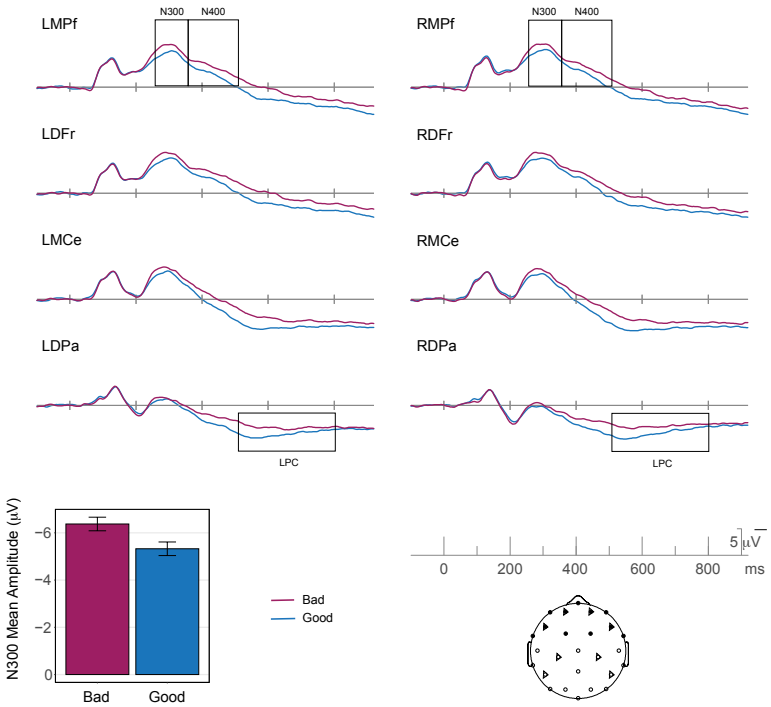

Figure 2. Grand average ERP waveforms for good (blue) and bad (maroon) exemplars in **Experiment 1** are shown at 8 representative electrode sites distributed over the head. Plotted channel locations are marked as triangles on the schematic of the scalp. Negative voltage is plotted upwards. The waveforms differ over frontal sites beginning in the N300 time-window (250-350 ms), with greater negativity for bad exemplars as compared to good exemplars. The bar plot gives mean amplitude over the 11 frontal electrode sites (darkened electrode sites on the schematic of the scalp) used for the

primary statistical analyses. The error bars plotted are within-subject confidence intervals. N=20.

Table 1. Experiment 1, mean amplitudes in the N300 time-window (250-350 ms) over 11 frontal electrode sites (see **Figure 2**), along with t-test and Bayes factor values. The N300 response to bad exemplars is more negative (larger) than that to good exemplars. The t- test and Bayes factor calculations compared the within subject Good/Bad difference to 0.

Condition	N	Mean (μV)	Mean Bad/Good Difference (μV)	Bad/Good Difference 95% C.I.	t(19)	p	Bayes Factor
Bad	20	-6.4±0.61	-1.05	-1.46 to -0.64	-5.4	3.3E-05	747.7
Good	20	-5.3±0.61					

Note: ± values reflect the normed standard deviation within subjects. C.I. = confidence interval.

Experiment 2

The good/bad difference observed in **Experiment 1** was elicited without any expectation regarding the specific category to be presented (i.e., category and good/bad status were completely randomized). Thus, the particular template or statistical regularity with which the image was compared must have been initially elicited by the input itself. In

Experiment 2, we preceded each image with a word cue that either matched or mismatched the upcoming category. If the N300 difference observed in **Experiment 1** reflects the matching of incoming stimuli to learned statistical regularities, we should be able to modulate that difference by activating either the appropriate (match cue) or inappropriate (mismatch cue) statistical regularity. In particular, since neither a good nor a bad exemplar (e.g., of a beach) should be a better match to an inappropriate category (e.g., a forest) we should see an elimination of the N300 difference between good and bad exemplars when the cue mismatches the current category.

Behavior

On each trial, participants were asked to respond if the stimulus matched the verbal cue ("Yes" or "No") via a button press. In the match condition, participants responded "Yes" with a higher degree of accuracy for good exemplars (mean = 98.7%, std. dev = 2.4%) than for bad exemplars (mean= 67.9% and std. dev = 14.6%). In the mismatch condition, wherein the exemplars did not fit the cued category, participants responded "No" with similar accuracy for good (mean = 95.9%, std. dev = 4.6%) and bad exemplars (mean = 94.0% and std. dev = 5.5%). All trials were used for the ERP analyses.

ERPs

Scenes elicited an N300 response (**Figure 3**) with similar timing, polarity and scalp distribution to that observed in **Experiment 1**; see the **Supplementary Information** for a formal distributional analysis. Analyses of N300 mean amplitudes were conducted using the same time window (250-350 ms) and frontal electrode sites as in **Experiment**

1, here comparing good and bad exemplars under the two cueing conditions: match and mismatch.

N300

In the match condition, when the scene is congruent with the verbal cue, we replicate the N300 effect of **Experiment 1** for the good and bad exemplars, with a frontally distributed negativity that is larger for the bad exemplars than the good exemplars (**Figure 3, Table 2**). Importantly, and as predicted, this N300 difference between good and bad exemplars is notably reduced – indeed, likely absent altogether (Bayes factor 0.33) – in the mismatch condition, consistent with the idea that the N300 is indexing the fit of the incoming stimulus to the template activated by the verbal cue; that is, neither a good or bad exemplar of category A represents a better match to a template for category B.

