Behavioral/Cognitive

Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain

Jiefeng Jiang,[1] **Christopher Summerfield,**[2] **and Tobias Egner**[1]

[1]Center for Cognitive Neuroscience and Department of Psychology and Neuroscience, Duke University, Durham, North Carolina 27708, and [2]Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom

Attention, the prioritization of goal-relevant stimuli, and expectation, the modulation of stimulus processing by probabilistic context, represent the two main endogenous determinants of visual cognition. Neural selectivity in visual cortex is enhanced for both attended and expected stimuli, but the functional relationship between these mechanisms is poorly understood. Here, we adjudicated between two current hypotheses of how attention relates to predictive processing, namely, that attention either enhances or filters out perceptual prediction errors (PEs), the PE-promotion model versus the PE-suppression model. We acquired fMRI data from category-selective visual regions while human subjects viewed expected and unexpected stimuli that were either attended or unattended. Then, we trained multivariate neural pattern classifiers to discriminate expected from unexpected stimuli, depending on whether these stimuli had been attended or unattended. If attention promotes PEs, then this should increase the disparity of neural patterns associated with expected and unexpected stimuli, thus enhancing the classifier's ability to distinguish between the two. In contrast, if attention suppresses PEs, then this should reduce the disparity between neural signals for expected and unexpected percepts, thus impairing classifier performance. We demonstrate that attention greatly enhances a neural pattern classifier's ability to discriminate between expected and unexpected stimuli in a region- and stimulus category-specific fashion. These findings are incompatible with the PE-suppression model, but they strongly support the PE-promotion model, whereby attention increases the precision of prediction errors. Our results clarify the relationship between attention and expectation, casting attention as a mechanism for accelerating online error correction in predicting task-relevant visual inputs.

Introduction

Visual attention increases neural selectivity for goal-relevant stimuli in category-selective regions of ventral visual cortex (Murray and Wojciulik, 2004; Yi et al., 2006; Mitchell et al., 2009), allowing the identity of attended stimuli to be more readily decoded from fMRI multivoxel patterns (Serences et al., 2009; Jehee et al., 2011; Chen et al., 2012). Attention thus appears to enhance the signal-to-noise ratio of neural population activity, facilitating readout of task-relevant visual information at downstream processing stages. Aside from relevance, the contextual probability of stimulus occurrence is also a key determinant of successful recognition (Biederman, 1972; Bar, 2004). One class of theory ("predictive coding") argues that contextual facilitation depends on reciprocal message passing between higher and lower processing stages, with recognition occurring when expected in-

formation and observed information are reconciled (Mumford, 1992; Rao and Ballard, 1999; Friston, 2005). Accordingly, visual regions would compute the prior probability of a stimulus (a prediction) and how this prediction should be revised given new sensory information (a prediction error [PE]; Friston, 2005). In support of this hypothesis, conditionally probable stimuli elicit reduced aggregate activity in sensory cortices (Näätänen et al., 1987; Garrido et al., 2008; Summerfield et al., 2008; den Ouden et al., 2009; Alink et al., 2010; Egner et al., 2010; Kok et al., 2012a) and different visual neurons code for whether anticipated and observed stimuli are matching or mismatching (Miller et al., 1993; Meyer and Olson, 2011; Keller et al., 2012).

One outstanding question, however, is how these neural signals encoding predictions and their violation (PEs) are modulated by visual attention (Summerfield and Egner, 2009). A canonical view is that attention acts as a filter, suppressing irrelevant information to focus on the most relevant signals (Broadbent, 1958). For example, visual search is facilitated if unanticipated information is suppressed (Seidl et al., 2012). Accordingly, attention might mitigate the influence of unexpected information by dampening visual PE signals (Rao and Ballard, 2005), which would obviate the reconciliation of expected and observed information, thus reducing the net disparity between neural signals for expected and unexpected percepts (the PE-suppression model). Because expected and unexpected stimuli are associated with distinct fMRI multivoxel patterns (Kok et al., 2012a; de Gardelle et al., 2013), the PE-suppression model pre-
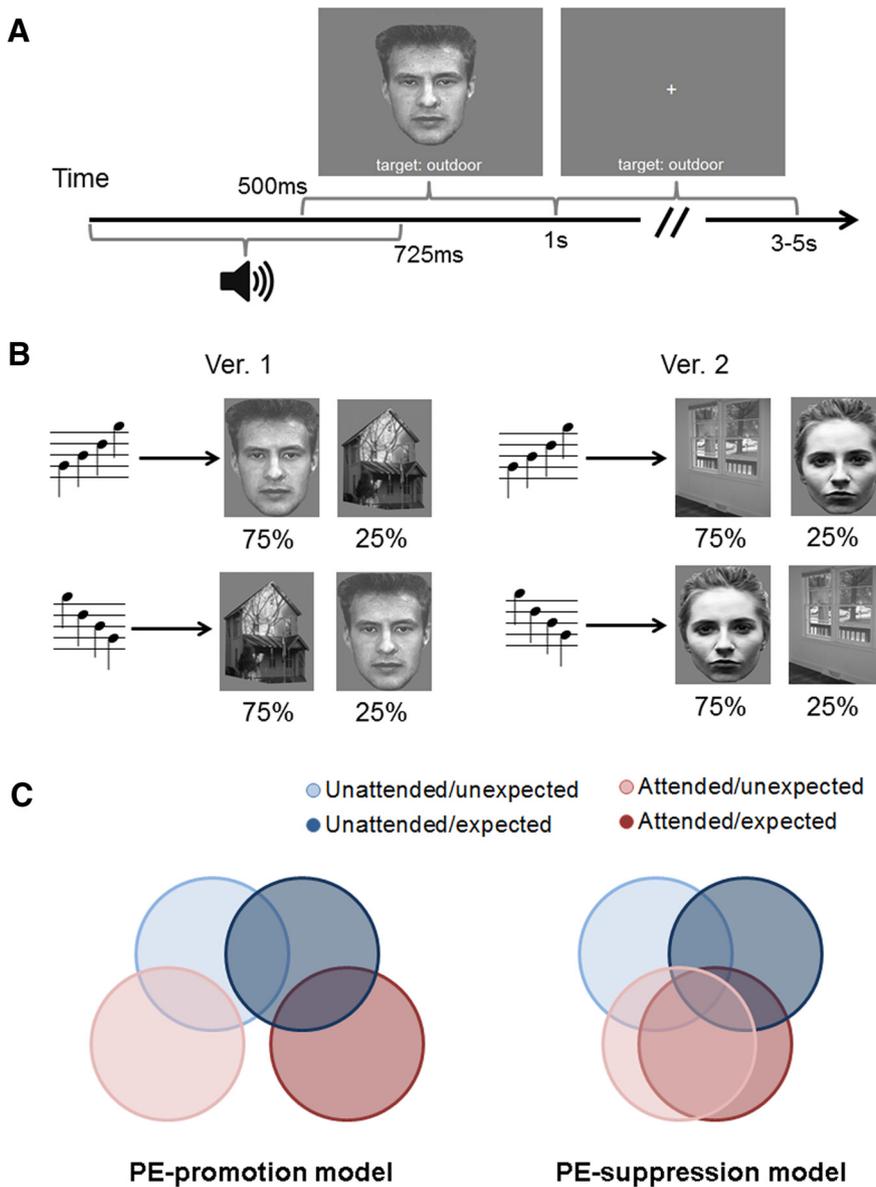
Figure 1. Experimental protocol and predictions. **A**, Timeline of an example nontarget trial. An auditory cue preceded each visual stimulus (face or scene), followed by a jittered intertrial interval. A reminder of the current target category (here: outdoor scenes) remained on screen throughout each block. **B**, Two versions of tone-picture associations were used, with each subject experiencing one version only. In version 1 (Ver. 1), a rising (falling) tone indicated a probability of 75% that the forthcoming nontarget stimulus was a male face (outdoor scene), respectively. In version 2 (Ver. 2), a rising (falling) tone indicated a probability of 75% that the forthcoming nontarget stimulus was an indoor scene (female face), respectively. **C**, Schematic illustration of model predictions: each disk represents the multivariate voxel pattern associated with a given experimental condition and the overlap between disks represents the degree of pattern similarity. If attention promotes error signals, then this should render the representations of unexpected stimuli more different from those of expected stimuli (left cluster), whereas the opposite would hold for attentional suppression of prediction errors (right cluster).

of multivoxel patterns associated with expected and unexpected information (the PE-promotion model).

