Cerebral Cortex Advance Access published February 24, 2010

Cerebral Cortex
doi:10.1093/cercor/bhq021

Moderate Excitation Leads to Weakening of Perceptual Representations

Ehren L. Newman[1] and Kenneth A. Norman[2]

[1]Center for Memory and Brain, Boston University, Boston, MA 02215, USA and [2]Department of Psychology, Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA

Address correspondence to Ehren L. Newman, Center for Memory and Brain, Boston University, 2 Cummington Street, Boston, MA 02215, USA. Email: enewman@gmail.com.

A fundamental goal of memory research is to specify the conditions that lead to the strengthening and weakening of neural representations. Several computational models of memory formation predict that learning effects should vary as a nonmonotonic function of the amount of excitation received by a neural representation. Specifically, moderate excitation should result in synaptic weakening, while strong excitation should result in synaptic strengthening. In vitro investigations of plasticity in rodents have provided support for this prediction at the level of single synapses. However, it remains unclear whether this principle scales beyond the synapse to cortical representations and manifests changes in behavior. To address this question, we used electroencephalogram pattern classification in human subjects to measure trial-by-trial fluctuations in stimulus processing, and we used a negative priming paradigm to measure learning effects. In keeping with the idea that moderate excitation leads to weakening, moderate levels of stimulus processing were associated with negative priming (slower subsequent responding to the stimulus), but higher and lower levels of stimulus processing were not associated with negative priming. These results suggest that the same principles that account for synaptic weakening in rodents can also account for diminished accessibility of perceptual representations in humans.

Keywords: cognition, EEG (electroencephalogram), pattern classification, learning and memory, plasticity

Introduction

One of the fundamental goals of learning and memory research is to specify the conditions that trigger the strengthening and weakening of neural representations (Hebb 1949). In this study, we tested the hypothesis that neural representations are weakened when the neurons that comprise those representations receive a moderate degree of excitation. Figure 1 illustrates this hypothesized relationship: For very low levels of postsynaptic excitation, connections into the postsynaptic neuron remain unchanged; for moderate levels of excitation, connections into that neuron from other active neurons are weakened; for higher levels of excitation, connections into that neuron from other active neurons are strengthened. Because the relationship between excitation and change in connection strength reverses as the level of excitation goes up, we refer to this as the "nonmonotonic plasticity hypothesis."

This nonmonotonic plasticity hypothesis has been instantiated in several neural network learning algorithms (Bienenstock et al. 1982; Diederich and Opper 1987; Gardner 1988; Vico and Jerez 2003; Senn and Fusi 2005; Norman et al. 2006). Simulation studies have used variants of this hypothesis to explain numerous neuroscientific findings (e.g., patterns of visual cortex plasticity; Cooper et al. 2004) and psychological findings (e.g., patterns of forgetting on episodic and semantic memory tests; Norman et al. 2007). Also, numerical and analytical studies have demonstrated that nonmonotonic learning algorithms can have desirable functional properties; for example, the nonmonotonic rule described by Diederich and Opper (1987) and Gardner (1988) yields higher associative memory storage capacity than the (monotonic) Hopfield learning rule (Hopfield 1982; Amit et al. 1985, 1987).

Currently, studies of individual synapses in rodents provide the most direct empirical support for the nonmonotonic plasticity hypothesis. For example, it has been found that moderate depolarizing currents and intermediate concentrations of postsynaptic Ca^{2+} ions (indicative of moderate excitatory input) generate long-term depression (i.e., synaptic weakening), whereas stronger depolarization and higher Ca^{2+} concentrations (indicative of greater excitatory input) generate long-term potentiation (i.e., synaptic strengthening) (Artola et al. 1990; Hansel et al. 1996).

The goal of this study was to test if the nonmonotonic plasticity hypothesis also applies at the level of distributed neural representations in humans. Simulation studies by Norman et al. (2006) and others suggest that nonmonotonic plasticity should scale up to neural ensembles: Moderate excitation of the neural ensemble responsible for representing a stimulus should lead to overall weakening of that ensemble (by weakening synapses within the ensemble and synapses coming into the ensemble). To test this prediction, we used pattern classifiers, applied to electroencephalogram (EEG) data, to measure stimulus processing (Haynes and Rees 2006; Philiastides and Sajda 2006; Philiastides et al. 2006), and we used a negative priming paradigm to measure learning effects (Tipper 1985).

In the negative priming paradigm, subjects are asked to ignore a stimulus (the prime) and later are asked to respond to that stimulus as quickly as possible; the basic negative priming effect is that subjects are slower to respond to the previously ignored prime than to a novel stimulus (for a review, see Fox 1995). The nonmonotonic plasticity hypothesis can explain this finding by positing that, on average, ignored primes are processed moderately and (consequently) are weakened, leading to slower subsequent processing. The key prediction of the nonmonotonic plasticity hypothesis, in this context, relates to how variability in processing of the ignored prime (across trials) should relate to the size of the negative priming effect: Moderately processed primes should show a robust negative priming effect, but primes that happen to receive lower or higher levels of processing should not show the predicted negative priming effect (see Fig. 1). This is the specific prediction that we set out to test in our experiment. We used EEG pattern classifiers to covertly measure processing of the ignored prime on each trial, and we related this trial-by-trial measure of prime processing to the size of the negative priming effect.

Materials and Methods

In this section, we first discuss the behavioral paradigm; next, we discuss our EEG data collection methods; after that, we discuss our EEG pattern classification methods; finally, we discuss the steps involved in our main analysis, where we related the output of the EEG pattern classifier to behavioral priming effects.

Subjects

Twenty subjects were recruited for this study using fliers hung around the Princeton University campus. Subjects were compensated with $30 for their participation in the 2-h session. Four subjects were excluded from all analyses because more than one-third of their trials (>200 out of 600) failed to meet the trial inclusion criteria outlined below. The remaining 16 subjects ranged in age from 19 to 38 years (mean 25.5); 4 of the 16 subjects were female; 14 of the 16 subjects reported themselves to be right-handed. Written informed consent was obtained in a manner approved by the Princeton Institutional Review Board.

Behavioral Paradigm

Figure 2 illustrates the design of the behavioral paradigm. Each trial included the following events (in order): a fixation cross (randomly sampled duration from 400 to 600 ms), a prime display (500 ms), a visual mask (1000 ms), a probe display (presented until subjects responded or until 5000 ms had passed), and a screen with the text "OK to blink" (1500 ms; not shown in Fig. 2). Both the prime and the probe displays consisted of a red-tinted centered target picture from 1 of 4 categories (face, house, shoe, or chair) superimposed on a grayscale offset distractor picture from 1 of the other 3 categories (see Stimulus Details below). Subjects had to indicate whether the target (red-tinted) stimulus in the probe display matched the target stimulus from the prime display. If the 2 stimuli were identical, subjects had to say "match"; otherwise, they had to name the category of the probe stimulus as quickly as possible. We measured priming effects by comparing reaction times from 2 types of trials: ignored repetition trials, where the probe target matched the prime distractor, and control trials, where the stimuli used in the prime display and the stimuli used in the probe display were sampled from different categories. Note that, on both ignored repetition and control trials, the prime target and probe target came from a different category. Thus, the correct response (to the probe) for both trial types was to name the category of the probe target image.

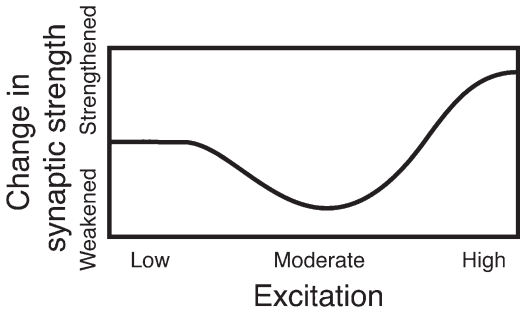

Figure 1. Hypothesized nonmonotonic relationship between the level of excitation of a neuron and modification of synapses involving that neuron. Moderate levels of excitation result in synaptic weakening, whereas stronger levels of excitation result in synaptic strengthening.