For completeness, and to compare the N300 in our experiment with its characterization in the existing literature, we also performed an ANOVA across multiple factors: Good/Bad x Cueing (Match/Mismatch) x Anteriority x Laterality x Hemisphere (see **Supplementary Information** for results of the distributional analysis). There was a main effect of Good vs. Bad (bad larger than good; ($F(1,19) = 15.34$) and an interaction between Good/Bad and Cueing ($F(1,19) = 5.87$), with larger Good/Bad effects when the scene matched the cue. The main effect of Cueing was not significant ($F(1,19) = 0$).

Figure 3

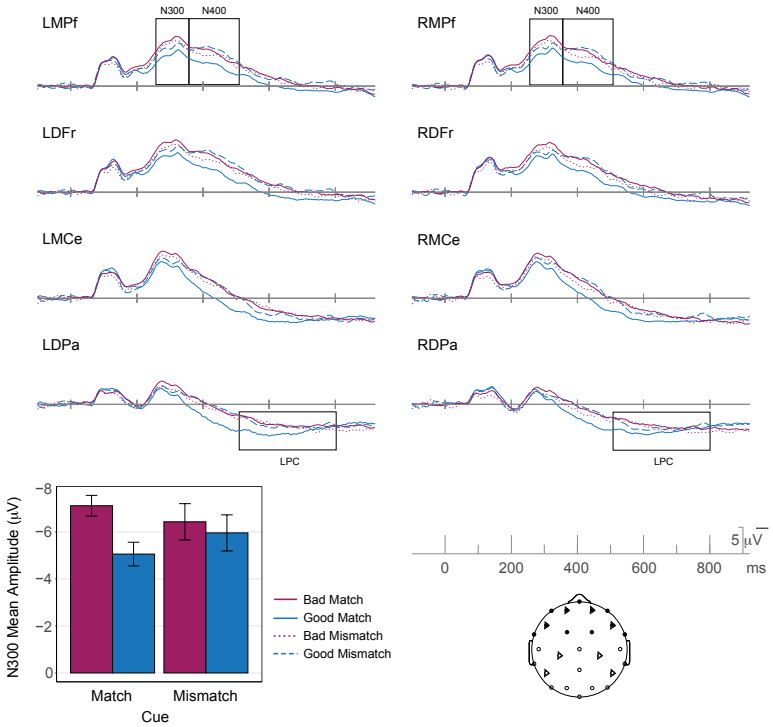

Figure 3. Grand average ERP waveforms for the good-match (solid-blue), bad-match (solid-maroon), good-mismatch (dashed-blue), and bad-mismatch (dotted-maroon) conditions in **Experiment 2** are shown at the same 8 representative electrode sites. In the match condition, responses to good and bad exemplars differ in the N300 time-window (250-350 ms), with greater negativity for bad exemplars as compared to good exemplars, over frontal sites (darkened electrode sites on the schematic of the scalp). In the mismatch condition, the differences between good and bad exemplars on the N300 are diminished/eliminated. The bar plot gives the grand average mean of the ERP amplitude over the 11 frontal electrode sites (darkened electrode sites on the schematic of the scalp) used for the primary statistical analyses (N = 20). The plotted error bars are within-subject confidence intervals.

Table 2. The grand average mean values, in the N300 time-window (250-350 ms), shown for 11 frontal electrode sites (see **Figure 3**), along with t-test and Bayes factor values. There is strong evidence (large Bayes factor) for greater negativity of the N300 for bad exemplars as compared to good exemplars when the cue matches the stimulus. When there is a mismatch between the cue and the stimulus there is no evidence (small Bayes factor) for the difference between good and exemplars in the N300 time-window. The t- test and Bayes factor calculations compared the within subject Good/Bad difference to 0.

Condition	Cue	N	Mean (μV)	Mean Bad/ Good Difference (μV)	Bad/Good Difference 95% C.I.	t(19)	p	Bayes Factor
Bad	Match	20	-7.1±0.94	-2.06	-2.6 to -1.5	-7.4	5.6E-07	30457
Good	Match	20	-5.1±1.07					
Bad	Mismatch	20	-6.4±1.64	-0.47	-1.7 to 0.73	-0.82	0.42	0.31
Good	Mismatch	20	-6.0±1.64					

Note: ± values reflect the normed standard deviation within subjects.

Post N300 Components

Again, for completeness, we also examined effects on the N400 (350-500 ms) and Late Positive Complex (LPC) (500-800 ms). These are presented in full in the **Supplementary Information** and summarized here. Given prior work (reviewed in

(Kutas & Federmeier, 2011)), we expected the N400 to be particularly sensitive to the match between the verbal cue and the scene category. Indeed, overall, N400 responses to good scenes that matched the verbal cue were facilitated (more positive: -3.5 μV) than to good scenes that mismatched their cues (-5.6 μV), consistent with the large literature on N400 semantic priming (See Table S3). Moreover, we replicated the effect in **Experiment 1**: N400 amplitudes were also larger for bad (-5.3 μV) than for good exemplars (-3.5 μV) in the match condition, although, again, we cannot rule out influence from the prior N300 effects on the observed pattern. We see an interaction of Good/Bad x Cuing in the N400 window ($F = 18.56$; $p = 0.0004$; $E = 1$), with the largest facilitation for good exemplars in the match condition. LPCs were larger (more positive) for good exemplars in the match condition (2.7 μV) compared to both bad exemplars (0.4 μV) in the match condition (replicating **Experiment 1**) and to either scene type in the mismatch condition (Good: 0.22 μV; Bad: 0.99 μV), presumably reflecting the increased ease and confidence of responding to the good match items (see **Table S4**).