Here, we characterized the manner in which attention and predictive processing interact by adjudicating between these multivariate predictions of the PE-promotion and PE-suppression accounts.

Materials and Methods

Participants. Twenty-one healthy, right-handed volunteers (7 males, 14 females, mean age = 25 years) with normal or corrected-to-normal vision gave informed consent in accordance with institutional guidelines.

Apparatus and stimuli. Stimulus delivery and behavioral data collection were performed using Presentation (http://www.neurobs.com/). Visual stimuli were presented on a back projection screen viewed via a head coil mirror, auditory stimuli were delivered via MRI-compatible headphones, and responses were collected using an MRI-compatible button box. Visual stimuli consisted of black-and-white photographs of four types: male faces, female faces, outdoor scenes, and indoor scenes (specifically, outside and inside views of buildings). These four stimulus types were selected to belong to two overarching stimulus categories, faces and scenes. Each stimulus type was represented by 60 unique photographs. Face images were aggregated from various databases (Egner et al., 2010). Scene stimuli were acquired from real estate websites, then cropped and adjusted to match the sizes and luminance of face stimuli (all stimuli subtended ~3° of horizontal and 4° of vertical visual angle). Auditory stimuli consisted of two tones (725 ms duration) composed of four consecutive notes (261.63, 392.44, 588.67, and 883.00 Hz) that were presented in ascending ("rising tone") or descending order ("falling tone").

Procedure. We independently manipulated feature-based attention to (i.e., relevance) and expectations of (i.e., probability) different stimulus categories (faces vs scenes). To manipulate attention, the protocol was designed as a rare target detection task. Specifically, the task was divided into five runs of six blocks each. At the beginning of each block, an instruction screen was shown for 4 s asking subjects to detect specific visual target stimuli (indicated by a button press). Target stimuli for a given block consisted of one of the four stimulus types (male faces, female faces, indoor scenes, outdoor scenes) and determined the stimulus features or category (faces, scenes) that the subject would pay attention to during that block. The instruction screen was followed by 16 nontarget trials and one to three target trials (mean, two) per block, randomly intermixed. Each trial consisted of a 725 ms auditory cue followed by 500 ms presentation of a visual face or scene stimulus (Fig. 1A). Trials were separated by exponentially jittered intertrial intervals (range = 3–5 s, step size = 1 s) during which a central fixation cross was displayed. As a reminder, the current target stimulus type was continuously displayed at the bottom of the screen.

In each block, the 16 nontarget trials (the focus of our analyses) consisted of eight stimuli belonging to the same category as the targets (e.g.,

dicts that attention will impair our ability to decode whether a stimulus was expected or unexpected. Another, complementary view is that attention promotes learning about the statistical structure of the world (Zhao et al., 2013), with classic theories proposing that attention increases the rate at which stimulus-stimulus associations are acquired (Rescorla and Wagner, 1972; Pearce and Hall, 1980). Under this view, attention acts not to suppress but to enhance PEs, acting as a multiplicative scaling factor on the impact of PEs on subsequent predictions (Feldman and Friston, 2010), which should increase (rather than decrease) the disparity
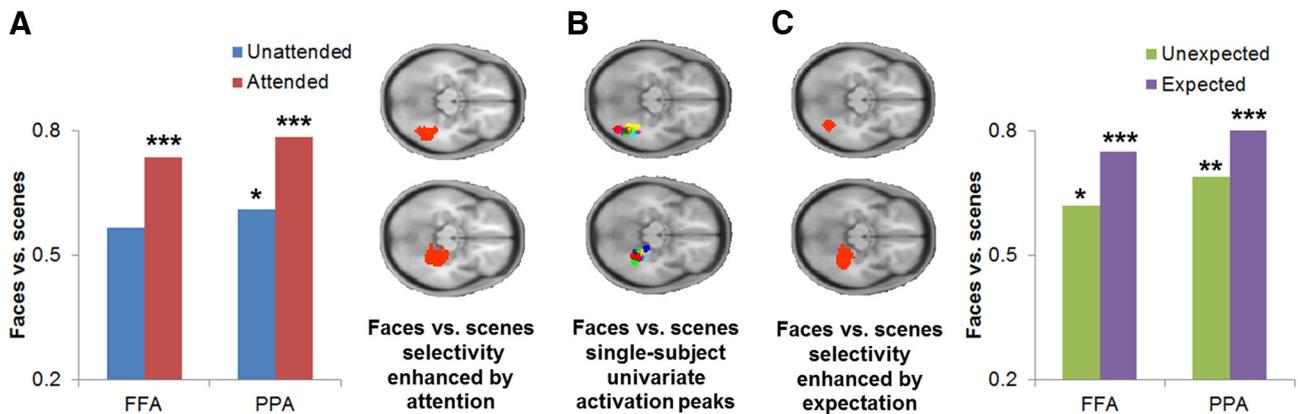
Figure 2. Category selectivity is enhanced by both attention and expectation. **A**, Searchlight MVPA ($p < 0.05$, corrected) identified regions in right fusiform gyrus (upper brain slice) and right parahippocampal gyrus (lower brain slice), where attention enhanced the discrimination between face and scene stimuli. **B**, For comparison, the upper and lower brain slices display single-subject activation peaks (one color per subject) in fusiform and parahippocampal gyri, respectively, as based on conventional functional definition of the FFA (univariate contrast of faces > scenes) and PPA (univariate contrast of scenes > faces). **C**, Searchlight MVPA ($p < 0.05$, corrected) identified regions in right fusiform gyrus (upper brain slice) and right parahippocampal gyrus (lower brain slice) where expectation enhanced the discrimination between face and scene stimuli. *$p < 0.05$; **$p < 0.001$; ***$p < 0.0001$.

male faces in a block where female faces are targets) and eight stimuli of one type of the opposing category (e.g., outdoor scenes). In this manner, nontarget trials in each block could be classified as being either a face or scene stimulus and as being either attended (e.g., male faces in a block where female faces are targets) or unattended (e.g., scenes in a block where faces are targets). To avoid target stimuli from one block proving distracting as nontargets in a different block, a given target stimulus type (e.g., female faces) was never used as a nontarget in other blocks. Specifically, for each subject, only one face stimulus type (either male or female) and one scene stimulus type (either indoor or outdoor) was used as a target and the other two stimulus types served as nontargets throughout the task; and targets and nontargets were counterbalanced across subjects. Face and scene target blocks were interleaved.

The manipulation of expectations consisted of probabilistic auditory tone-to-visual stimulus associations. For nontarget trials, one of the auditory cues (e.g., the rising tone) implied a 75% probability that the incoming stimulus was a face and the other cue (e.g., the falling tone) indicated a 75% probability that the forthcoming stimulus was a scene. This cue manipulation created expected (probable) and unexpected (improbable) nontarget trials. The specific cue-stimulus associations were consistent within subjects across blocks, but counterbalanced across subjects (Fig. 1B). For target trials, the cue-stimulus association was noninformative (50% probability), which was pointed out explicitly before the experiment. Subjects received two practice blocks (one for each target category) of prescan training. The above manipulations resulted in a 2 (stimulus: face vs scene) × 2 (attention: attended vs unattended) × 2 (expectation: expected vs unexpected) factorial design for nontarget trials. Trial counts for expected and unexpected stimuli were 90 and 30, respectively, for each of the attended and unattended conditions. Note, though, that the differential trial count was controlled for in the multivoxel pattern analysis (MVPA) described below because the number of features representing expected versus unexpected stimuli were equated.