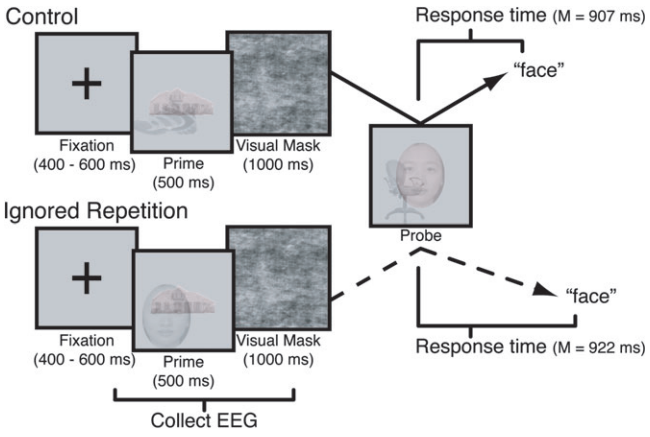

Figure 2. Negative priming task design comparing control trials and ignored repetition trials. In both the prime display and the probe display, subjects were instructed to attend to the centered red-tinted image (i.e., the target) and to ignore the offset grayscale image (i.e., the distractor). On control trials, the categories used in the probe display did not overlap with the categories used in the prime display. On ignored repetition trials, the probe target was identical to the prime distractor, and the probe distractor was sampled from a category that was not used in the prime display.

In addition to ignored repetition and control trials, we also included catch trials where the probe target was sampled from the same category as the prime target. Half of the catch trials used the exact same image for the prime target and the probe target (match trials), and half of the catch trials used different images from the same category (mismatch trials). These catch trials were included to ensure that subjects were performing the delayed-match-to-sample task.

The experiment was composed of six 100-trial blocks. Between blocks, subjects were given a break during which the experimenter checked that the subject was comfortable and alert. In the first 5 blocks, ignored repetition, control, match, and mismatch trials each made up one-sixth of the trials. In the remaining trials, the probe distractor was identical to the prime target; these trials were left over from an earlier (pilot) version of the experiment and were excluded from the analyses described here. The final (sixth) block of trials consisted of target images only (i.e., no distractors). These target-only trials were used to test the classifier's ability to detect the presence of each image category when presented as a target image, as described below, under Testing Classifier Sensitivity to the Target Category.

Stimulus Details

The stimuli consisted of 195 luminance-matched grayscale images from each of 4 categories: faces, houses, shoes, and chairs. The image order and the assignment of images to conditions (i.e., target vs. distractor on either the prime vs. the probe display) were randomized across subjects. Each stimulus was used on either 2 or 3 different trials within the experiment (the mean stimulus presentation frequency was 2.5 trials). Within a particular category, all of the stimuli were presented once before any stimuli were repeated across trials, and all of the stimuli were presented twice before any stimuli were presented on a third trial. For a given subject, each image was always used in the same way across repetitions (e.g., if an item served as a probe distractor in the control condition on its first presentation, it also served as a probe distractor in the control condition on its second presentation).

Both prime displays and probe displays were generated by superimposing a target image and a distractor image over a uniform gray background. On all displays, the target images were presented at 60% of their normal contrast value (by linearly combining the luminance values of the individual pixels of the image with a flat gray image in a ratio of 3:2). On prime displays, half of the distractors were presented at 50% contrast and the other half were presented at 30%. Multiple levels of contrast were used to ensure variability in the level of distractor processing. On probe displays, distractors were always presented at 60% of their normal contrast. The target was centered on the screen and was given a red tint by increasing the magnitude of the red channel of the RGB image by 10% of the maximum (see Fig. 2). The distractor image was offset from the target toward 1 of the 4 corners of the screen by 1.8° of visual angle (a new corner was chosen randomly for each trial). Each image occupied 5.7° of visual angle, and the total stimulus display subtended 7.5° of visual angle when presented on a 17-inch CRT monitor (100-Hz refresh rate) positioned at eye height 60″ away from the subject.

A visual mask was presented between the prime and probe displays on each trial to reduce the possibility that subjects could use the afterimage of the prime to perform the task. A unique mask was generated for each trial by recombining phase-shifted components of

a spatial Fourier decomposition of different images from each of the 4 categories. The experiment was presented using E-Prime 1.2 (Psychological Software Tools), run on a Windows PC.

EEG Data Acquisition

The EEG was recorded in an electrostatically shielded subject testing room from 77 Ag/AgCl scalp electrodes placed according to the international 10-20 system using custom 87-electrode caps (Electro-Cap International). Signals were recorded using two 64-channel preamplifiers from Sensorium Inc. with a bandpass of 0.02–300 Hz and digitized at 1000 Hz with two 64-channel 12-bit National Instruments data acquisition cards controlled by a custom Windows software package developed in the laboratory. All channels were referenced to the left mastoid (input impedance <5 kΩ) online. The signal was rereferenced to the common mean and notch filtered at 60 and 120 Hz in subsequent processing.

Trial Inclusion Criteria

Trials were excluded from use in the EEG analyses if they were contaminated by movement or eye-blink artifacts, if the subject's response was incorrect, or if the response time (RT) was excessively fast (<200 ms) or excessively slow (>3 standard deviations [SDs] above the mean). Eye-blink or other movement artifacts were identified by computing a weighted running mean of the electrooculogram activity based on Net Station's eye-blink and movement detection algorithm (Electrical Geodesics 2003). Trials were removed if the running mean crossed a threshold of 40 μV within a window beginning 200 ms before the cue onset and 500 ms after the prime offset. The total number of excluded trials per subject ranged from 7 in the best case to more than 400 trials in the worst case. Subjects were excluded from the study if more than 200 trials (out of a total of 600) had to be excluded. Of the 20 subjects run in the experiment, 4 were removed according to this criterion. Overall, subjects' error rate was low (less than 5%; see Results), so our decision to remove error trials from the analysis did not greatly reduce the total number of trials.

Preprocessing of EEG Data

The contiguous block of rereferenced EEG data was broken into epochs starting 1200 ms prior to the onset of the prime and extending 2000 ms after prime onset. This window included a 1000-ms buffer at the beginning and end of each trial to avoid edge effects when performing the spectral decomposition. The spectral decomposition was performed using a set of 49 Morlet wavelets that were logarithmically spaced from 2 to 128 Hz. Each wavelet was built to have a width that was 6 times the period of its center frequency. After performing the decomposition, the 1000-ms buffers were dropped from either end of each epoch, leaving us with a 1200-ms window, starting 200 ms prior to stimulus onset and extending 1000 ms after stimulus onset. The magnitude of each complex coefficient (i.e., the power) was then computed and downsampled to 50 Hz by computing the mean power of each component for every 20-ms bin within the 1200-ms recording window. Thus, the spectral decomposition transformed each of the 60 time bins of a trial into the power values of 49 frequency bands for each of the 77 electrodes. Each unique combination of frequency/electrode/time bin was z scored, such that (across trials) the feature's mean was zero and the feature's SD was one. These spectral features were used as inputs to the EEG pattern classifiers.

Overview of EEG Pattern Classification Analysis

The goal of our EEG pattern classification analysis was to measure processing of the prime distractor on each trial. The EEG analysis was composed of the following steps: First, we trained a set of classifiers to detect the patterns of spectral features associated with processing each category as the target (attended) stimulus. Next, we tested the classifiers' ability to detect processing of both target and distractor stimuli that had not been presented at training. For example, we used the classifier trained to detect shoes as targets to measure processing when the distractor category was shoe. As an initial index of classifier performance, we computed classifier sensitivity (i.e., how well classifier output discriminates between trials where the category was on-screen vs. when it was totally absent). Finally, for our main analysis, we used the classifier's trial-by-trial readout of distractor processing to predict behavioral negative priming effects. As is typical for classification analyses, the classification procedure was run within individual subjects (i.e., the classifier was trained on one subject's data and then applied to other data from that subject). These steps of the pattern classification analysis are described in more detail below. Figure 3 provides an overview of our EEG analysis procedures.