Discussion

We proposed that the N300 component of the ERP reflects the impact of hierarchical predictive coding on late stage visual processing during which global features of the stimulus are being processed. Across many studies, larger (more negative) N300 responses have been observed for conditions that might be characterized as statistically irregular (Mudrik et al., 2010; Pietrowsky et al., 1996; Schendan & Kutas, 2002, 2003, 2007; Vo & Wolfe, 2013). However, thus far, the literature has examined only the response to objects, objects in scenes, or to artificially degraded stimuli. If the N300

more broadly reflects a late visual stage predictive coding, then it should be sensitive to statistical regularity represented at later visual processing stages more generally, not just in the context of object processing and artificial manipulations of global structure. To this end, here we showed that the N300 is also sensitive to the difference between good (statistically regular) and bad (statistically irregular) exemplars of natural scenes. Predictive coding posits a larger inference error in processing statistically irregular items (bad exemplars) as compared to statistically regular items (good exemplars), and, consistent with this, in **Experiment 1**, N300 responses were larger for the statistically irregular exemplars. These data, in combination with prior experiments, show that the N300 indeed manifests the expected response properties for a general index of predictive coding mechanisms for late stage visual processing of complex objects and scenes.

Across the literature, the kinds of stimuli distinguished by the N300 do not just differ in low level attributes but encompass global structure, canonical viewpoints, probable views of objects in scene contexts, and, in our own experiment, the category-level representativeness of the stimuli. We would like to collectively refer to these properties as learned statistical regularities. It is this collection of statistical regularities, which might be called a *template*, that can aid the rapid categorization and identification of stimuli. Thus, we can think of the differences on the N300 component as an indicator of the degree to which an incoming exemplar can be matched with a template, with greater negativity for a stimulus when it doesn't match a template as compared to when it does.

In **Experiment 1**, neither scene category nor exemplar status (good or bad) was predictable from trial to trial, and thus the statistical regularity driving the observed effect must have been acquired over the life time (i.e., learning what does and does not

411 constitute a good exemplar of a category), rather than within the context of the
412 experiment. However, a key attribute of PHT models, of which predictive coding is a
413 popular example, is that the hypotheses that are generated are sensitive to the current
414 context. If the N300 reflects a template matching process, such that the input is
415 compared against a contextually relevant learned statistical regularity, then the N300
416 sensitivity to statistical regularity should vary by context.

417 In **Experiment 2,** therefore, we set up expectations for a particular category on each trial
418 using a word cue, with the aim of pre-activating a particular scene category template.
419 Critically, on 25% of trials the scene did not match the cued category. We found that the
420 N300 is indeed sensitive to regularities cued by the current context. When the scenes,
421 both good and bad exemplars, were congruent with the cued category, we observed a
422 significant effect of statistical regularity (good versus bad) in the N300 time-window,
423 replicating the results from **Experiment 1**. Here the good exemplar provides a better
424 match to the activated template than the bad exemplars, and thus the reduced inference
425 error or iterative matching is reflected in the amplitude of the N300. In the mismatching
426 condition, however, the presented stimulus, again whether a good or bad exemplar of its
427 own category, does not match the cued template (e.g., a “Forest” template has been
428 cued but a good or bad beach scene was presented). In this case, notably, we failed to
429 observe a reliable difference between the N300 to good and bad exemplars. In the
430 language of predictive coding models, similar inference errors would be generated for
431 both statistically regular (good) and irregular (bad) exemplars that mismatch the
432 activated template, as they would both violate the predicted regularities or at least
433 neither good nor bad exemplars of another category should violate the predicted
434 regularities more than the other. Thus, the N300 not only indexes statistical regularities

learned over a lifetime, but also shows sensitivity to the expectations (predictions) generated within the current context. Thus, again the N300 conforms to characteristics expected of a neural signature of predictive coding.

The N300 Indexes Perceptual Hypothesis Testing

We can think of visual identification and categorization as a cascade of processes, starting with identification of low level visual features, followed by perceptual grouping of features, and then appreciation of the "whole" visual form of objects and scenes, after which processing moves beyond the visual modality into multi-modal semantics and decision making. PHT mechanisms can work within and across each of these stages. In the context of object processing, prior work on the N300 has posited it as an index of object model selection, an intermediate stage in the process of object identification and categorization (Schendan, 2019; Schendan & Kutas, 2002, 2003, 2007). Having extended the N300 differences to natural scenes, we propose that the N300 reflects PHT mechanisms in this intermediate stage more broadly, not just object selection. Similar to other work (Schendan, 2019), we believe that the N300 reflects processing at the point wherein the input is matched to items in memory with similar perceptual structures. However, our data show that this process is not limited to objects and that it makes use of variety of statistical regularities learned from the world, including those critical for processing both objects and scenes.