Image acquisition and preprocessing. Images were acquired parallel to the AC-PC line on a 3T scanner (General Electric). Structural images were scanned using a T1-weighted SPGR axial scan sequence (120 slices, slice thickness = 1 mm, TR = 8.124 ms, FoV = 256 mm * 256 mm, in-plane resolution = 1 mm * 1 mm). Functional images were scanned using a T2*-weighted single-shot gradient EPI sequence of 36 contiguous axial slices (slice thickness = 3 mm, TR = 2 s, TE = 28 ms, flip angle = 90 °, FoV = 192 mm * 192 mm, in-plane resolution = 3 mm * 3 mm). Functional data were acquired in 5 runs of 226 images each. Preprocessing was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/). After discarding the first five scans of each run, the remaining images were realigned to their mean image and corrected for differences in slice-time acquisition. Each subject's structural image was coregistered to the mean functional image and normalized to the Montreal Neurological Institute

(MNI) template brain. The transformation parameters of the structural image normalization were then applied to the functional images.

Univariate image analyses. For each subject, a task model was created via vectors of visual stimulus onsets corresponding to the eight nontarget trial types, along with vectors for target trials, errors, head-motion parameters, and grand means of each run. Vectors were convolved with SPM8's canonical hemodynamic response function to produce a design matrix, against which the BOLD signal at each voxel was regressed. To relate MVPA findings to conventional functional definitions of the fusiform face area (FFA; Kanwisher et al., 1997) and parahippocampal place area (PPA; Epstein and Kanwisher, 1998), we computed univariate contrasts for face > scene stimuli (and vice versa) and determined single-subject peak activations within the fusiform gyrus (FFA) and parahippocampal gyrus (PPA; Fig. 2B) based on anatomical masks from the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). For these conventional analyses, normalized functional images were resampled to a resolution of 2 mm × 2 mm × 2 mm and smoothed using an 8 mm Gaussian kernel.

Multivariate image analyses. We performed MVPA on data from fusiform, parahippocampal, and inferior temporal gyri using an iterative leave-one-subject-out cross-validation scheme (Clithero et al., 2011; Haxby et al., 2011), gauging whether a classifier trained on fMRI data from all-but-one subjects could successfully decode neural responses in the left-out subject. This between-subject approach imposes additional constraints and is thus more conservative than the more commonly used within-subject MVPA. Although in the latter, a significant group finding for a given brain region can be obtained even if the individual subjects display completely distinct (including opposite) activation patterns, the former will only identify regions where multivariate patterns are replicable across subjects, thus ensuring the generalizability of our findings (Clithero et al., 2011; Haxby et al., 2011). In addition to the advantage of generalizability, unlike within-subject MVPA, the between-subjects approach also allowed us to collapse over male and female faces and indoor and outdoor scenes in the classification analyses. Therefore, the between-subject MVPA results also reflected more generalizable patterns with respect to stimulus types (faces vs scenes) than would have been achieved in a within-subject MVPA (where in a given subject, classifiers would be trained only on, e.g., male faces vs outdoor scenes or female faces vs indoor scenes).

To extract multivariate information content, the same models were fit to unsmoothed preprocessed images in their native resolution to reduce the blending of information patterns in the raw fMRI data. Then, for each trial type in each subject, a one-sample *t* test across runs was performed to produce a *t*-image. The *t*-images were further normalized across trial types by removing from each condition the cross-condition mean and dividing the resulting values by the cross-condition SD. This normaliza-

tion removed trial-type-independent, individual baseline activity that may confound the leave-one-subject-out cross-validation while retaining the activation differences between trial types. As a result, for each subject, we obtained one pattern (i.e., one t-image) for each nontarget trial type. Therefore, although there was a higher total number of expected than unexpected trials, these trial types equally contributed a single t-image per subject to the pattern classification analyses, so this analysis was not biased by unbalanced data points between trial types. The resulting t-images were defined as features containing task-relevant information (Jiang and Egner, 2013) on which a searchlight MVPA (Kriegeskorte et al., 2006) was conducted. Each searchlight was a spherical cluster with a radius of 2 voxels (6 mm) and contained up to 33 cortical voxels. A linear support vector machine (SVM) was used as the classifier and a default constrain value of 1 was used for all SVMs. The performance of SVMs was evaluated with an iterative leave-one-subject-out cross-validation procedure. After searchlight MVPA, a group classification accuracy image was obtained, in which each gray matter voxels encoded the average classification accuracy of the searchlight centered at that voxel.

Using this procedure, we first investigated the modulation of attention on the distinction between face and scene representations. If attention enhances the neural selectivity toward preferred categories in ventral visual regions, then face and scene stimuli should be more easily discriminated, resulting in higher classification accuracy in attended compared with unattended conditions (Serences et al., 2009; Chen et al., 2012). Therefore, the modulation of category selectivity by attention can be tested by comparing classification accuracy (faces vs scenes) between attended and unattended conditions. Compared with a typical two-way ANOVA approach for testing this type of interaction in univariate analyses, the present approach tests not just the effect of interaction (modulation), but also the "directionality" of the interaction (i.e., factor A's modulation on factor B or vice versa), which represents an additional advantage of using MVPA in this context.

Specifically, one set of classifiers were trained to classify face versus scene trials (or features) using only unattended trials and a second set of classifiers were trained to classify face versus scene trials using only attended trials. Therefore, for each searchlight we obtained 2 independent observations of accuracy, O_1 and O_2, for unattended and attended trials, respectively. A challenge here is that for each condition, between-subject MVPA produces only a single classification accuracy value for all subjects. Therefore, we cannot simply apply a paired t-test for drawing statistical inferences. To test the differences between classification results statistically, we therefore used a Bayesian approach: a null hypothesis of no modulation of attention on expectation can then be formulated as O_1 and O_2 being based on the same information content x, which is quantified in classification accuracy (i.e., the "true" classification accuracy for that searchlight). Therefore, given O_1, the probability of observing a classification accuracy O can be calculated as follows:

$$p(O|O_1) = \int p(O|x)\,p(x|O_1)\,dx$$

Here, $p(O|O_1)$ was calculated by iterating all possible values of the underlying true classification accuracy x. For each value of x, we first calculated the belief of x being the true classification accuracy given the observation O_1 or $p(x|O_1)$. We then applied this belief to calculate the probability of observing O using $p(O|x)$. The computation of $p(O|x)$ can be derived from a binomial distribution as follows:

$$p(O|x) = C_N^{NO} x^{NO} (1-x)^{N(1-O)}$$

where N is the number of cases in MVPA (e.g., 21 subjects × 2 classes × 2 cases [e.g., attended vs unattended trials and/or expected vs unexpected trials]). $p(x|O_1)$ can be calculated using Bayes' rule as follows:

$$p(x|O_1) = \frac{p(O_1|x)\,p(x)}{p(O_1)}$$

where $p(O_1)$ is a constant and can be replaced by normalizing $p(O|O_1)$ in the final step. $p(x)$ is the distribution of the true classification accuracy.

Assuming that the observations were unbiased estimates of x, $p(x)$ can be approximated by the distribution of O_1 and O_2 across the whole brain. Using $p(O|O_1)$, we then tested the null hypothesis by calculating the probability of observing an accuracy that is no less than O_2 (if $O_2 \geq O_1$; or no greater than O_2 if $O_2 < O_1$), given O_1 and the assumption that O_1 and O_2 were derived from the same x. The probability of observing an accuracy of no less than O_2 can be calculated as follows:

$$p(O \leq O_2|O_1) = \sum_{O=O_2}^{N} p(O|O_1)$$

The p-value for the "no greater than" case can be calculated correspondingly. Given our focus on face- and scene-selective regions, we constrained these analyses to a mask of the ventral visual cortex, defined as a conjunction of the bilateral fusiform, parahippocampal, and inferior temporal gyri, as delineated in the AAL template (Tzourio-Mazoyer et al., 2002). This mask was further dilated by 2 voxels (6 mm) to account for potential discrepancy of gray matter classification between the AAL template and our analyses. All MVPA results we report were corrected for multiple comparisons at $p < 0.05$ for combined searchlight classification accuracy and cluster extent thresholds, using the Analysis of Functional NeuroImages AlphaSim algorithm (http://afni.nimh.nih.gov/). A total of 5000 Monte Carlo simulations determined that an uncorrected searchlight accuracy p-value threshold of <0.01 in combination with a searchlight cluster size 6–7 searchlights ensured a false discovery rate of <0.05. In addition, for MVPA ROI-based analysis, we investigated whether a given mean ROI classification accuracy O_{ROI} differed from chance level (50%) using a two-tailed test as follows:

$$P = p(O > \max[O_{ROI}, 1 - O_{ROI}]|x = 0.5)$$
$$+ p(O \leq \min[O_{ROI}, 1 - O_{ROI}]|x = 0.5)$$