Training Category-Specific Classifiers

Classifier training was implemented using the ridge regression algorithm (Hoerl and Kennard 1970; Hastie et al. 2001). The ridge regression algorithm learns a linear mapping between a set of input features and an outcome variable. Like standard multiple linear regression, the ridge regression algorithm adjusts feature weights to minimize the squared error between the predicted label and the correct label. Unlike standard multiple linear regression, ridge regression also includes an L2 regularization term that biases it to find a solution that minimizes the sum of the squared feature weights. Ridge regression uses a parameter (λ) that determines the impact of the regularization term. In our analysis, λ was adaptively set to be equal to 5% of the number of input features for each classifier. Ridge regression was implemented using the Matlab Multi-Voxel Pattern Analysis toolbox (available for download at www.pni.csbmb.princeton.edu/mvpa). We chose ridge regression instead of logistic regression (an algorithm commonly used for 2-category classification) because we needed a method that would allow us to distinguish moderate levels of stimulus processing (where we predict negative priming) from lower and higher levels of stimulus processing (where we do not predict negative priming). Logistic regression transforms the weighted sum of classifier inputs using a nonlinear link function, effectively clustering the outputs of the classifier around zero and one; we were concerned that this bimodal property of logistic regression would impede our ability to distinguish between low, moderate, and high levels of stimulus processing. Ridge regression does not have this nonlinearity, and thus, it should do a better job of tracking intermediate levels of stimulus processing.

A separate ridge regression classifier was trained for each combination of stimulus category (face, house, shoe, and chair) and time bin (all 20-ms time bins within the 1200-ms trial window). Each classifier was trained to discriminate between trials where the category of interest was on-screen as the prime target (these trials were labeled with a one) and those where the category of interest was not on-screen at all (these trials were labeled with a zero). For example, one classifier was trained to detect shoe activity within the 40-ms poststimulus-onset time bin. After being trained in this fashion, the classifier can be used to generate a scalar estimate of how much the subject is processing the category of interest for a particular time bin in a particular trial: Smaller values (close to zero) indicate less processing, while greater values (close to one) indicate more processing.

Feature Selection

To reduce the number of features being fed into the classifier, we used a feature selection procedure, whereby we discarded features that (individually) did not discriminate between the 2 conditions of interest. This feature selection procedure was performed separately for each classifier that we built. A nonparametric t statistic was used to decide whether to keep each feature (defined as a particular frequency/electrode/time bin pairing). The P value for each feature was computed by comparing the output of a t-test run on the trials with their proper condition labels to the distribution of outputs generated when the labels were randomly shuffled 200 times. Features that discriminated between our conditions of interest at the $P < 0.01$ level were included in the classification process. We used a fixed statistical threshold across time bins and subjects, rather than using a fixed number of features, in order to better match (across time bins and subjects) the signal-to-noise quality of the data going into the pattern classification. Note that our feature selection procedure was restricted to the data points that

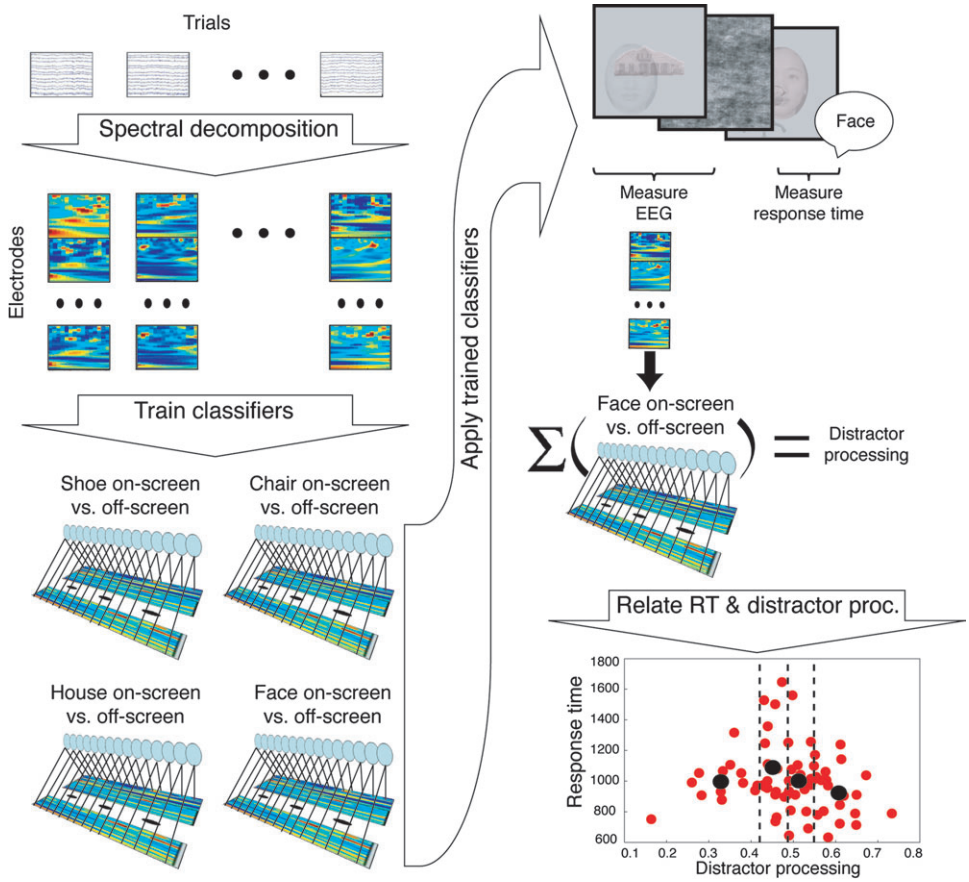

Figure 3. Overview of the EEG data analysis procedure. This procedure was run separately for each individual subject. The spectral power of the rereferenced EEG around each trial was computed and downsampled to 50 samples per second. The spectral features from all of the electrodes were concatenated and used as inputs to pattern classifiers. The classifiers were trained using ridge regression to recognize when each stimulus category was being processed as the target image; a separate classifier was trained for each combination of stimulus category and time bin. Next, the trained classifiers were used to measure how strongly the prime distractor image was processed on each trial. Finally, subjects' RTs (to the probe image) on individual trials were aligned to the classifier output from the respective trials. To assess the relationship between probe image RTs and prime image distractor processing, trials were sorted into quartiles based on the processing of the prime distractor (as measured by the classifier). We then computed the mean RT to the probe for each of the distractor-processing quartiles.

were used for classifier training; the feature selection procedure never used data points that were (subsequently) used for testing the performance of the trained classifier.

Testing Classifier Sensitivity to the Target Category

The primary goal of our EEG pattern classification analysis was to measure processing of prime distractor images and then relate this measure to subjects' behavior. However, before doing this, we evaluated the classifiers' ability to detect processing of the target and distractor images; the classifiers needed to show above-chance sensitivity to category-specific processing in order to be useful in predicting behavior. In this section, we describe how we evaluated the classifiers' sensitivity to processing of target stimuli; in the next section, we describe how we evaluated the classifiers' sensitivity to processing of distractor stimuli.