The broadened view of the N300 as being reflective of a general visual template matching process would suggest that its source be occipitotemporal visual areas. Indeed, the N300 response to objects has been source localized to occipitotemporal visual areas (Schendan & Lucia, 2010; Sehatpour et al., 2006). Although the N300 for

458 scenes has not yet been source localized, a high-density ERP study on scene
459 categorization localized activity in the 200-300 ms time window to these same
460 occipitotemporal visual areas (Greene & Hansen, 2020). Moreover, our prior fMRI work
461 with good and bad scene exemplars (Torralbo et al., 2013) would suggest that the N300
462 for scenes originates in the PPA, a region known to preferentially process natural scenes
463 (Epstein & Kanwisher, 1998). Using the same good and bad scene exemplars as in our
464 experiments, we found that, in the PPA, bad exemplars elicited a greater BOLD signal
465 than good exemplars (Torralbo et al., 2013), mirroring the effect we observed for the
466 N300. Interestingly, in that same PPA region of interest we observed that good
467 exemplars were better decoded than bad exemplars; that is, we were better able to
468 predict the scene category presented on the basis of activity patterns when the scene
469 was a good exemplar than when it was bad (Torralbo et al., 2013). In other words, it was
470 not the case that reduced activity for good exemplars reflected a weaker representation
471 but instead likely reflected a more efficient representation, an interpretation that aligns
472 nicely with our characterization of the N300 effect as one of visual template matching in
473 occipitotemporal cortex. These templates can be rapidly activated via pattern completion
474 mechanisms in occipitotemporal cortex (Cowell et al., 2019; Cutler et al., 2019; Hindy et
475 al., 2016; Norman & O'Reilly, 2003). We suggest that the N300 may be interpreted as a
476 component that reflects the iterative processing, as posited by PHT, in occipitotemporal
477 cortical regions, which helps match previously learned regularities of objects and scenes
478 with the incoming stimulus.

479 Although we are arguing that the N300 indexes PHT for late stage visual processing of
480 complex visual objects and scenes, it is possible that other components could index PHT
481 at other stages of processing. For example, PHT matching low level sensory features,

such as gratings (Kok et al., 2012), to hypotheses about such low level features should occur at earlier stages in the processing hierarchy. These PHT processes maybe reflected in other ERP components, such as the visual mismatch negativity (Oxner et al., 2019), or even earlier ERP components that index expected features, like the P1 (Boutonnet & Lupyan, 2015).

The N300 may thus index the "final" stage of visual PHT, at the transition into multimodal, semantic processing. Immediately after the N300, ERP responses to complex objects and scenes are characterized by an N400, which we also observe in our experiment. The N400 is widely accepted as a signature of multi-modal semantic processing, elicited by not only visual words and pictures, but also meaningful stimuli in other modalities (see review Kutas & Federmeier, 2011), whereas the N300 seems to be about visual perceptual structure (Schendan, 2019; Schendan & Kutas, 2002, 2003, 2007). In some cases, it may be difficult to disentangle the precise contributions of the N300 and N400 to observed effects of object categorization and match to object knowledge (Gratton et al., 2009; Schendan, 2019; Schendan & Maher, 2009) since the N400 is known to be sensitive to the fit between, e.g., a picture and its context (Federmeier & Kutas, 2002; Ganis et al., 1996). Importantly, however, this does not impact the critical effect of our good versus bad scenes, as neither contain contextually inappropriate items, nor, in **Experiment 1,** did we set up any context prior to an image (i.e., the scene category is unpredictable).

Conclusion

We have argued that the N300 serves as an index of PHT at the level of whole-objects and scenes. Using statistically regular and irregular exemplars of natural scenes, we

showed that items that do not match our known regularities, the statistically irregular exemplars, and hence should lead to larger inference errors evoked a large N300 amplitude than statistically regular exemplars. We not only showed that the N300 is sensitive to current context, by pre-activating a context with a verbal cue, but that it behaves like a template matching process in which statistically regular images procure their advantage by virtue of matching the current contextual prediction; that is, the N300 was differentially sensitive to good and bad exemplars of a category only when that category was expected. Our work not only extends prior work on the N300 to natural scenes but it suggests that the N300 reflects a general template/model selection process of the sort proposed by PHT models, such as predictive coding. We propose that the N300 indexes visual inference processing in a late visual time-window that occurs at the boundary between vision and the next stage of multi-modal semantic processing. Further studies will be needed to explore the full range of the N300 response. For example, does it require that the object or scene is attended or might it proceed more automatically. Can it be modulated by contexts setup in different modalities (e.g., auditory inputs: speech, sounds)? Regardless, we propose that the N300 can serve as a useful marker of knowledge guided visual processing of objects and scenes, with templates based on prior knowledge serving as hypotheses for visual inference as posited by PHT.

Methods

Participants

The data for **Experiment 1** came from 20 right-handed college-age subjects (mean age = 24.36 years, range = 18 to 33 years, 12 women), and the data for **Experiment 2** from a separate set of 20 right-handed subjects (mean age = 22.44; range 18-30 years; 14 women). In both experiments, participants gave written, informed consent and were compensated for their participation in the study with course credit or cash. The study was approved by the Institutional Review Board of the University of Illinois at Urbana-Champaign. All participants were right-handed, as assessed by the Edinburgh Inventory (Oldfield, 1971) and none had a history of neurological disease, psychiatric disorders, or brain damage.

Materials and Procedures

ERP-eliciting stimuli were pictures of natural scenes from six categories: beaches, forests, mountains, city streets, highways and offices (**Figure 1C**). These images were collected from the internet and rated for their representativeness of the named category on Amazon Mechanical Turk (Torralbo et al., 2013). The 60 top rated images were used as good exemplars for each category, and the 60 lowest rated images were used as bad exemplars for each category (for details on the choice of good and bad exemplars see (Torralbo et al., 2013)). Images were resized to 340 x 255 pixels and presented on a black background with a fixation cross at the center. The images were randomly presented at one of three locations: the center of the scene, 2 degrees to the left of

fixation, or 2 degrees to right of fixation, with a total of 120 good images and 120 bad images presented at each location. Here, we report only results for centrally-presented images (lateral presentation was used to separately answer questions about hemispheric biases in scene processing that are outside the scope of this manuscript).

