Behavioral/Cognitive

From Image Statistics to Scene Gist: Evoked Neural Activity Reveals Transition from Low-Level Natural Image Structure to Scene Category

Iris I.A. Groen,[1,2] **Sennay Ghebreab,**[2,3] **Hielke Prins,**[2] **Victor A.F. Lamme,**[1] **and H. Steven Scholte**[1,2]

[1]Cognitive Neuroscience Group, Department of Psychology, [2]Amsterdam Center for Brain and Cognition, Institute for Interdisciplinary Studies, and [3]Intelligent Systems Laboratory Amsterdam, Institute of Informatics, University of Amsterdam, 1018 WS, Amsterdam, The Netherlands

The visual system processes natural scenes in a split second. Part of this process is the extraction of "gist," a global first impression. It is unclear, however, how the human visual system computes this information. Here, we show that, when human observers categorize global information in real-world scenes, the brain exhibits strong sensitivity to low-level summary statistics. Subjects rated a specific instance of a global scene property, naturalness, for a large set of natural scenes while EEG was recorded. For each individual scene, we derived two physiologically plausible summary statistics by spatially pooling local contrast filter outputs: contrast energy (CE), indexing contrast strength, and spatial coherence (SC), indexing scene fragmentation. We show that behavioral performance is directly related to these statistics, with naturalness rating being influenced in particular by SC. At the neural level, both statistics parametrically modulated single-trial event-related potential amplitudes during an early, transient window (100–150 ms), but SC continued to influence activity levels later in time (up to 250 ms). In addition, the magnitude of neural activity that discriminated between man-made versus natural ratings of individual trials was related to SC, but not CE. These results suggest that global scene information may be computed by spatial pooling of responses from early visual areas (e.g., LGN or V1). The increased sensitivity over time to SC in particular, which reflects scene fragmentation, suggests that this statistic is actively exploited to estimate scene naturalness.

Introduction

The remarkable speed at which humans can perceive natural scenes has been studied for decades (Potter, 1975; Intraub, 1981; Thorpe et al., 1996; Fei-Fei et al., 2007). Many theories of visual processing propose that a global impression of the scene accompanies (Rousselet et al., 2005; Wolfe et al., 2011) or precedes (Hochstein and Ahissar, 2002) detailed feature extraction ("coarse-to-fine" processing) (Hegdé, 2008). This global percept is often described as visual *gist* (Torralba and Oliva, 2003; Oliva, 2005). Behavioral results confirm that global scene properties are indeed perceived very rapidly (Greene and Oliva, 2009a).

An important example of a global scene property is manifested in a difference between images with man-made versus natural content. Scene naturalness can be judged with less exposure time compared with other global properties (e.g., openness, depth) (Oliva and Torralba, 2001; Greene and Oliva, 2009b) or basic-level categories (e.g., mountain vs forest) (Rousselet et al., 2005; Joubert et al., 2007; Loschky and Larson, 2010; Kadar and Ben-Shahar, 2012). In computer vision, it was shown that this distinction coincides with low-level regularities of natural scenes (e.g., the distribution of spatial frequencies in an image) (Baddeley, 1997; Oliva et al., 1999), leading to the suggestion that humans might use such "image statistics" in rapid scene recognition (Schyns and Oliva, 1994; McCotter et al., 2005; Kaping et al., 2007). It is unclear, however, to which image statistics the brain is sensitive and how it uses them to extract scene naturalness: besides spatial frequency, color (Oliva and Schyns, 2000; Goffaux et al., 2005), and local edge alignment (Wichmann et al., 2006; Loschky et al., 2007) are examples of image features that may play a role.

We propose that, to elucidate the role of image statistics in scene perception and their neural representations, we need to address two challenges. First, a plausible computational mechanism by which image statistics are extracted by the brain must be specified. For example, estimating spatial frequency distributions using a global Fourier transformation is a biologically implausible mechanism, given that the visual system receives localized information from the retina (Graham, 1979; Field, 1987). Second, if image statistics are indeed used to extract global properties, they should predict not only general differences between scene categories but representations of individual images. This means that single-trial differences, in both neural responses and behavior, should correlate with differences in image statistics.

Here, we addressed these problems by combining computational modeling of physiologically plausible image statistics with

Received July 23, 2013; revised Oct. 7, 2013; accepted Oct. 24, 2013.

Author contributions: I.I.A.G., S.G., V.A.F.L., and H.S.S. designed research; I.I.A.G. and H.P. performed research; S.G. and H.S.S. contributed unpublished reagents/analytic tools; I.I.A.G. analyzed data; I.I.A.G., S.G., V.A.F.L., and H.S.S. wrote the paper.

This work is part of the Research Priority Program "Brain & Cognition" at the University of Amsterdam and was supported by an Advanced Investigator Grant from the European Research Council to V.A.F.L. and the Dutch national public-private research program COMMIT to S.G.