To evaluate the classifiers' sensitivity to processing of the target stimuli, we used a 10-fold cross-validation procedure consisting of the following steps: First, we identified the eligible trials for the classifier type of interest. For example, when training a face classifier, we found all of the trials where faces were on-screen-as-target and all of the trials where faces were not on-screen. Classifier training was limited to trials with superimposed target and distractor stimuli (i.e., target-only trials were never used for classifier training). Next, we randomly divided the eligible trials into 10 sets. Nine out of the 10 sets were selected to comprise the training set, and the remaining set was put aside to be used as the testing set. Before training the classifier (on the 9 sets), we made sure that equal numbers of trials were present from each condition in the training set (i.e., that there were equal numbers of category-on-screen-as-target and category-not-on-screen trials). Extra trials were eliminated in a pseudorandom fashion from the condition with more trials. We then ran feature selection on the training set (as described above) to determine which features would be used to perform the mapping. The classifier was trained on data from the training set and applied to trials from the testing set (i.e., one-tenth of the data not used at training). The classifier was also tested on the target-only trials (which, as mentioned above, were never used for training). As is standard practice in *n*-fold cross-validation, we cycled through the preceding steps 10 times, each time selecting a different set to "leave out" and use as the testing set. To ensure that our results were not dependent on a particular way of dividing up the trials, we repeated the entire 10-fold cross-validation procedure 10 times, each time using a different random division of the data into tenths.

For each trial, all of the classifier estimates for a given time bin were averaged together into a single estimate. The outputs were then smoothed by replacing each value with the mean of the surrounding 5 time bins (i.e., the mean of the surrounding 100-ms window). To measure the classifier's ability to decode the category of the target stimulus, we used area under the receiver operating characteristic curve (AUC) to quantify how well classifier output discriminated between trials that had the category on-screen as a target and those that did not have the category on-screen (Fawcett 2006).

Measuring Distractor Processing

To measure distractor processing, we took all of the classifiers that were trained using the 10-fold cross-validation procedure described above, and we applied these classifiers to trials where the category of interest was on-screen as the prime distractor. For example, the classifier trained to detect processing of faces was applied to the trials in which faces were presented as the distractor image. Note that the category-specific classifiers were trained on trials where the category was present as a target or totally absent but never on those where the category was present as a distractor; as such, for a given classifier, there was no overlap between the trials used at training and those used to measure distractor processing. As with the target sensitivity analysis described in the previous section, we averaged all of the estimates for each time bin and then temporally smoothed the estimates. We then used AUC to quantify how well classifier output discriminated between trials that had the category on-screen as a distractor and those that did not have the category on-screen.

Combining Classifier Estimates across Time Bins

The methods described above yield a classifier estimate for each 20-ms time bin within a trial. For our analyses relating classifier output to behavior, we wanted estimates that summarize, for each trial (across time bins), the degree of processing for a particular category. To obtain a single estimate of the level of category-specific processing for each trial, we computed the mean output of the 48 time bin classifiers beginning with the first classifier poststimulus onset and ending with the last classifier (i.e., the classifiers from bins 20- to 960-ms poststimulus onset).

Relating Classifier Estimates to Behavior

Our primary analysis involved testing whether the magnitude of the negative priming effect varied as a nonmonotonic function of distractor processing, as predicted by the nonmonotonic plasticity hypothesis. To measure the relationship between distractor processing and priming, we split the trials into quartiles based on the estimated level of distractor processing (i.e., the magnitude of the output of the appropriate classifier) on each trial. The quartile boundaries were computed separately for each subject by pooling the trials from the control and ignored repetition conditions. These quartile boundaries were then applied to each of the 2 conditions. We computed the priming effect for each quartile by subtracting the mean RT of ignored repetition trials in that quartile from that of control condition trials in that quartile. We also ran a variant of the analysis where we divided up trials into quartiles based on the degree of prime target processing (instead of distractor processing). For both the distractor-processing and the target-processing quartile analyses, we used a repeated measures one-way analysis of variance (ANOVA) to test whether the size of the priming effect varied significantly across quartiles and performed post hoc tests between individual quartile effects using paired Student's *t*-tests. Also, to evaluate the reliability of the predicted nonmonotonic pattern (correcting for multiple comparisons), we used a nonparametric permutation test (see Results for details). The significance of each test was evaluated using 2-tailed distributions with $n = 16$ and $\alpha = 0.05$. All data were confirmed to approximate a normal distribution using a Kolmogorov–Smirnov test prior to performing any statistics.

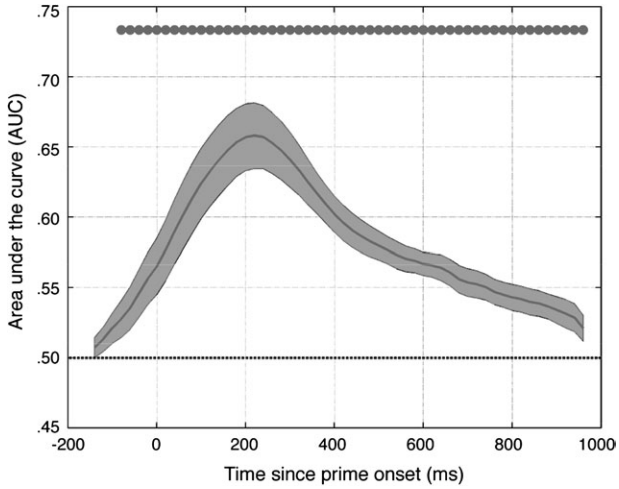

Figure 4. Average sensitivity of the classifier to the category of the prime target stimulus, combining across the 4 categories. Sensitivity was computed in a cross-validated fashion (i.e., different sets of trials were used for classifier training and testing). The solid black line plots the AUC computed at each time bin. The shaded region around this line indicates the standard error across subjects. The dotted black line along 0.5 marks chance performance. The dots along the top of the figure indicate which of the time bin–specific classifiers performed significantly above chance at the $P < 0.05$ level.

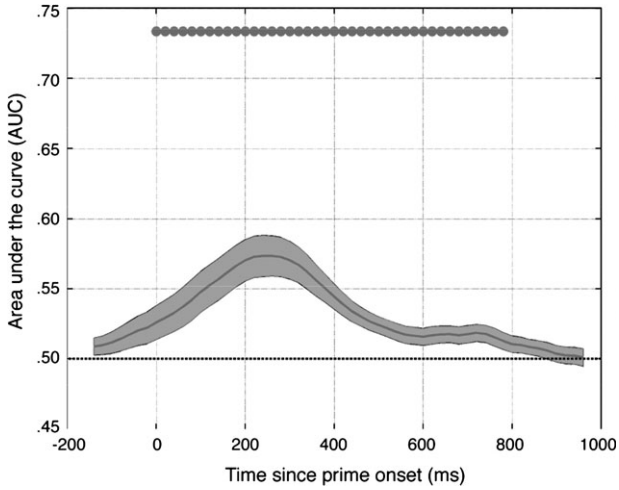

Figure 5. Average sensitivity of the classifier to the category of the prime distractor stimulus, combining across the 4 categories. The classifier was trained on trials where the category was on-screen as the "target" stimulus and then tested on (other) trials where the category was on-screen as the "distractor" stimulus. The solid black line plots the AUC computed at each time bin. The shaded region around this line indicates the standard error across subjects. The dotted black line along 0.5 marks chance performance. The dots along the top of the figure indicate which of the time bin–specific classifiers performed significantly above chance at the $P < 0.05$ level.

Results

Classifier Sensitivity Analysis

The results of the classifier sensitivity analysis are shown in Figures 4 and 5. Sensitivity was indexed using AUC, where chance = 0.5. Figure 4 shows the classifier's sensitivity to the category of the prime target stimulus, as a function of time-elapsed poststimulus onset. When we averaged classifier output across time bins, the classifier's sensitivity to the target category was 0.64 (standard error of mean [SEM] = 0.02), which was significantly above chance, $t_{15} = 8.45$, $P < 0.00001$. Figure 5 shows the classifier's sensitivity to the category of the prime distractor stimulus, as a function of time-elapsed poststimulus onset. When we averaged classifier output across time bins, the classifier's sensitivity to the target category was 0.55 (SEM = 0.01), $t_{15} = 4.82$, $P < 0.001$. Overall, these results show that the classifier was significantly above chance at detecting the category of both target and distractor stimuli.