In **Experiment 1**, participants were instructed at the beginning of the study that they would be seeing good and bad exemplars of six scene categories and that their task at the end of each trial was to indicate via button press whether the image was a good or a bad exemplar of its category. Participants first practiced with 9 trials to acclimatize to the task environment, and these images were not repeated in the main experiment. Then, they completed 3 blocks each consisting of an equal number of trials, for a total of 240 centrally presented trials (trials were also presented to the left and right visual fields in each block). The trial counts for centrally presented stimuli, for each category (good and bad combined) are as follows: beaches = 39; cities = 41; forests = 38; highways = 42; mountains = 36; offices = 44. Participants were seated at a distance of 100 cm from the screen, and the images subtended a visual angle of 7.65° x 5.73° (width x height). Subjects were instructed to maintain fixation on the central fixation cross and to try to minimize saccades and eye blinks during stimulus presentation. As depicted in **Figure 1A**, each trial began with a fixation cross presented on a blank screen for a duration jittered between 1000-2000 seconds (to reduce the impact of slow, anticipatory components on the ERP signal). The scene image, either a good exemplar or a bad exemplar from one of the six categories, was presented for a duration of 200 ms, followed by a fixation cross on a blank screen for 500 ms. At the end of the trial a prompt with "Good or Bad?" was displayed on the screen, and participants pressed one of two

response buttons, held in each hand (counterbalanced across participants), to indicate their judgment. The experiment lasted for approximately one hour and fifteen minutes. Subjects were given two five-minute breaks at roughly 25 minutes and 60 minutes from the start of the experiment.

Experiment 2 was identical to **Experiment 1**, except that each trial began with a word cue, presented for 500 ms (**Figure 1B**), which corresponded to one of the six scene categories used in the experiment: Beach, City Street, Forest, Highway, Mountain, and Office. For each category, we ensured that five trials of each type (good and bad exemplars) were mismatched. There were thus 75% matched trials (15 trials each of good and bad within each of the six scene categories) and 25% mismatched trials, for a total of 90 matched trials and 30 mismatched trials. Instead of making a good or bad judgment, at the end of each trial participants were prompted to respond “yes” or “no,” with a button press, to the question of whether or not the picture had matched the cue. Hand used to respond “yes” or “no” was counterbalanced.

<u>ERP Setup and Analysis</u>

EEG was recorded from 26 channels of passive electrodes that were equidistantly arranged on the scalp, referenced online to the left mastoid and re-referenced offline to the average of the left and right mastoids. Additional electrodes placed on the outer cantus of each eye and on the orbital ridge below the left eye were used to monitor saccadic eye movements and blinks. Impedances were kept below 5 KΩ for scalp channels and 10 KΩ for eye channels. The signal was bandpass filtered online (0.02 Hz

- 100 Hz) and sampled at 250 Hz. Trials with artifacts due to horizontal eye movements or signal drift were rejected using fixed thresholds calibrated for individual subjects. Trials with blinks were either rejected, or, for subjects with higher numbers of blink artifacts (12 in **Experiment 1** and 8 in **Experiment 2**), were corrected using a blink correction algorithm (Dale, 1994). We confirmed that the analytical results were unchanged if blinks were rejected instead of corrected. On average, in **Experiment 1**, 6.83% of good exemplar trials and 9.04% of bad exemplar trials were rejected due to artifacts and no condition had fewer than 63 trials per subject in the analysis. In **Experiment 2**, in the match condition, 10.8% of good exemplar trials and 11.09% of bad exemplar trials were rejected due to artifacts and no condition had fewer than 56 trials per subject in the analysis. In the mismatch condition, 10.38% of good exemplar trials and 13.89% of bad exemplar trials were rejected due to artifacts and no condition had fewer than 19 trials per subject in the analysis.

ERPs were epoched for a time period spanning 100 ms before stimulus onset to 920 ms after stimulus onset, with the 100 ms prestimulus interval used as the baseline. This processed signal was then averaged for each condition within each subject. A digital bandpass filter (0.2 Hz - 30 Hz) was applied before measurements were taken from the ERPs. Based on prior work showing that the N300 is frontally distributed and occurs between 250 ms to 350 ms (Federmeier & Kutas, 2001; Schendan & Kutas, 2002, 2003), we measured N300 mean amplitudes in this time window across the 11 frontal electrode sites: MiPf, LLPf, RLPf, LMPf , RMPf , LDFr , RDFr , LMFr , RMFr , LLFr , and RLFr (first letter: R=right, L=left, Mi=midline; second letter: L=lateral, M=medial, D=dorsal; Pf = prefrontal and Fr= frontal). Statistics were computed using R (R Core

Team, 2020). Specifically, we used the functions t.test, to compute t-tests, and ttestbf (from the package: BayesFactor) to compute Bayes Factors. The t- test and Bayes factor calculations compared the measured condition difference to 0. For within-subject calculations of confidence intervals, we used the function summarySEwithin() that is based on (Morey, 2008).

For completeness, we also analyzed two ERP components in the time-window after the N300: the N400 and the Late Positive Complex (LPC). Prior work examining the N400 to pictures has shown a frontal distribution (Ganis et al., 1996), and thus we again used the 11 frontal electrode sites, but now in the time-window 350-500 ms. For the LPC we chose posterior sites in the time-window of 500-800 ms based on prior work characterizing the distribution and timing of the LPC (Finnigan et al., 2002).

Acknowledgments

We would like to thank Yanqi Zhang for assistance with running subjects in **Experiment 1**, and Resh Gupta, and Nirupama Mehrotra for helping with **Experiment 1** data collection. We also thank Rami Alsaqri, Johan Saelens, Daria Niescierowicz, and Benjamin D. Schmitt for helping with data collection in **Experiment 2**. We would like to acknowledge funding provided by ONR MURI (DMB) and by NIH (R01 AG026308) and the James S. McDonnell foundation (KDF).