Note that due to the nature of this between-subject classification analysis, the sample only contained a single observation in each ROI-mean classification accuracy analysis, and therefore no estimates of variance across observations (i.e., error bars) are displayed in the corresponding figures (Figs. 2 A, C, 3 A, B, and 4D). Instead, to indicate the classification accuracies' positions with respect to the variance of the null distributions, we cite z-scores of the classification accuracies. After the test of attentional modulation of category selectivity, we reiterated the same analyses to gauge the effects of expectation on category selectivity, training one set of classifiers to discriminate face versus scene features using only unexpected trials and a second set using only expected trials, and then identifying searchlights in which the latter was more successful than the former. Note that given the above approach to statistical inference, it was not feasible to conduct a three-way interaction test with the factors of attention, expectation, and stimulus category. Probing this type of three-way interaction would involve four classification accuracy estimates, requiring a 4D null distribution of $\sim$3,400,000 (43^4) data cells. Given that the fMRI images contained only $\sim$45,000 gray matter searchlights, it was impractical to sample a robust null distribution at this level.

To determine whether attention and expectation boosted category selectivity by shared or distinct mechanisms, we trained searchlights that displayed attentional enhancement of category selectivity in the above analysis to discriminate between attended face and scene stimuli and then tested these classifiers' ability to discriminate these categories in unattended/expected and unattended/unexpected trials. The converse analysis was also performed, training searchlights of significant expectation-enhanced category selectivity on expected trials and testing their ability to discriminate attended or unattended unexpected stimuli. Finally, we tested the modulation of attention on the effects of expectation. Here, instead of classifying between face and scene stimuli, we trained classifiers to distinguish between expected and unexpected stimuli. Then, we used the same approach as above to test whether expected stimuli could be distinguished from unexpected stimuli with higher accuracy under attended than unattended conditions or vice versa. To assess categorical stimulus specificity of the FFA/PPA, two MVPAs were conducted, one using only
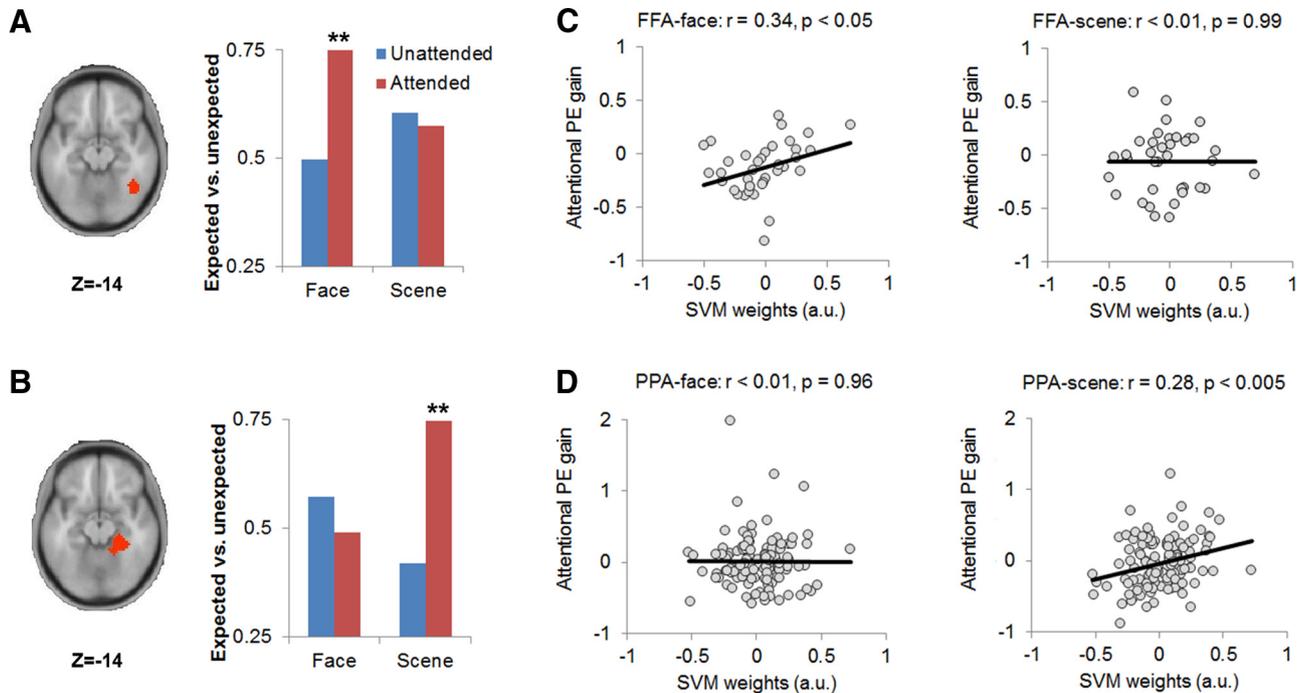
Figure 3. Attention enhances the distinction between unexpected and expected stimuli. *A*, *B*, Searchlight MVPA ($p < 0.05$, corrected) identified regions in right fusiform gyrus (*A*) and right parahippocampal gyrus (*B*) where attention enhanced the discrimination between expected and unexpected faces and scenes, respectively. *C*, *D*, The (univariate) attentional modulation of expected versus unexpected stimuli for each individual FFA/PPA voxel (*y*-axis) is plotted against the voxels' SVM classifier weight in the attended condition. *C*, FFA voxel weights for classifying face stimuli ("FFA-face," left) and scene stimuli ("FFA-scene," right). *D*, PPA voxel weights for classifying face stimuli ("PPA-face," left) and scene stimuli ("PPA-scene," right). **$p < 0.001$.

face trials and one using only scene trials. All MVPA results were corrected for multiple comparisons at $p < 0.05$ (see above).

To determine whether superior distinction between expected and unexpected stimuli under attention (Fig. 3 *A*,*B*) was attributable to multivariate signals, we repeated the above analyses on univariate FFA/PPA data by substituting each searchlight's multivariate signal pattern with its mean univariate *t*-value (Jiang and Egner, 2013). In addition, we ran standard ANOVAs involving the factors of stimulus category, attention, and expectation on these mean values (collapsed across FFA and PPA; Fig. 4*A*). Moreover, we addressed the possibility that our initial classification results were driven by a single, outlying condition (the attended/expected condition; Fig. 4*C*). In other words, perhaps attention and expectation would jointly make activation patterns more distinct from the unattended/unexpected conditions. It follows from this hypothesis that, in expected trials, attended and unattended stimuli should be distinguished with higher accuracy than in unexpected trials. To rule out this possibility we tested the ability of classifiers trained on FFA/PPA voxels to decode whether a stimulus was attended or unattended in expected compared with unexpected conditions using the same methods described above (Fig. 4*D*). Finally, we visualized univariate effects by running a one-sample *t* test on attention × expectation interaction contrasts for the FFA and PPA voxels that had been identified in the attention × expectation MVPA (Fig. 4*B*).