The finding that classifier sensitivity was higher for targets than for distractors was expected; insofar as targets are

processed more strongly than distractors, we would expect target processing to be more detectable than distractor processing. The most important point here is that distractor sensitivity was above floor, which licenses us to proceed with analyses relating distractor processing (as measured by the classifier) to behavior. In the Supplementary Material, we present additional analyses comparing classifier sensitivity for targets and distractors; we also show that classifier sensitivity was significantly above chance for all 4 categories (for both targets and distractors), and we present classifier "importance maps" showing which features were used by the classifiers to detect each of the 4 categories.

The finding of above-chance classification prior to stimulus onset in Figures 4 and 5 is an artifact of our spectral decomposition procedure, which used 6-cycle wavelets. Six-cycle wavelets have better frequency resolution than shorter wavelets, but they induce a higher degree of temporal smearing. The finding of above-chance classification prior to stimulus onset goes away when we use 3-cycle wavelets.

Negative Priming Results: Behavior

Consistent with previous negative priming studies (e.g., Tipper 1985; Fox 1995), we found a significant negative priming effect for reaction times: Subjects were 15 ms slower overall to name images in the ignored repetition condition (M = 922 ms) compared with the control condition (M = 907 ms), SEM of the ignored repetition–control difference = 5.5 ms, t_{15} = 2.71, P < 0.05. Response accuracy (indexed in terms of percent correct) was effectively at ceiling in both the ignored repetition condition (M = 98.4) and the control condition (M = 97.8) and did not differ significantly across these conditions, SEM of the ignored repetition–control difference = 0.7, t_{15} = –0.90, n.s.

Negative Priming Results: Relating Distractor Processing to Behavior

The key question that we wanted to address was how the size of the negative priming effect varied as a function of prime distractor processing (measured via the EEG pattern classification analysis). Consistent with the nonmonotonic plasticity hypothesis, we found that the priming effect varied nonmonotonically as a function of prime distractor processing. The results of our quartile analysis (where we split trials into quartiles based on the level of prime distractor processing) are shown in Figure 6. A repeated measures ANOVA showed that the magnitude of the priming effect was significantly different across distractor-processing quartiles ($F_{3,45}$ = 4.21, P < 0.05). The priming effects within each quartile were as follows: The low distractor-processing quartile showed a nonsignificant –15 ms priming effect (t_{15} = 1.37, n.s.); the medium-low distractor-processing quartile showed a significant –51 ms priming effect (t_{15} = 4.87, P < 0.001); the medium-high distractor-processing quartile showed a nonsignificant –10 ms priming effect (t_{15} = 1.01, n.s.), and the high distractor-processing quartile showed a nonsignificant 11 ms priming effect (t_{15} = 0.69, n.s.). Pairwise comparisons between the quartile priming effects reveal a nonmonotonic pattern: The priming effect in the medium-low distractor-processing quartile was significantly more negative than those in the other 3 quartiles. The statistics for the pairwise comparisons are reported in the Supplementary Material; we also present the

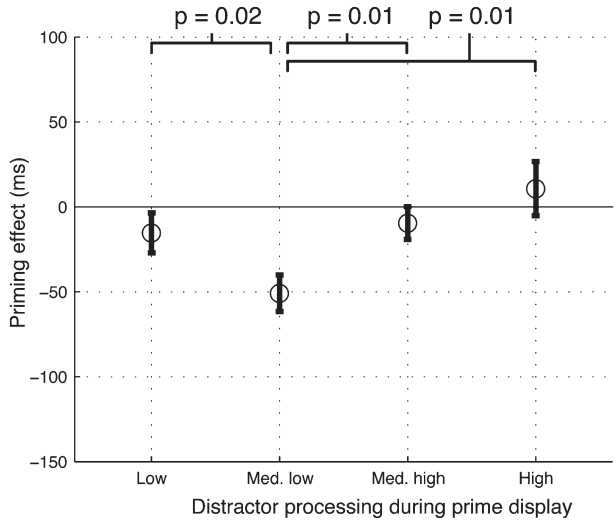

Figure 6. Comparison of priming effects as a function of prime distractor processing. Trials were split into quartiles based on the level of prime distractor processing, and then, priming effects were computed separately for each quartile. The priming effect in the medium-low distractor-processing quartile was significantly more negative than that in the other 3 quartiles. Significance values reflect the reliability of the difference across subjects. Error bars indicate standard errors (across subjects) on the mean priming effect within each quartile.

results of the analysis split by prime distractor category (face, house, shoe, and chair) in the Supplementary Material.

Next, we used a nonparametric test to assess the probability of obtaining this nonmonotonic pattern due to chance (correcting for multiple comparisons). Going into the experiment, we had predicted that the negative priming effect would be largest given moderate processing compared with high and low processing, but we did not have a prediction regarding which of the moderate-processing quartiles (i.e., medium-low or medium-high) would show the largest negative priming effect. Thus, there were 2 patterns of quartile priming effects that we would have regarded as equally consistent with our hypothesis (pattern 1: significantly more negative priming in the medium-high quartile than in the high and low quartiles, P < 0.05 for each pairwise comparison; pattern 2: significantly more negative priming in the medium-low quartile than in the high and low quartiles, P < 0.05 for each pairwise comparison). For our nonparametric test, we assessed the likelihood of obtaining either of these patterns by chance, by shuffling each subject's RT data with respect to the classifier output data (note that the shuffle was done within condition, so each subject's mean control RT and ignored repetition RT were unchanged); this shuffle instantiates the null hypothesis that prime distractor processing has no impact on probe RTs. We shuffled the RT data 10 000 times. For each of the 10 000 shuffled versions of the data, we recomputed the quartile priming scores and measured whether there were statistically significant differences between quartiles. We found that the probability of obtaining a nonmonotonic effect in the shuffled data (i.e., either pattern 1 or pattern 2 listed above) was less than 0.005.

As mentioned in the Trial Inclusion Criteria part of the Materials and Methods, our EEG analysis excluded trials where subjects responded incorrectly. The error rate in this study was very close to zero, and it did not differ significantly across the ignored repetition and control conditions, so we did not

expect this exclusion to have a large effect. Nonetheless, to ensure that this exclusion was not influencing our results, we subsequently ran a version of the analysis where error trials were included. The results of this analysis are reported in the Supplementary Material. Including these error trials did not change any of our key results. Most importantly, the priming effect for the medium-low quartile was still significantly more negative than those for other quartiles. Also, we found in this analysis that error rates did not vary across conditions (control vs. ignored repetition) within any of the quartiles, and the effect of condition on error rates did not vary across quartiles.

Negative Priming Results: Role of the Control Condition

In the analyses reported above, we evaluated the nonmonotonic plasticity hypothesis by subtracting ignored repetition RTs from control RTs within each distractor-processing quartile. The control condition and the ignored repetition condition were matched in every respect except for the fact that the prime distractor was repeated as the probe target in the ignored repetition condition (but not in the control condition). Thus, the effect of subtracting ignored repetition RTs from control RTs was to isolate the effect of stimulus repetition on RTs, unconfounded by other factors that might be shared across the 2 conditions.