References

Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience*, *35*(25), 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Caddigan, E., Choo, H., Fei-Fei, L., & Beck, D. M. (2017). Categorization influences detection: A perceptual advantage for representative exemplars of natural scene categories. *Journal of Vision*, *17*(1), 21. https://doi.org/10.1167/17.1.21

Caddigan, E., Walther, D. B., Fei-Fei, L., & Beck, D. M. (2010). Perceptual differences between natural scene categories. OPAM 2010 18th Annual Meeting. *Visual Cognition*, *18*(10), 1498–1502.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477

Cowell, R. A., Barense, M. D., & Sadil, P. S. (2019). A Roadmap for Understanding Memory: Decomposing Cognitive Processes into Operations and Representations. *ENeuro*, *6*(4). https://doi.org/10.1523/ENEURO.0122-19.2019

Cutler, R. A., Duff, M. C., & Polyn, S. M. (2019). Searching for Semantic Knowledge: A Vector Space Semantic Analysis of the Feature Generation Task. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00341

Dale, A. M. (1994). *Source localization and spatial discriminant analysis of event-related potentials: Linear approaches (brain cortical surface).* Dissertation Abstracts International.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

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. http://psycnet.apa.org/journals/xlm/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. http://kutaslab.ucsd.edu/people/kutas/pdfs/2002.N.730.pdf

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. http://www.sciencedirect.com/science/article/pii/S0028393202001136

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The Search for "Common Sense": An Electrophysiological Study of the Comprehension of Words and Pictures in Reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89

Gratton, C., Evans, K. M., & Federmeier, K. D. (2009). See what I mean? An ERP study of the effect of background knowledge on novel object processing. *Memory & Cognition*, *37*(3), 277–291. https://doi.org/10.3758/MC.37.3.277

Greene, M. R., & Hansen, B. C. (2020). Disentangling the Independent Contributions of Visual and Conceptual Features to the Spatiotemporal Dynamics of Scene

Categorization. *The Journal of Neuroscience*, JN-RM-2088-19. https://doi.org/10.1523/JNEUROSCI.2088-19.2020

Gregory, R. L. (1980). Perceptions as Hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090

Helmholtz, H. von. (1925). *Treatise on physiological optics, Bd. 3: The perceptions of vision* (English translation of the 3rd edition., Vol. 3). The Optical Society of America. https://echo.mpiwg-berlin.mpg.de/ECHOdocuView?url=/permanent/library/HS7FH69N/pageimg&tocMode=thumbs&viewMode=index&pn=17&mode=imagepath

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), 665–667. https://doi.org/10.1038/nn.4284

Hochberg, J. (1981). On cognition in perception: Perceptual coupling and unconscious inference. *Cognition*, *10*(1), 127–134. https://doi.org/10.1016/0010-0277(81)90035-4

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lange, F. P. de, Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Lauer, T., Willenbockel, V., Maffongelli, L., & Võ, M. L.-H. (2020). The influence of scene and object orientation on the scene consistency effect. *Behavioural Brain Research*, *394*, 112812. https://doi.org/10.1016/j.bbr.2020.112812

Lupyan, G. (2017). Changing What You See by Changing What You Know: The Role of Attention. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00553

Mcpherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65. http://onlinelibrary.wiley.com/doi/10.1017/S0048577299971196/abstract

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005)." 4.2 (2008):61. Web. *Reason*, 61.

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

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oxner, M., Rosentreter, E. T., Hayward, W. G., & Corballis, P. M. (2019). Prediction errors in surface segmentation are reflected in the visual mismatch negativity, independently of task and surface features. *Journal of Vision*, *19*(6), 9–9. https://doi.org/10.1167/19.6.9

Pietrowsky, R., Kuhmann, W., Krug, R., Molle, M., Fehm, H. L., & Born, J. (1996). Event-related brain potentials during identification of tachistoscopically presented pictures. *Brain and Cognition*, *32*(3), 416–428. http://www.sciencedirect.com/science/article/pii/S0278262696900743

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* https://www.R-project.org/

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rock, I. (1983). *The Logic Of Perception*. Cambridge: MIT Press.

Schendan, H. E. (2019). Memory influences visual cognition across multiple functional states of interactive cortical dynamics. In *Psychology of Learning and Motivation* (Vol. 71, pp. 303–386). Elsevier. https://doi.org/10.1016/bs.plm.2019.07.007

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931–945. http://www.sciencedirect.com/science/article/pii/S0028393201001762

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. http://www.mitpressjournals.org/doi/abs/10.1162/089892903321107864

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. http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2007.19.5.734

Schendan, H. E., & Lucia, L. C. (2010). Object-sensitive activity reflects earlier perceptual and later cognitive processing of visual objects between 95 and 500ms. *Brain Research*, *1329*, 124–141. https://doi.org/10.1016/j.brainres.2010.01.062

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. https://doi.org/10.1016/j.neuroimage.2008.09.061

Sehatpour, P., Molholm, S., Javitt, D. C., & Foxe, J. J. (2006). Spatiotemporal dynamics of human object recognition processing: An integrated high-density electrical mapping and functional imaging study of “closure” processes. *NeuroImage*, *29*(2), 605–618. https://doi.org/10.1016/j.neuroimage.2005.07.049

Smith, M. E., & Loschky, L. C. (2019). The influence of sequential predictions on scene-gist recognition. *Journal of Vision*, *19*(12), 14–14. https://doi.org/10.1167/19.12.14