Results

We adjudicated between the PE-promotion and PE-suppression models of attention-expectation interaction by applying a between-subjects (Clithero et al., 2011; Haxby et al., 2011) searchlight MVPA (Kriegeskorte et al., 2006; see Materials and Methods) to fMRI data acquired from the category-selective visual regions, the FFA and the PPA, during a task that independently manipulated attention to, and expectation of, face and scene stimuli (Fig. 1*A*,*B*). In alternating blocks, participants searched for rare female face targets (face attention blocks) or

rare outdoor scene targets (scene attention blocks) in a stream of frequent male face and indoor scene nontarget stimuli (target/nontarget category assignments were counterbalanced across subjects, see Materials and Methods). Each stimulus was preceded by an auditory cue (Fig. 1*A*) that was 75% predictive of the forthcoming stimulus category (face vs scene; Fig. 1*B*). Therefore, nontarget faces and scenes (the foci of our analyses) could be classified as either attended (e.g., male faces in a block where female faces are targets) or unattended (e.g., scene stimuli in a block where face stimuli are targets) and as being either expected or unexpected (i.e., probable or improbable in relation to the auditory cue). This design allowed us to assess whether, under attended (vs unattended) conditions, neural pattern classifiers could distinguish expected from unexpected stimuli with greater accuracy (PE-promotion hypothesis; Feldman and Friston, 2010) or reduced accuracy (PE-suppression hypothesis; Rao and Ballard, 2005; Fig. 1*C*). Subjects ($n = 21$) detected targets with very high accuracy (mean target detection = 99%; mean false alarm rate = <1%), suggesting that they performed the task as instructed.

In the following fMRI analyses, we first seek to establish that our manipulations of attention and expectation were successful by attempting to replicate previously reported effects of attention and expectation on category selectivity in ventral visual cortex. This is followed by additional, novel analyses gauging whether these effects are mediated by the same underlying mechanisms. The results of these first sets of analyses provide a solid foundation on which to base the test of our main question of interest: how attention modulates the effects of expectation on stimulus processing. Finally, the results of that analysis are then subjected to a number of control analyses.
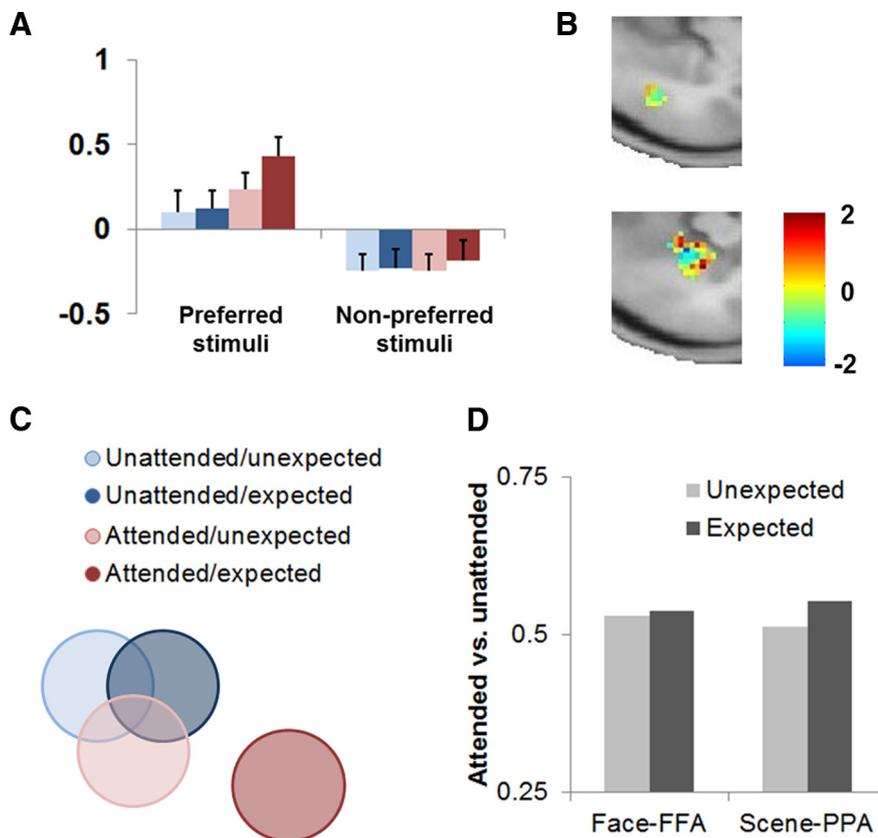
Figure 4. Control analyses. **A**, Univariate fMRI effects of attention, expectation, and stimulus categories (preferred vs nonpreferred, collapsed across FFA and PPA). The bar graphs show group mean activation estimates in MVPA feature values (t-values of activation normalized across experimental conditions within each subject) plotted as a function of the attention, expectation, and stimulus category conditions. The FFA and PPA ROIs over which these data are collapsed were defined by the clusters reported in our analysis of attentional modulation on expectation (Fig. 3 A, B). **B**, Voxelwise t-values for the univariate attention × expectation interaction group contrast in voxels of the FFA (top) and PPA (bottom) as defined by MVPA results in Figure 3, A and B. **C**, Schematic illustration of how MVPA results could in theory be driven by a single distinct multivariate pattern in the attended/ expected stimulus condition. Each disk represents the multivariate voxel pattern associated with a given experimental condition and the overlap between disks represents the degree of pattern similarity. **D**, MVPA results for classifiers attempting to distinguish attended from unattended preferred stimuli, plotted as a function of expectation.

Attention enhances the neural distinction between stimulus categories

We began by verifying that attention enhanced selectivity for object categories, as described previously (Serences et al., 2009; Jehee et al., 2011; Chen et al., 2012). Separate multivariate classifiers were trained on fMRI data from ventral visual cortex (fusiform, parahippocampal, and inferior temporal gyri; see Materials and Methods) to discriminate between face and scene stimuli under attended and unattended conditions, respectively. We then identified searchlights where the distinction between face and scene stimuli was significantly enhanced under attention (these analyses collapsed across expected and unexpected stimuli). As shown in Figure 2A, attention to face stimuli resulted in significantly improved classification accuracy between faces and scenes in the right fusiform gyrus (2-way interaction, $p < 0.05$, corrected; mean attended accuracy = 74%, $z = 4.4$, above chance at $p < 0.0001$; mean unattended accuracy = 56%, $z = 1.1$, $p > 0.1$), whereas attention to scene stimuli significantly improved this discrimination in the right parahippocampal gyrus (2-way interaction, $p < 0.05$, corrected; mean attended accuracy = 79%, $z = 5.3$, above chance at $p < 0.0001$; mean unattended accuracy = 61%, $z = 2.0$, above chance at $p < 0.05$). As an anatomical reference, Figure 2B displays single-subject activation peaks in fusiform and

parahippocampal gyri based on conventional functional definition of the FFA (univariate contrast of faces > scenes) and PPA (univariate contrast of scenes > faces), respectively. It can be seen that the ventral visual regions sensitive to attentional enhancement of category selectivity overlapped closely with the FFA and PPA territories as defined by conventional within-subject category contrasts.

Expectation enhances the neural distinction between stimulus categories

Next, we performed a corresponding analysis to assess whether expectation may also boost category selectivity. Such a benefit is implied by behavioral facilitation effects and has found support in a recent study of multivoxel patterns in primary visual cortex (Kok et al., 2012a), but has not been tested previously for higher-level object categories. Collapsing across the attention factor, we trained separate classifiers to discriminate between face and scene stimuli under expected and unexpected conditions, respectively, and then located searchlights where this discrimination was significantly enhanced for expected stimuli. As shown in Figure 2C, discrimination between face and scene stimuli in right fusiform gyrus and parahippocampal gyrus sites was more accurate for expected than for unexpected stimuli (FFA: 2-way interaction, $p < 0.05$, corrected; mean expected accuracy = 75%, $z = 4.6$, above chance at $p < 0.0001$; mean unexpected accuracy = 62%, $z = 2.2$, $p < 0.05$; PPA: 2-way interaction, $p < 0.05$, corrected; mean expected accuracy = 81%, $z = 5.7$, above chance at $p < 0.0001$; mean unexpected accuracy = 69%, $z = 3.5$, above chance at $p < 0.001$). The clusters identified in this analysis (Fig. 2C) were again located in the same territory as those identified by conventional FFA/PPA definition (Fig. 2B) and by the attentional modulation analysis (Fig. 2A). Note that we did not conduct a three-way interaction test for attention, expectation, and stimulus category because sampling a robust null distribution to test this interaction against was not feasible in our analysis scheme (see Materials and Methods).