In this section, we report the results of quartile analyses conducted separately on control RTs and ignored repetition RTs; these analyses allow us to see whether there were effects of distractor processing on RT that were hidden by the control versus ignored repetition subtraction. We had expected that the ignored repetition condition would show a nonmonotonic effect of prime distractor processing on probe RTs (akin to the pattern shown in Fig. 6) and that the control RTs would be relatively flat as a function of distractor processing. However, this is not the pattern that we found. The mean quartile RTs in the ignored repetition condition (considered on its own) did show the same nonmonotonic pattern as the mean quartile priming effects, but the effect was weak and did not reach significance: A repeated measures ANOVA did not find that RTs varied significantly over levels of distractor processing ($F_{3,45}$ = 1.50, n.s.), and the pairwise comparisons of these quartile effects did not cross the $P < 0.05$ significance threshold. These mean quartile RTs are shown in panel *A* of Figure 7. The pattern of mean RTs across quartiles in the control condition was the opposite of the pattern observed in the ignored repetition condition. That is, RTs associated with medium-low levels of prime distractor processing were the fastest, whereas those associated with high levels of prime distractor processing were slowest. Again, a repeated measures ANOVA did not find that RTs varied significantly over levels of distractor processing ($F_{3,45}$ = 2.74, n.s.). However, statistical pairwise comparisons showed that mean RTs in the medium-low distractor-processing quartile were significantly faster than those in the low distractor-processing quartile and the high distractor-processing quartile. These results are shown in panel *B* of Figure 7. The statistics for the pairwise comparisons from both conditions are reported in the Supplementary Material.

The unexpected presence of quartile differences in the control condition (and the lack of robust quartile differences in the ignored repetition condition) highlights the importance of including this control when evaluating nonmonotonic plasticity; the predicted pattern only emerges when we compute the difference between control RTs and ignored repetition RTs. These results also pose an interesting question: Why were there quartile differences in the control condition? In the Discussion, we explain these differences in terms of a location-based negative priming effect that affected processing in both the control condition and the ignored repetition condition.

Relating Target Processing to Behavior

Finally, we ran a version of the analysis where we split trials into quartiles based on how strongly the target was processed during the prime (instead of the distractor). The results of this analysis are shown in Figure 8. There were no significant differences in priming across quartiles, and the quartile means did not show the nonmonotonic pattern (see the Supplementary Material for detailed statistics). This suggests that the nonmonotonic pattern shown in Figure 6 is specifically attributable to differences in distractor processing, as opposed to more general factors (e.g., fluctuations in subjects' attentional state) that impact both target and distractor processing.

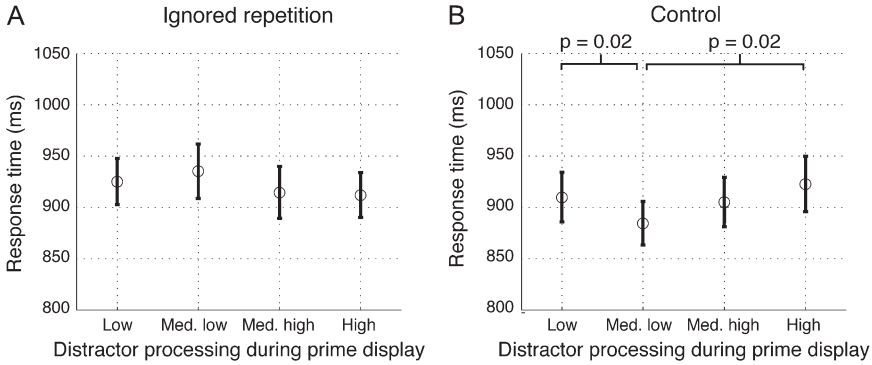

Figure 7. Comparison of RTs as a function of prime distractor processing for the ignored repetition and control conditions. Each panel shows the mean RT of each quartile of trials when sorted by distractor processing during the prime. The relationship between distractor processing and RT in the ignored repetition condition (*A*) follows the same trend as was observed between distractor processing and negative priming. The relationship between distractor processing and RT in the control condition (*B*) shows the opposite relationship (i.e., the RTs were fastest for trials with medium-low distractor processing and slowest for trials with high distractor processing). Significance values reflect the reliability of the difference across subjects, calculated using a 2-tailed paired-samples *t*-test. Error bars indicate standard errors (across subjects) on the mean RT within each quartile.

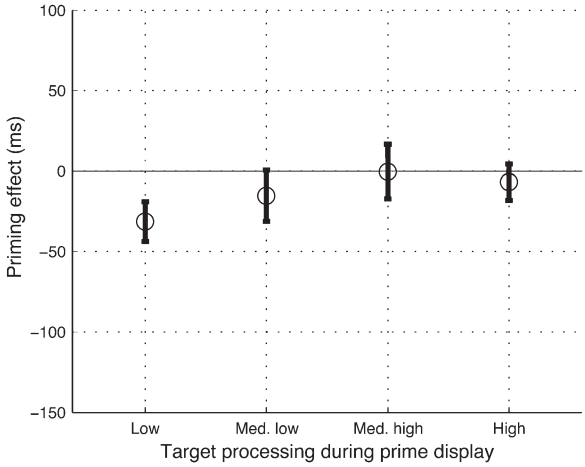

Figure 8. Comparison of priming effects as a function of prime target processing. Trials were split into quartiles based on the level of prime target processing, and then, priming effects were computed separately for each quartile. None of the pairwise comparisons of the quartile priming effects were significant. Error bars indicate standard errors (across subjects) on the mean priming effect within each quartile.

Discussion

According to the nonmonotonic plasticity hypothesis, moderate excitation of a representation should trigger weakening of that representation (Bienenstock et al. 1982; Diederich and Opper 1987; Gardner 1988; Vico and Jerez 2003; Senn and Fusi 2005; Norman et al. 2006). In keeping with this hypothesis, we found a robust negative priming effect given moderate (but not higher or lower) levels of distractor processing.

The nonmonotonic plasticity hypothesis was first proposed by Bienenstock et al. (1982) in the form of a computational model of plasticity in visual cortex. Subsequently, this hypothesis has been instantiated in several other computational theories of learning (e.g., Diederich and Opper 1987; Gardner 1988; Vico and Jerez 2003; Senn and Fusi 2005; Norman et al. 2006), and slice electrophysiology studies have provided evidence in support of the nonmonotonic plasticity hypothesis at the level of single synapses (e.g., Artola et al. 1990; Hansel et al. 1996). The current study provides the first evidence that the nonmonotonic plasticity hypothesis scales up to the level of distributed representations (i.e., the summed activity over millions of neurons as measured by scalp EEG) and has a measurable effect on behavioral performance, as predicted by Norman et al. (2006). These results link the conditions known to induce synaptic long-term depression in rodents (Artola et al. 1990; Hansel et al. 1996) to diminished accessibility of perceptual representations in humans. Additional investigations will be required to test if the latter effect is dependent on the former effect.

It is worth noting that it is not possible, given our current results, to define the absolute level of processing at which weakening begins or at which strengthening takes over. While slice electrophysiology studies have documented the absolute levels of membrane potential at which weakening (e.g., Artola et al. 1990; Hansel et al. 1996) occurs, it is difficult to obtain absolute measures of stimulus processing from the scalp. Our results serve as an existence proof that memory strength can vary in a nonmonotonic fashion as a function of stimulus processing; more work is needed to identify absolute markers of the level of processing that triggers weakening.

Explaining Results from the Control Condition

Going into the experiment, we had not expected to find an effect of prime distractor processing on probe RTs in the control condition. We were therefore surprised to find a nonmonotonic effect of distractor processing on control RTs, whereby RTs in the medium-low distractor-processing quartile were significantly faster than those in the low distractor-processing quartile and the high distractor-processing quartile.

The pattern of RT results in the control condition can be explained by a location-based negative priming effect: Several studies have found that subjects are slower to respond to an item when it is placed where a previously ignored stimulus had been located (Tipper et al. 1990; Connelly and Hasher 1993; Milliken et al. 1994). In both the control condition and the ignored repetition condition, the prime distractor and the probe distractor appeared in the same location. As such, any suppression that accrues to the location of the prime distractor will transfer to the (same-location) probe distractor; this decrease in processing of the probe distractor should result in faster processing of the probe target. RTs in the ignored repetition condition are therefore subject to 2 distinct negative priming effects: location-based negative priming (which leads to faster probe RTs) and item-based negative priming (which leads to slower probe RTs). The control condition is subject to the former location-based effect (since the distractor location is the same in the prime and the probe) but not to the latter item-based effect (since the prime distractor itself is not repeated). Insofar as ignored repetition RTs reflect location-based priming plus item-based priming, and control RTs reflect location-based priming (alone), this implies that computing the difference between control RTs and ignored repetition RTs will recover the item-based priming effect, which shows the predicted nonmonotonic shape.