The authors declare no competing financial interests.

Correspondence should be addressed to Iris I.A. Groen, Nieuwe Achtergracht 129-131, 1018 WS, Amsterdam, The Netherlands. E-mail: i.i.a.groen@uva.nl.

DOI:10.1523/JNEUROSCI.3128-13.2013

 0270-6474/13/3318814-11$15.00/0

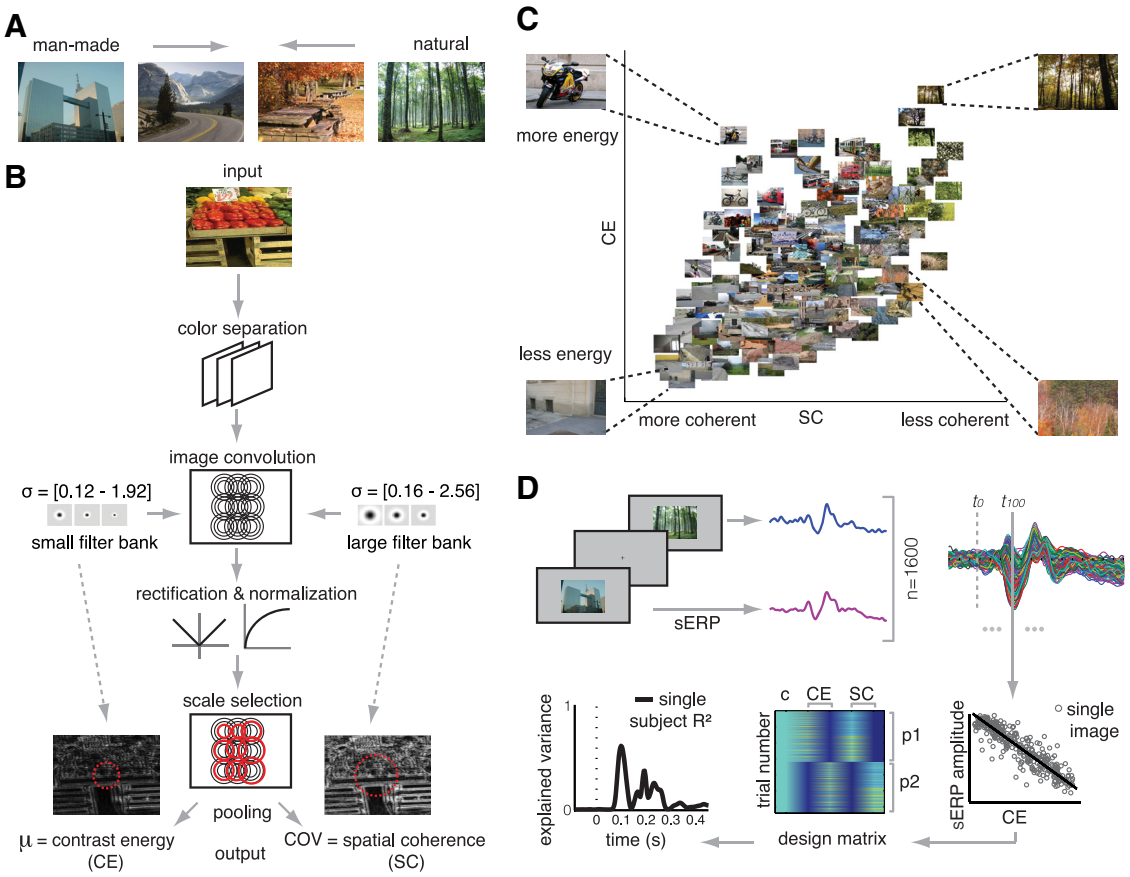

Figure 1. Stimuli, model, and methods. ***A***, Examples of images used in the experiment. Images varied considerably in the degree to which they contained exclusively man-made (left) or natural (right) elements, and the set included images for which the distinction may be unclear (middle). ***B***, A feedforward filtering model was used to derive two summary statistics: CE and SC. Three opponent (grayscale, blue-yellow, red-blue) contrast magnitude maps were computed by convolving the image with multiscale filters (black circles). For CE, a range of smaller filter sizes (σ = diameters in degrees) was used; for SC, a range of larger filter sizes. For each image location and parameter, a single filter response was selected (red circles) from each range using minimum reliable scale selection (see Materials and Methods). These responses were then pooled across a selection of the visual field (red dotted circles): for CE, the resulting responses were averaged; for SC, the coefficient of variation (COV) was computed. These values were averaged across the three color-opponent maps resulting in one CE and SC value per image. ***C***, A subset of 160 images (10% of the whole stimulus set, randomly selected) plotted against their CE and SC values. CE (the approximation of the β parameter of the Weibull function) describes the scale of the contrast distribution: it varies with the distribution of local contrasts strengths. SC (the approximation of the γ parameter of the Weibull function) describes the shape of the contrast distribution: it varies with the amount of scene fragmentation (scene clutter). Four representative pictures are shown in each corner of the parameter space. Images that are highly structured (e.g., a street corner) are found on the left, whereas highly cluttered images (e.g., a forest) are on the right. Images with higher figure-ground separation (depth) are located on the top, whereas flat images are found at the bottom. ***D***, sERPs to images presented at the center of the screen were computed for each subject. The resulting estimates of sERP amplitude were regressed on CE and SC at each time sample and electrode separately. The design matrix for the regression contained five columns: a constant term (c) for the intercept, two columns for CE and two for SC, each containing the same parameter values for the first and second image presentation (p1 and p2). These were modeled as separate predictors to examine reliability of the obtained effects across repetitions. The outcome of the analysis is a measure of model fit (explained variance or R^2) separately over subjects, time (samples), and space (electrodes); an example is shown for one subject at electrode Oz.