Attention and expectation boost category selectivity via distinct mechanisms

Even though attention and expectation were manipulated orthogonally in our design, the fact that both factors enhanced category selectivity ultimately raises the question of whether they do so via the same mechanism (Rao, 2005; Yu and Dayan, 2005) or via distinct mechanisms (Kok et al., 2012a; Wyart et al., 2012). To address this question in the current dataset, we trained searchlights that displayed significant attentional enhancement of category selectivity in the above analysis to discriminate between attended face and scene stimuli. We then tested these classifiers' ability to discriminate stimulus categories in unattended/expected and unattended/unexpected trials. If attention and expec-

tation operate via shared mechanisms, then these searchlights should display better classification of expected trials because expected stimuli should have a similar neural signature to the attended stimuli that comprised the training data. Contrary to this prediction, however, only two of the 105 searchlights in the PPA (and none in the FFA) showed significantly greater classification accuracy ($p < 0.01$) in expected compared with unexpected unattended stimuli (vs three searchlights showing the opposite effect). Similarly, in the converse analysis, training searchlights of significant expectation-enhanced category selectivity on expected trials and testing their ability to discriminate attended or unattended unexpected stimuli, only five of the 120 searchlights in the PPA (and none in the FFA) showed greater classification accuracy for attended compared with unattended unexpected stimuli (vs 14 searchlights showing the opposite effect). In other words, voxel patterns that allowed faces and scenes to be discriminated most effectively under attention were quite different from those that allowed maximal discrimination under expectation. These data support the view that attention and expectation boost category selectivity via different underlying mechanisms.

The results so far replicate previous findings of attentional (Serences et al., 2009; Jehee et al., 2011; Chen et al., 2012) and expectation-driven (Kok et al., 2012a) enhancement of (multivariate) category selectivity in the ventral visual stream and also suggest that these effects are mediated by distinct underlying mechanisms (Kok et al., 2012a; Wyart et al., 2012). This sets the stage for our prime analysis of interest: to determine how attention may modulate the neural distinction between expected and unexpected faces and scenes in FFA and PPA.

Attention enhances the neural distinction between expected and unexpected stimuli

To adjudicate between the PE-promotion and PE-suppression hypotheses of attention-expectation interactions, we estimated whether attention improves (PE-promotion model) or dampens (PE-suppression model) the classification of expected versus unexpected stimuli from multivoxel data in ventral visual cortex (Fig. 1C). Adopting a comparable strategy to the preceding analyses, we began by training classifiers on searchlights from ventral visual cortex to distinguish expected from unexpected stimuli under attended and unattended conditions, respectively, performing these analyses separately for face and scene stimuli. We then identified searchlights where attention significantly modulated the classification accuracy for distinguishing expected from unexpected stimuli. Note that because of our factorial design, these analyses are orthogonal to the analyses on attention- and expectation-based modulation of category selectivity reported above.

As can be seen in Figure 3A, attention greatly enhanced the distinction between expected and unexpected faces in the FFA (2-way interaction: $p < 0.05$, corrected), raising discrimination performance from chance (50% accuracy, $z = 0$) to 75% ($z = 3.2$, above chance at $p < 0.001$). However, the same FFA searchlights were incapable of classifying expected versus unexpected scenes regardless of attention (unattended scenes accuracy = 0.57, $z = 0.9$, attended scenes accuracy = 0.60, $z = 1.3$; neither different from chance, $ps > 0.1$). Similarly, attention amplified the discrimination between expected and unexpected scenes in the PPA (Fig. 3B, 2-way interaction: $p < 0.05$, corrected), from chance (accuracy = 0.42, $z = -1.0$, no different from chance, $p > 0.1$) to 75% accuracy ($z = 3.2$, above chance at $p < 0.001$), but the same searchlights were unable to classify faces regardless of attention (PPA unattended faces accuracy = 0.49, $z = -0.1$, PPA attended

faces accuracy = 0.60, $z = 1.3$; neither different from chance, $ps > 0.1$). Both the fusiform and parahippocampal searchlight clusters identified in these analyses coincide closely with the territories obtained in the MVPA on attention- and expectation-based modulation of stimulus category selectivity (above), as well as with the conventional, univariate FFA and PPA definitions (cf. Fig. 2).

To provide an intuition of how individual voxels contributed to these results, in Figure 3, C and D, we regressed the univariate attentional modulation of expected versus unexpected stimuli for each individual FFA/PPA voxel against the corresponding classifier weight in the attended condition. Consistent with the rationale for our MVPA, voxels displaying stronger attentional modulation also carry greater weight in contributing to classification (in a stimulus category-selective fashion), driving a positive correlation. In sum, these data indicate that attention facilitates the neural distinction between expected and unexpected stimuli in category-specific regions of the ventral visual stream. This result is compatible with a PE-promoting, but not a PE-suppressing, role of attention.

Are these results attributable to multivariate or univariate effects?

The above results document that attention facilitated the neural distinction between expected and unexpected stimuli, but what exactly underlies this effect? Predictive coding proposes that computations underlying predictions and error-driven adjustments cooccur in local neural circuits (Rao and Ballard, 1999; Friston, 2005; Bastos et al., 2012). Although a given fMRI voxel in the FFA/PPA can then be assumed to contain an intermingled set of predictive and error signals, these may, due to random sampling, be biased to respond preferentially to expected or unexpected stimuli in a manner akin to, for example, biased orientation selectivity in voxels of primary visual cortex (Kamitani and Tong, 2005). Strong support for this contention can be found in a recent fMRI study showing intermingled voxels in the FFA that consistently (across scanner runs) respond to repeated face stimuli with either a suppressed or enhanced signal (de Gardelle et al., 2013). Given this scenario, the present results can be parsimoniously interpreted as attention promoting error signals (and thus activity in error-biased voxels), thereby rendering the multivoxel patterns associated with unexpected stimuli more distinct from those elicited by expected stimuli.

However, the mere fact that MVPA was more successful at decoding expected versus unexpected stimuli under attended conditions does not imply that the signals exploited by the classifiers for this discrimination were actually multivariate in nature nor that they relied on interspersed voxels of differential sensitivities to prediction and prediction error signals. Instead, our results might simply reflect a univariate (mean signal) advantage for the attended/expected or attended/unexpected conditions in category-specific brain regions. To rule out this possibility, we explored the data at the univariate level. First, we analyzed activation estimates (collapsed across FFA and PPA) in terms of mean MVPA feature values (i.e., t-values of activation normalized across experimental conditions within each subject) in a conventional ANOVA involving the factors of stimulus category, attention, and expectation (Fig. 4A). We observed a main effect of stimulus category ($F_{1,20} = 117.6$, $p < 0.001$), due to higher activation to preferred than nonpreferred stimuli, and a main effect of attention ($F_{1,20} = 5.9$, $p < 0.05$), due to higher activation when stimuli were attended than unattended. These effects were qualified by a marginally significant interaction between attention

and stimulus category ($F_{1,20} = 4.3$, $p = 0.052$), reflecting a greater effect of stimulus category for attended than unattended stimuli, although category effects were robust in either condition (attended: $F_{1,20} = 74.6$, $p < 0.001$; unattended: $F_{1,20} = 29.2$, $p < 0.001$). Crucially, no attention by expectation interaction effect of the kind we detected in the multivariate analyses was observed. To ensure that these differences between multivariate and univariate results were not attributable to univariate data being extracted from ROIs based on the multivariate findings, we reran these analyses based on ROIs defined by univariate face versus scene stimuli contrasts. The results were qualitatively equivalent to those seen in Figure 4A (data not shown).

Second, to ensure maximum comparability with multivariate analyses, we reran the above classification analysis substituting each searchlight's multivariate signal pattern with its mean t-value (Jiang and Egner, 2013). Based on these data, however, FFA and PPA searchlights could neither distinguish expected from unexpected stimuli at above chance levels in the attended condition (FFA attended faces accuracy = 0.58, $z = 1.0$; PPA attended sense accuracy = 0.57, $z = 0.9$; neither different from chance, $ps > 0.2$) nor in the unattended condition (FFA unattended faces accuracy = 0.54, $z = 0.5$; PPA unattended sense accuracy = 0.44, $z = -0.8$; neither different from chance, $ps > 0.2$), with no difference between conditions. In an additional analysis, we selected all of the FFA/PPA searchlights that displayed a significant interaction effect between attention and expectation in the multivariate analyses (all $ps < 0.01$) and tested whether they would also show this effect when considering only their mean signal. This was the case for only 18% of those searchlights, indicating that the multivariate results cannot be accounted for by univariate signals alone.