One feature of the above account requires further explanation: If RTs in the control condition were being modulated by location-based priming, why did these RTs vary nonmonotonically with the classifier's readout of distractor identity processing? In this paradigm, we expect that location and object identity processing will be highly correlated: Processing of the distractor's identity is likely to also involve processing of its location; as such, the classifier's readout of distractor identity processing can be used as a proxy for distractor location processing. Furthermore, we hypothesize that nonmonotonic plasticity is a general principle that should apply to location representations in addition to object identity representations. These claims imply that we should observe a nonmonotonic relationship between our proxy measure of location processing (i.e., classifier output) and RTs in the control condition, which is what we found. This account also helps to explain why RTs were relatively flat across quartiles in the ignored repetition condition: If identity and location processing are highly correlated, then levels of distractor identity processing that lead to an item-based negative priming effect should be accompanied by levels of distractor location processing that lead to a negative location-based priming effect. As described above, these 2 effects push RTs in opposite directions on ignored repetition trials and thus should (approximately) cancel each other out. These ideas could be tested in future work by orthogonally manipulating whether the identity of the distractor is repeated and whether the location of the distractor is repeated; this would make it possible to

separately assess how location- and identity-based priming effects vary as a function of prime distractor processing.

Relationship to Other Work on Negative Priming

The link between negative priming and nonmonotonic plasticity was first made in a conference presentation by Gotts and Plaut (2005). Their approach to testing these predictions was to manipulate the relative brightness of targets and distractors (as a means of manipulating distractor processing). Using this purely behavioral approach, they found that increasing distractor processing led to a transition between negative and positive priming, but they were not able to trace out the full nonmonotonic curve. One limitation of the approach taken by Gotts and Plaut (2005) is that distractor processing can vary within conditions defined by particular stimulus brightness values; subjects might succeed at completely ignoring a bright distractor, and they might clearly perceive a dim distractor. This within-condition variability makes it harder to detect the predicted nonmonotonic effect as a function of stimulus brightness. A key benefit of our approach is that it allows us to measure—and thus account for—within-condition variability in distractor processing (see Benefits of Pattern Classification for additional discussion of this point).

Our preferred interpretation of our results is that moderate distractor processing weakened the distractor's representation, resulting in slower RTs when the subject (subsequently) had to name the item's category. However, there are other theories of negative priming (besides nonmonotonic plasticity) that can explain the nonmonotonic pattern of results observed here. For example, the temporal discrimination theory set forth by Milliken et al. (1998) predicts that reaction times to moderately processed primes will be slow because these primes were processed well enough to be familiar (thereby inducing the perceptual system to try to retrieve prior instances of the item), but the level of processing was not high enough to support successful recollection of the prior instance; the time wasted with this failed retrieval attempt leads to slow RTs. Within the domain of perceptual priming, the predictions of this theory mimic those of the nonmonotonic plasticity hypothesis. The key difference between the Milliken et al. (1998) account of the nonmonotonicity and the nonmonotonic plasticity hypothesis is that the latter is a domain-general learning principle—According to the nonmonotonic plasticity hypothesis, nonmonotonic learning effects should be observed across multiple memory paradigms and dependent measures, ranging from long-term cued-recall accuracy to perceptual task RTs. By contrast, the nonmonotonic relationship predicted by Milliken et al. (1998) applies to a much narrower range of situations (RTs to repeated stimuli).

Regardless of which explanation is correct, our finding that priming effects are nonmonotonically related to (initial) prime processing may help to explain numerous puzzling results in the negative priming literature. For example, manipulations of target–competitor perceptual grouping (i.e., whether or not the target and competitor are perceived as a single object) have been shown to magnify the negative priming effect in some studies (Fox 1998) and to reverse it in others (Fuentes et al. 1998). A nonmonotonic relationship between distractor processing and priming could account for these apparently contradictory results: If distractor processing starts out low, then increasing the degree of distractor processing could push it from the low-excitation (no-learning) region of the plasticity curve to the moderate-excitation (weakening) region, thereby boosting the negative priming effect. However, if distractor processing starts out at a moderate level, then increasing distractor processing could push the distractor from the moderate-excitation (weakening) region to the high-excitation (strengthening) region, thereby reversing the negative priming effect.

Negative Priming as a Measure of Learning

Given that the nonmonotonic plasticity hypothesis should apply across multiple learning domains, why did we choose the negative priming paradigm to test the hypothesis? The main benefit of the negative priming paradigm is that the average level of competition between the target and the distractor can be adjusted in a straightforward fashion (e.g., by manipulating stimulus visibility). However, one drawback of using negative priming is that there is some ambiguity about the degree to which learning (i.e., lasting synaptic adjustment) is responsible for observed priming effects. While the negative priming effect has been shown to persist for as long as a month (DeSchepper and Treisman 1996), suggesting that the slowed RTs reflect a learning-based mechanism, this does not mean that all negative priming findings reflect lasting synaptic modification. Given the short time lag between the prime and probe in our study, it is possible (in principle) that priming effects in our study could be due to short-term carryover of activation (or habituation) from the prime rather than lasting adjustment of synapses.

However, in practice, it is difficult to see how short-term activation/habituation effects could account for the particular pattern of results observed here. In recent work using a short-term repetition priming paradigm (with zero lag between the prime and the probe), Huber and others have demonstrated that short prime durations facilitate probe processing and longer prime durations inhibit probe processing (possibly due to neural habituation effects; for a review, see Huber and O'Reilly 2003). This pattern of results is the exact opposite of the pattern observed here: In our study, the function relating prime processing to the size of the priming effect (Figs 1 and 6) initially dips below zero (negative priming) and then rises nonsignificantly above zero (positive priming). By contrast, in the short-term priming studies reviewed by Huber and O'Reilly, the function relating prime duration to the size of the priming effect initially rises above zero (positive priming) and then dips below zero (negative priming). This difference suggests that our results and the Huber and O'Reilly results are attributable to different mechanisms.

Extensions to Other Paradigms

To further evaluate the nonmonotonic plasticity hypothesis, we plan to explore nonmonotonic plasticity effects using long-term memory paradigms, where it is unambiguous that learning effects are driven by lasting synaptic changes. One potentially relevant paradigm is the think–no think paradigm (Anderson and Green 2001; Anderson et al. 2004). In this paradigm, subjects are given a cue that was previously associated with another item, and they are instructed to avoid retrieving the memory associated with that cue; the basic finding is that trying not to retrieve a memory (when it is strongly cued)

makes that memory more difficult to retrieve later. As with negative priming, the literature is mixed: While numerous studies have observed a forgetting effect, others have not (Bulevich et al. 2006). According to the logic of the present study, it may be possible to explain both successes and failures in terms of variability in the excitation of the to-be-forgotten representation: Successful forgetting (weakening) should occur with moderate excitation and forgetting should be unsuccessful with higher and lower levels of excitation.

Benefits of Pattern Classification

More generally, the work described here adds to a growing body of literature that leverages advances in machine learning to explore the neural mechanisms of cognition (e.g., Polyn et al. 2005; Philiastides et al. 2006; McDuff et al. 2009). Pattern classification methods make it possible to covertly track otherwise-uncontrolled variance in latent cognitive states, which we can then relate to behavior. Other studies have used neural data to measure differences in the average level of distractor processing across conditions (Yi et al. 2004; Vogel et al. 2005). As described above, a key advantage of our pattern classification approach is that it allows us to derive a trial-by-trial readout of prime distractor processing within a particular condition, which we can then relate to trial-by-trial variance in probe reaction times. In this study, our ability to measure the level of distractor processing on a trial-by-trial basis allowed us to isolate a 51-ms negative priming effect (in the medium-low distractor-processing quartile), whereas the basic negative priming effect (computed across all trials) was less than one-third of that magnitude (15 ms). The present results suggest that measuring neural processing with classifiers may help to resolve fundamental questions regarding how neural dynamics shape learning.