Spratling, M. W. (2010). Predictive Coding as a Model of Response Properties in Cortical Area V1. *Journal of Neuroscience*, *30*(9), 3531–3543. https://doi.org/10.1523/JNEUROSCI.4911-09.2010

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing*, *17*(3), 279–305. https://doi.org/10.1007/s10339-016-0765-6

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. https://doi.org/10.1126/science.1132028

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. https://doi.org/10.1371/journal.pone.0058594

Vo, M. L.-H., & Wolfe, J. M. (2013). Differential Electrophysiological Signatures of Semantic and Syntactic Scene Processing. *Psychological Science*, *24*(9), 1816–1823. https://doi.org/10.1177/0956797613476955

Voss, J. L., Federmeier, K. D., & Paller, K. A. (2012). The potato chip really does look like Elvis! Neural hallmarks of conceptual processing associated with finding novel shapes subjectively meaningful. *Cerebral Cortex (New York, N.Y.: 1991)*, *22*(10), 2354–2364. https://doi.org/10.1093/cercor/bhr315

Voss, J. L., & Paller, K. A. (2007). Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. *Learning & Memory*, *14*(4), 259–267. https://doi.org/10.1101/lm.529807

Voss, J. L., Schendan, H. E., & Paller, K. A. (2010). Finding meaning in novel geometric shapes influences electrophysiological correlates of repetition and dissociates perceptual and conceptual priming. *NeuroImage*, *49*(3), 2879–2889. https://doi.org/10.1016/j.neuroimage.2009.09.012

The N300: An Index For Predictive Coding Of Complex Visual Objects and Scenes

Manoj Kumar, Kara D. Federmeier, Diane M. Beck

Corresponding Author: Manoj Kumar

Email: mk35@princeton.edu

This PDF file includes:

Supplementary text

Tables S1 to S4

SI References

824 **Supplementary Information**

825 **Experiment 1**

826 **N300 Distributional Analysis**

827 To compare the N300 in our experiment with its characterization in the existing literature,
828 we also performed an ANOVA in the time-window 250-350 ms across the 16 scalp
829 channels that could be divided into 2 levels of Hemisphere (right and left scalp sites), 2
830 levels of Laterality (lateral and medial scalp sites), and 4 levels of Anteriority (prefrontal,
831 frontal, central/parietal, and occipital scalp sites). In addition to confirming the main
832 effect of Good vs. Bad (bad larger than good) ($F(1,19) = 11.97$; $p=0.0026$, $E=1$), there
833 were interactions of Good/Bad with both Laterality ($F(1,19)=7.35$; $p=0.014$; $E=1$) and
834 Anteriority ($F(3,57) = 11.97$; $p<0.0001$; $E=0.5336$), as well as three-way interactions:
835 Good/Bad x Hemisphere x Anteriority ($F(3,57)=3.13$; $p=0.0324$; $E=0.8353$), and
836 Good/Bad x Laterality x Anteriority ($F(3,57)=6.79$; $p=0.0005$; $E=0.7571$). Overall, N300
837 responses were observed over both left and right hemisphere sites and were largest
838 over the front of the head and larger over medial compared to lateral electrode sites.
839 Over frontal sites, the effect of Laterality was more pronounced and there was a
840 tendency for larger effects over left compared to right hemisphere sites. These results
841 show that the topographic distribution of the N300 effect for natural scenes has a similar
842 distributional profile as the N300 for objects (Schendan & Kutas, 2002, 2003, 2007).

843 **Post N300 Components**

844 **Table S1.** The grand average mean values in the N400 time-window (350-500 ms),
845 shown for 11 frontal electrode sites along with t-test and Bayes factor values. The N400

for the bad exemplars is larger (more negative) than that for the good exemplars. The t-test and Bayes factor calculations compared the within-subject Good/Bad difference to 0.

Condition	N	Mean (μV)	Bad/Good Difference Mean (μV)	Bad/Good Difference 95% C.I.	t(19)	p	Bayes Factor
Bad	20	-3.3±0.96	-1.14	-1.78 to -0.50	-3.74	0.0014	27.3
Good	20	-2.2±0.96					

Note: ± values reflect the normed standard deviation within subjects.

LPC

Table S2: The grand average mean values, in the LPC time-window (500-800 ms), shown for 15 posterior electrode sites (LMCe, RMCe, LDCe, RDCe, MiCe, MiPa, LLTe, RLTe, LDPa, RDPa, LLOc, RLOc, LMOc, RMOc, MiOc) along with t-test and Bayes factor values. The LPC for bad exemplars has a smaller mean amplitude than for the good exemplars. The t- test and Bayes factor calculations compared the within subject Good/Bad difference to 0.

Condition	N	Mean (μV)	Bad/Good Difference Mean (μV)	Bad/Good Difference 95% C.I.	t(19)	p	Bayes Factor
Bad	20	3.3±1.1	-1.17	-1.91 to -0.43	-3.3	0.0036	12.1
Good	20	4.5±1.1					

Note: ± values reflect the normed standard deviation within subjects.