To further corroborate the assumption that the classifiers capitalized on multivariate, functionally heterogeneous responses within these ventral visual stream regions (that were nevertheless stable across subjects), we computed voxelwise t-values for the univariate attention $\times$ expectation interaction group contrast in FFA and PPA. Consistent with this premise, we observed interspersed positive and negative voxelwise t-values for this interaction effect in both the FFA and PPA (Fig. 4B). These data support the idea that perceptual inference relies on regionally intermingled expectation and error signals in visual cortex (Rao and Ballard, 1999; Friston, 2005; Bastos et al., 2012). This data pattern also indicates that mean population signals from standard univariate analyses can obscure interactions between expectation and attention (cf. Fig. 4A).

Are these results driven by a single outlier condition?
A second concern is that better classification under attended conditions could be driven by a unique multivariate signature for the attended/expected condition, rather than enhanced classification of expected and unexpected signals under attention per se (cf. Figs. 4C, 1C). If this were the case, then we should not only observe enhanced decoding sensitivity for expected versus unexpected stimuli under attended conditions, but also enhanced decoding sensitivity for attended versus unattended stimuli in the expected condition (Fig. 4C). To rule out this alternative, we tested the ability of classifiers trained on FFA/PPA voxels (as defined by the above MVPA results; Fig. 3A,B) to decode whether a stimulus was attended or unattended in expected compared with unexpected conditions. This control MVPA established that the performance of attention classifiers did not vary as a function of expectation condition (Fig. 4D; mean accuracy gain on decoding attention status [expected − unexpected] = 0.01/

0.04 for FFA/PPA respectively, both $zs < 0.5$, both $ps > 0.33$). Therefore, our findings of an attention-enhanced neural distinction between expected and unexpected stimuli cannot be attributed to the modulation of expectation on attention and are more parsimoniously accounted for by attentional enhancement of PEs.

Discussion
Task relevance and contextual probability are both known to enhance the recognition of, and neural selectivity for, visual stimuli. However, the functional relationship between these key determinants of visual cognition is uncertain (Summerfield and Egner, 2009). We adjudicated between two rival hypotheses. The first hypothesis is that attention could filter out the processing of unexpected stimuli (Rao and Ballard, 2005), suppressing neural error signals and obviating reconciliation between predicted and observed percepts, thereby rendering neural representations of expected and unexpected stimuli more similar. The second hypothesis is that attention could enhance the processing of surprising stimuli, promoting neural error signaling and prediction updating (Feldman and Friston, 2010), thereby rendering neural representations of expected and unexpected stimuli more distinct. In strong support of the latter theory, we found that attention greatly boosted our ability to distinguish between multivoxel patterns of expected versus unexpected stimuli. This finding is consistent with attention enhancing perceptual PEs, potentially promoting both rapid online belief updating and longer-term learning about the statistical structure of goal-relevant properties of the environment (Chun and Turk-Browne, 2007). In other words, although expectations supply prior beliefs about the most likely causes for a percept, attention may determine the rate and efficacy with which these predictions are adjusted to reflect the most probable state of the world (Summerfield and Egner, 2013).

Our findings concur with the emerging view that attention enhances the precision (inverse of variance) of PEs by increasing the gain of error processing in a multiplicative fashion (Feldman and Friston, 2010). This perspective draws support from a number of recent findings. First, psychophysical data in humans indicate that stimulus relevance, when de-confounded from stimulus probability, leads to a reduction in internal processing noise (Wyart et al., 2012) equivalent to enhanced precision of bottom-up (error) signals in the predictive coding framework. Second, the claim that attention acts via internal noise reduction is also supported by monkey electrophysiological studies demonstrating that attentional facilitation of neuronal signaling in visual cortex is attributable primarily to the suppression of noise correlations (shared variability) across local neuronal populations, rather than to enhanced neuronal firing rates (Cohen and Maunsell, 2009; Mitchell et al., 2009). Finally, the present results tie attention intimately to learning processes, because the promotion of PEs should lead to accelerated online error correction (belief updating) in interpreting task-relevant visual inputs. This proposition is highly compatible with the well established effects of attention in promoting memory encoding (Craik et al., 1996) and robust long-term memory benefits (Chun and Turk-Browne, 2007).

Our study shares similarities with recent work by Kok et al (2012a) in which enhanced decoding of grating orientations was observed in V1 when orientations were validly cued (i.e., expected). Notably, that study found independent, noninteracting decoding benefits of attention and expectation, whereas we here report the decoding of expected versus unexpected stimuli to be enhanced by attention. Although these results are superficially

contradictory, the two studies addressed distinct questions: Kok et al. (2012a) were interested in decoding the identity of particular gratings rather than decoding their status of being expected or unexpected. In contrast, we investigated whether the neural differentiation between expected versus unexpected stimulus category members would be enhanced or suppressed by attention. Although the two studies' implications for attention-expectation relations are therefore not directly comparable, the current study nevertheless extends Kok et al.'s findings of expectation-based enhancement of category selectivity from simple stimulus features in early visual cortex to complex object representations at higher levels of the ventral visual stream. Specifically, Kok et al. (2012a) found that expectation benefited the decoding of grating orientations in V1, but not in areas V2 and V3. The investigators proposed two potential explanations: either improved decoding in V1 reflected that region's preference for simple oriented stimuli or higher visual regions are generally less susceptible to predictive processing. Our results argue against the latter possibility, because expectation greatly enhanced classification accuracy in high-level areas of the ventral visual stream. Therefore, enhanced selectivity for expected features appears to be a general purpose mechanism by which context modulates perception across the visual hierarchy.

The present study used a multivariate analysis approach ideally suited to addressing how the category selectivity of large-scale neural population activity is modulated by stimulus relevance, stimulus probability, and their interaction. An additional merit of this strategy is that it facilitated a clean juxtaposition of clearly distinct predictions derived from PE-promotion versus PE-suppression views of attention's role in perceptual inference, which may be more difficult when considering mean (mass-univariate) neural population signal. First, prior univariate fMRI studies on the interaction between attention and expectation have produced ambivalent results. For example, in one study, the suppression of visual neural responses to expected relative to unexpected stimuli was only observed when those stimuli were in the focus of attention (Larsson and Smith, 2012), whereas in another study, attention actually led to a greater neural response to expected compared with unexpected stimuli (Kok et al., 2012b, but see Kok et al., 2012a), a data pattern that, descriptively, was also found in the present univariate analysis (Fig. 4A). Second, although, prima facie, the PE-promotion model appears to imply larger prediction errors under attention, the model has at times been construed as predicting a relative suppression of mean neural population responses to unexpected relative to expected stimuli under attention (Kok et al., 2012b), a data pattern that would similarly be anticipated by the PE-suppression model. In contrast, the multivariate predictions under the two models are clearly divergent.

It is important to emphasize that the present classification results were obtained using a between-subject MVPA approach, in which neural responses in a given participant are decoded on the basis of classifiers trained on data from all other participants (Clithero et al., 2011; Haxby et al., 2011). This approach is more conservative than typical within-subjects MVPA because it imposes the additional constraint of the patterns in question being replicable over subjects, thus testing for neuroanatomically stable functional organization of responses, which ensures the generalizability of results. Cross-subject reliability of multivariate neural signals has been demonstrated previously in ventral visual cortex for a variety of complex visual stimuli (Haxby et al., 2011); however, the present results are the first to show that the attentional and contextual top-down biasing of such stimulus category selec-

tivity, and the attentional promotion of perceptual PEs, is also anatomically replicable across participants. This implies a high degree of similarity over subjects in the anatomical organization of predictive and PE signaling units, at least at the macroscopic level of voxel-based signals, and represents an important extension of previous data documenting temporally stable prediction/error signaling for single FFA voxels within subjects (de Gardelle et al., 2013). An important question for future studies is at which level of spatial resolution between-subject similarities in functional neuroanatomy can be observed and at which level (smaller or larger than the present scale) they may break down.