It is worth noting that the pattern classification approach carries a cost both in terms of methodological transparency and in required computation time. Yet, we believe that these costs are negated with the availability of freely downloadable pattern classification toolboxes and the benefits that accompany this approach. The main benefit of the pattern classification approach is that it makes full use of the information in the data. In contrast, a more standard event-related potential–based approach, in which the amplitude of an evoked response is used as the estimate of the level of processing, may not. For example, the extent to which a face was processed could be estimated based on the amplitude of the evoked N170. However, the N170 is clearly not the only type of neural activity that is informative regarding face processing. The pattern classification method involves an explicit feature discovery stage (feature selection) that makes it possible to discover other types of informative activity (in addition to components related to the N170) and then leverage them to estimate the level of face processing. This property of the pattern classification approach makes it particularly powerful for estimating the level of processing of stimuli for which well-defined evoked responses have not been well characterized (e.g., chairs or shoes).

Conclusions

In summary, the present study used an EEG pattern classification analysis to characterize the relationship between the level to which a stimulus was processed and its subsequent accessibility. Consistent with the nonmonotonic plasticity hypothesis, we found that the negative priming effect was largest, given moderate (as opposed to higher or lower) levels of stimulus processing. In addition to providing an existence proof for this nonmonotonic relationship, these results link the conditions known to generate synaptic weakening to diminished accessibility of perceptual representations in humans.

Funding

National Institutes of Health (R01MH069456 to K.A.N. and MH077469 to E.L.N.).

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Notes

We thank Greg Detre, Per Sederberg, Chris Moore, Joel Quamme, and Francisco Pereira for their advice and assistance regarding data analysis; Jim Haxby and Ida Gobbini for their help with the design of the study and for providing data sets for the initial tests of our methods; Andrew Saxe for elucidating links between learning algorithms; Andrew Conway for helping to orient us to the world of negative priming; and Rafael Escobedo for his help in the collection of the EEG data. *Conflict of Interest*: None declared.

References

Amit DJ, Gutfreund H, Sompolinsky H. 1985. Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys Rev Lett. 55:1530–1533.

Amit DJ, Gutfreund H, Sompolinsky H. 1987. Statistical mechanics of neural networks near saturation. Ann Phys (NY). 173:30–67.

Anderson MC, Green C. 2001. Suppressing unwanted memories by executive control. Nature. 410:366–369.

Anderson MC, Ochsner KN, Kuhl B, Cooper J, Robertson E, Gabrieli SW, Glover GH, Gabrieli JD. 2004. Neural systems underlying the suppression of unwanted memories. Science. 303:232–235.

Artola A, Bröcher S, Singer W. 1990. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature. 347:69–72.

Bienenstock EL, Cooper LN, Munro PW. 1982. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J Neurosci. 2:32–48.

Bulevich JB, Roediger HL, Balota DA, Butler AC. 2006. Failures to find suppression of episodic memories in the think/no-think paradigm. Mem Cognit. 34:1569–1577.

Connelly SL, Hasher L. 1993. Aging and the inhibition of spatial location. J Exp Psychol Hum Percept Perform. 19:1238–1250.

Cooper LN, Intrator N, Blais BS, Shouval HZ. 2004. Theory of cortical plasticity. Hackensack (NJ): World Scientific Publishing.

DeSchepper BG, Treisman A. 1996. Visual memory for novel shapes: implicit coding without attention. J Exp Psychol Learn Mem Cogn. 22:27–47.

Diederich S, Opper M. 1987. Learning of correlated patterns in spin-glass networks by local learning rules. Phys Rev Lett. 58:949–952.

Electrical Geodesics. 2003. Net station waveform tools technical manual. Eugene (OR): Electrical Geodesics, Inc. Technical report No. S-MAN-200-WTFR-001.

Fawcett T. 2006. An introduction to ROC analysis. Pattern Recognit Lett. 27:861–874.

Fox E. 1995. Negative priming from ignored distractors in visual selection: a review. Psychon Bull Rev. 2:145–173.

Fox E. 1998. Perceptual grouping and visual selective attention. Percept Psychophys. 60:1004–1021.

Fuentes LJ, Humphreys GW, Agis IF, Encarna C, Catena A. 1998. Object-based perceptual grouping affects negative priming. J Exp Psychol Hum Percept Perform. 24:664–672.

Gardner E. 1988. The space of interactions in neural network models. J Phys A: Math Gen. 21:257–270.

Gotts SJ, Plaut DC. 2005. Neural mechanisms underlying positive and negative repetition priming [abstract]. In: Cognitive Neuroscience Society 12th annual meeting; 2005 Apr 15–19; New York. Davis (CA): Cognitive Neuroscience Society.
Hansel C, Artola A, Singer W. 1996. Different threshold levels of postsynaptic [ca2+]i have to be reached to induce LTP and LTD in neocortical pyramidal cells. J Physiol Paris. 90:317–319.
Hastie T, Tibshirani R, Friedman J. 2001. The elements of statistical learning. New York: Springer.
Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. Nat Rev Neurosci. 7:523–534.
Hebb DO. 1949. The organization of behavior. New York: Wiley.
Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 12:69–82.
Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A. 79:2554–2558.
Huber DE, O'Reilly RC. 2003. Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregation through synaptic depression. Cogn Sci. 27:403–430.
McDuff S, Frankel H, Norman KA. 2009. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. J Neurosci. 29: 508–516.
Milliken B, Joordens S, Merikle PM, Seiffert AE. 1998. Selective attention: a reevaluation of the implications of negative priming. Psychol Rev. 105:203–229.
Milliken B, Tipper SP, Weaver B. 1994. Negative priming in a spatial localization task: feature mismatching and distractor inhibition. J Exp Psychol Hum Percept Perform. 20:624–646.
Norman KA, Newman EL, Detre GJ. 2007. A neural network model of retrieval-induced forgetting. Psychol Rev. 114:887–953.
Norman KA, Newman EL, Detre GJ, Polyn SM. 2006. How inhibitory oscillations can train neural networks and punish competitors. Neural Comput. 18:1577–1610.
Philiastides MG, Ratcliff R, Sajda P. 2006. Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. J Neurosci. 26:8965–8975.
Philiastides MG, Sajda P. 2006. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex. 16:509–518.
Polyn SM, Natu VS, Cohen JD, Norman KA. 2005. Category-specific cortical activity precedes retrieval during memory search. Science. 310:1963–1966.
Senn W, Fusi S. 2005. Learning only when necessary: better memories of correlated patterns in networks with bounded synapses. Neural Comput. 17:2106–2138.
Tipper SP. 1985. The negative priming effect: inhibitory priming by ignored objects. Q J Exp Psychol A. 37:571–590.
Tipper SP, Brehaut JC, Driver J. 1990. Selection of moving and static objects for the control of spatially directed action. J Exp Psychol Hum Percept Perform. 16:492–504.
Vico FJ, Jerez JM. 2003. Stable neural attractors formation: learning rules and network dynamics. Neural Process Lett. 18:1–16.
Vogel EK, McCollough AW, Machizawa MG. 2005. Neural measures reveal individual differences in controlling access to working memory. Nature. 438:500–503.
Yi DJ, Woodman GF, Widders D, Marois R, Chun MM. 2004. Neural fate of ignored stimuli: dissociable effects of perceptual and working memory load. Nat Neurosci. 7:992–996.