Experiment 2

N300 Distributional Analysis with Cuing

A novel feature of **Experiment 2** was the use of a verbal precue. Prior work on the N300, including our own **Experiment 1**, has not used a verbal precue. Thus, to see how the distributional properties of the N300 effect under cued conditions compare to that in prior work, we performed an ANOVA analysis using identical electrode sites as in the ANOVA analysis for **Experiment 1**, with identical groupings: 2 levels of Hemisphere (right and left scalp sites), 2 levels of Laterality (lateral and medial scalp sites), and 4 levels of Anteriority (prefrontal, frontal, central/parietal, and occipital scalp sites). We replicate the main effect of **Experiment 1** with Good vs. Bad (bad larger than good; $F(1,19) = 15.34$; $p= 0.0009$, $E=1$). The topographic distribution of this effect is significant

in the following interactions: Good/Bad x Hemisphere ($F(1,19) = 5$; $p = 0.0375$, $E = 1$); Good/Bad x Laterality ($F(1,19) = 21.1$; $p = 0.0002$, $E = 0.5336$); Good/Bad x Anteriority ($F(3,57) = 7.13$; $p = 0.0004$, $E = 0.4030$). N300 effects were observed over both left and right hemisphere sites, with the left hemisphere showing somewhat larger effects as compared to the right hemisphere, and, as in **Experiment 1**, were larger over medial than lateral sites and larger over the front of the head.

We also see an interaction of Good/Bad x Cueing ($F(1,19) = 5.87$; $p = 0.0255$, $E = 1$) with large N300 effects the stimuli matched the cue, but negligible effects in the mismatch condition. There was a 3-way interaction of Good/Bad x Cueing x Laterality ($F(1,19) = 4.83$; $p = 0.0406$, $E = 1$), because N300 effects were larger over medial than lateral sites in the match condition, with the effect being eliminated in the mismatch condition.

N400: The Effects of Cuing

The N400 indexes a match to semantic expectancy, and we performed an ANOVA analysis in the N400 time-window (350-500 ms) to examine the effects of semantic expectancy in **Experiment 2**, although, we do not claim that the N400 effects are completely independent of the preceding N300 effects in our experiment. Prior work has shown that the N300 and N400 are functionally dissociable (Federmeier & Kutas, 2002; Gratton et al., 2009), but their similar response pattern to, e.g., incongruent and congruent items, can make separating them challenging under some circumstances (Draschkow et al., 2018). For the N400, we see a main effect of Cuing ($F = 13.35$; $p = 0.0017$; $E = 1$) with a smaller N400 amplitude for the matched condition as compared to

the mismatched condition. We also see a significant interaction of Good/Bad x Cuing (F =18.56; p =0.0004; E =1), with the good exemplars in the match condition having the smallest amplitude as compared to the good and bad exemplars in the match- and-mismatched condition. There is also a 3-way interaction of Good/Bad x Laterality x Anteriority (F(3,57) = 2.96; p = 0.0411; E = 0.8875) in the N400 time-window.

Table S3. The grand average mean values, in the N400 time-window (350-500 ms), shown for 11 frontal electrode sites. The t- test and Bayes factor calculations compared the within subject Good/Bad difference to 0.

Condition	Cue	N	Mean (μV)	Difference (μV)	Bad/Good Difference 95% C.I.	t(19)	p	Bayes Factor
Bad	Match	20	-5.3±0.98	-1.82	-2.5 to -1.15	-5.67	1.8E-05	1287.4
Good	Match	20	-3.5±1.41					
Bad	Mismatch	20	-4.7±1.65	-0.47	-0.42 to 2.25	-1.44	0.17	0.56
Good	Mismatch	20	-5.6±2.06					

Note: ± values reflect the normed standard deviation within subjects.

LPC

For the LPC, we examined differences in the 500-800 ms time window, encompassing the Late Positive Complex (LPC), over 15 posterior sites (identical to the sites for the analysis in **Experiment 1**). We replicate the main effect of Good/Bad with a larger LPC amplitude for good as compared to bad exemplars (F = 4.84; p = 0.0403; E =1). The LPC is known to index confidence in decision making (Finnigan et al., 2002) and the main effect of Cuing in **Experiment 2** aligns with this understanding**,** with the match condition showing a larger LPC amplitude as compared to the mismatch condition (F = 19.69; p =0.0003; E=1). The interaction of Good/Bad x Cuing is also significant (F =9.82; p =0.0003; E =1) with the good match having the largest LPC amplitude as compared to the Good mismatch, Bad match, and Bad mismatch conditions.

Table S4. The grand average mean values, in the LPC time-window (500-800 ms), shown for 15 posterior electrode sites. The t- test and Bayes factor calculations compared the Good/Bad difference to 0.

Condition	Cue	N	Mean (μV)	Mean Bad/Good Difference (μV)	Bad/Good Difference 95% C.I.	t(19)	p	Bayes Factor
Bad	Match	20	0.4±0.98	-2.24	-2.94 to 1.54	-6.69	2.1E-06	9014.4
Good	Match	20	2.7±1.09					
Bad	Mismatch	20	0.9±1.66	0.68	-0.56 to 1.93	1.15	0.26	0.41
Good	Mismatch	20	0.22±1.61					

Note: ± values reflect the normed standard deviation within subjects.

SI References

Draschkow, D., Heikel, E., Võ, M. L.-H., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia*, *120*, 9–17. https://doi.org/10.1016/j.neuropsychologia.2018.09.016

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*(7), 730–747. http://kutaslab.ucsd.edu/people/kutas/pdfs/2002.N.730.pdf

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. http://www.sciencedirect.com/science/article/pii/S0028393202001136

Gratton, C., Evans, K. M., & Federmeier, K. D. (2009). See what I mean? An ERP study of the effect of background knowledge on novel object processing. *Memory & Cognition*, *37*(3), 277–291. https://doi.org/10.3758/MC.37.3.277

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931–945. http://www.sciencedirect.com/science/article/pii/S0028393201001762

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. http://www.mitpressjournals.org/doi/abs/10.1162/089892903321107864

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. http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2007.19.5.734