Finally, a perhaps surprising feature of our results is that classification of unattended expected versus unexpected stimuli was at chance, suggesting attention to be a precondition for discriminating expected from unexpected stimuli. The prior literature on this is mixed, with some studies suggesting expectation-based effects to be strongly dependent on attention (Larsson and Smith, 2012) and others report robust learning of statistical structure in the absence of attention and even awareness (Fischer et al., 1999; Brázdil et al., 2001; Turk-Browne et al., 2009). An important mediating factor may be the level at which expectations are acquired and expressed. For example, violations of local, low-level regularities can be detected in early sensory processing in the absence of attention, whereas those concerning more global regularities, requiring longer temporal integration over stimulus events, are detected at later processing stages and dependent on attention (Bekinschtein et al., 2009). The present design likely involved relatively "global" expectations, because they were based on the frequency of cue-stimulus pairings observed across trials. Moreover, the null effect in question was obtained in the context of a between-subject classification analyses. A parsimonious interpretation of this observation is that there exists a large degree of variance in error signaling across subjects when stimuli are unattended (thus thwarting classification attempts), but that attentional enhancement of the precision of PE signals substantially reduces this source of noise (thus enabling successful between-subject classification). One way of testing this interpretation would be to contrast within- and between-subject classification performance directly (Clithero et al., 2011). However, due to a small number of runs and a relatively low count of expected and unexpected trials per stimulus category per attentional condition within each run, the present study design is suboptimal for within-subject MVPA, so an appropriate comparison between within- and between-subject analytic sensitivity must be left for future study.

In conclusion, our data provide robust support for an emerging view of visual cognition in which learned predictions of probable inputs are used to "explain away" expected bottom-up signals and attention serves to boost the precision of unexpected stimulus information. This facilitates the reconciliation of that information with predictions concerning the most probable cause of a current percept and promotes longer-term learning about the statistical structure of goal-relevant properties of the environment.

References

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. J Neurosci 30: 2960–2966. CrossRef Medline

Bar M (2004) Visual objects in context. Nat Rev Neurosci 5:617–629. CrossRef Medline

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711. CrossRef Medline

Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious processing of auditory regularities. Proc Natl Acad Sci U S A 106:1672–1677. CrossRef Medline

Biederman I (1972) Perceiving real-world scenes. Science 177:77–80. CrossRef Medline

Brázdil M, Rektor I, Daniel P, Dufek M, Jurák P (2001) Intracerebral event-related potentials to subthreshold target stimuli. Clin Neurophysiol 112:650–661. CrossRef Medline

Broadbent DE (1958) Perception and Communication. London: Pergamon.

Chen AJ, Britton M, Turner GR, Vytlacil J, Thompson TW, D'Esposito M (2012) Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. Front Hum Neurosci 6:187. CrossRef Medline

Chun MM, Turk-Browne NB (2007) Interactions between attention and memory. Curr Opin Neurobiol 17:177–184. CrossRef Medline

Clithero JA, Smith DV, Carter RM, Huettel SA (2011) Within- and cross-participant classifiers reveal different neural coding of information. Neuroimage 56:699–708. CrossRef Medline

Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. Nat Neurosci 12:1594–1600. CrossRef Medline

Craik FI, Govoni R, Naveh-Benjamin M, Anderson ND (1996) The effects of divided attention on encoding and retrieval processes in human memory. J Exp Psychol Gen 125:159–180. CrossRef Medline

de Gardelle V, Waszczuk M, Egner T, Summerfield C (2013) Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. Cereb Cortex 23:2235–2244. CrossRef Medline

den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. Cereb Cortex 19:1175–1185. CrossRef Medline

Egner T, Monti JM, Summerfield C (2010) Expectation and surprise determine neural population responses in the ventral visual stream. J Neurosci 30:16601–16608. CrossRef Medline

Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598–601. CrossRef Medline

Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci 4:215. CrossRef Medline

Fischer C, Morlet D, Bouchet P, Luaute J, Jourdan C, Salord F (1999) Mismatch negativity and late auditory evoked potentials in comatose patients. Clin Neurophysiol 110:1601–1610. CrossRef Medline

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836. CrossRef Medline

Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T, Kilner JM (2008) The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42:936–944. CrossRef Medline

Haxby JV, Guntupalli JS, Connolly AC, Halchenko YO, Conroy BR, Gobbini MI, Hanke M, Ramadge PJ (2011) A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72:404–416. CrossRef Medline

Jehee JF, Brady DK, Tong F (2011) Attention improves encoding of task-relevant features in the human visual cortex. J Neurosci 31:8210–8219. CrossRef Medline

Jiang J, Egner T (2013) Using neural pattern classifiers to quantify the modularity of conflict-control mechanisms in the human brain. Cereb Cortex. Advance online publication. doi:10.1093/cercor/bht029. CrossRef Medline

Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685. CrossRef Medline

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311. Medline

Keller GB, Bonhoeffer T, Hübener M (2012) Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. Neuron 74:809–815. CrossRef Medline

Kok P, Jehee JF, de Lange FP (2012a) Less is more: expectation sharpens representastions in the primary visual cortex. Neuron 75:265–270. CrossRef Medline

Kok P, Rahnev D, Jehee JF, Lau HC, de Lange FP (2012b) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197–2206. CrossRef Medline

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868. CrossRef Medline

Larsson J, Smith AT (2012) fMRI repetition suppression: neuronal adaptation or stimulus expectation? Cereb Cortex 22:567–576. CrossRef Medline

Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci U S A 108:19401–19406. CrossRef Medline

Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J Neurosci 13:1460–1478. Medline

Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. Neuron 63:879–888. CrossRef Medline

Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66:241–251. CrossRef Medline

Murray SO, Wojciulik E (2004) Attention increases neural selectivity in the human lateral occipital complex. Nat Neurosci 7:70–74. CrossRef Medline

Näätänen R, Paavilainen P, Alho K, Reinikainen K, Sams M (1987) The mismatch negativity to intensity changes in an auditory stimulus sequence. Electroencephalogr Clin Neurophysiol Suppl 40:125–131. Medline

Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532–552. CrossRef Medline

Rao RP (2005) Bayesian inference and attentional modulation in the visual cortex. Neuroreport 16:1843–1848. CrossRef Medline

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87. CrossRef Medline

Rao RP, Ballard DH (2005) Probabilistic models of attention based on iconic representations and predictive coding. In: Neurobiology of attention (Itti L, Rees G, Tsotsos JK, eds), pp 553–561. Waltham, MA: Elsevier Academic.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical conditioning II (Black AH, Prokasy WF, eds), pp 64–99. New York: Appleton-Century-Crofts.

Seidl KN, Peelen MV, Kastner S (2012) Neural evidence for distracter suppression during visual search in real-world scenes. J Neurosci 32:11812–11819. CrossRef Medline

Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. Neuroimage 44:223–231. CrossRef Medline

Summerfield C, Egner T (2009) Expectation (and attention) in visual cognition. Trends Cogn Sci 13:403–409. CrossRef Medline

Summerfield C, Egner T (2014) Attention and decision-making. In: The Oxford handbook of attention (Nobre A, Kastner S, eds), pp 837–864. Oxford: Oxford UP.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T (2008) Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004–1006. CrossRef Medline

Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK (2009) Neural evidence of statistical learning: efficient detection of visual regularities without awareness. J Cogn Neurosci 21:1934–1945. CrossRef Medline

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289. CrossRef Medline

Wyart V, Nobre AC, Summerfield C (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. Proc Natl Acad Sci U S A 109:3593–3598. CrossRef Medline

Yi DJ, Kelley TA, Marois R, Chun MM (2006) Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces. Brain Res 1080:53–62. CrossRef Medline

Yu AJ, Dayan P (2005) Inference, attention, and decision in a Bayesian neural architecture. In: Advances in neural information processing systems 17: proceedings of the 2004 conference (Saul LK, Weiss Y, Bottou L, eds), pp 1577–1584. Cambridge, MA: MIT.

Zhao J, Al-Aidroos N, Turk-Browne NB (2013) Attention is spontaneously biased toward regularities. Psychol Sci 24:667–677. CrossRef Medline