single-trial EEG measurements. We approximated image statistics by summarizing the outputs of model receptive fields (Torralba, 2003; Renninger and Malik, 2004; Ghebreab et al., 2009) using two parameters, contrast energy (CE), and spatial coherence (SC) that carry information about contrast strength and scene fragmentation (Scholte et al., 2009; Groen et al., 2012b). We tested whether these parameters were related to behavioral ratings of scene naturalness, and examined how they affected the time course of single-trial event-related potentials (sERPs) during naturalness categorization. Additionally, we examined how they affected discriminability of man-made versus natural ratings based on EEG activity.

Materials and Methods

Subjects

Sixteen subjects (age mean ± SD, 26 ± 6 years; 3 males) participated in the experiment, which was approved by the Ethical Committee of the Psychology Department at the University of Amsterdam. All participants gave written informed consent before participation and were rewarded with study credits or financial compensation (7 euro/h). Two subjects were excluded from analysis because their recordings were incomplete.

Visual stimuli

A stimulus set of 1600 color images (bit depth 24, JPG format, 640 × 480 pixels) was composed from several existing online databases. The set included images from a previous fMRI study on scene categorization (Walther et al., 2009), as well as images from various datasets used in computer vision: the INRIA holiday database (Jegou et al., 2008), the GRAZ dataset (Opelt et al., 2006), ImageNet (Deng et al., 2009), and the McGill Calibrated Color Image Database (Olmos and Kingdom, 2004). These different sources assured maximal variability of the stimulus set (Fig. 1*A*): it contained a wide variety of indoor and outdoor scenes, landscapes, forests, cities, villages, roads, images with and without animals, objects, and people. The images were selected such that one half of the set contained mostly man-made, and the other mostly natural elements.

Computational modeling

General approach. Natural images exhibit much statistical regularity. One instance of such regularity is present in the distribution of contrast

strengths, which ranges between power law and Gaussian and therefore conforms to a Weibull distribution (Simoncelli, 1999; Geusebroek and Smeulders, 2002). This regularity can thus be described by fitting a Weibull function to the contrast distribution, which yields two "summary" parameters that represent the scale (β) and shape (γ) of the distribution (Scholte et al., 2009). The β parameter varies with the range of contrast strengths present in the image, whereas γ varies with the degree of correlation between contrasts. In previous work, we found that these parameters can be approximated in a physiologically plausible way by summarizing responses of receptive field models to local contrast (Scholte et al., 2009). Specifically, summing the responses from a model of the two main parvocellular and magnocellular pathways in the LGN led to accurate approximations of β and γ values, respectively.

Here, we used an improved version of this previous model (Fig. 1*B*) where we refer to the approximated β value as CE and to the approximated γ value as SC. In this new model, two major changes have been introduced: (1) contrast is computed at multiple spatial scales; and (2) CE is estimated by averaging local contrast values, whereas SC is taken to be the coefficient of variation (mean divided by SD). Importantly, these changes have led to improved results in separate EEG datasets from the one reported here, with different images and different EEG recordings. Specifically, using CE and SC led to an increase in effect size in EEG (see below) by ~20% (Scholte et al., 2009 vs Ghebreab et al., 2009); and the CE and SC derived with the present model correlated most strongly with fitted Weibull parameters (H.S.S. et al., unpublished observations). In the present dataset, the approximations (CE and SC) correlated with the fitted parameters (β and γ) with $r = 0.95$ and $r = 0.73$, respectively. The computations involved in deriving CE and SC are described below.

Step 1: Local contrast detection. Contrast filtering was done separately for the three image color layers, which were first converted to opponent color space (grayscale, blue-yellow, red-green) (Koenderink et al., 1972). We computed image contrast by convolving each image layer with exponential filters (Zhu and Mumford, 1997) at five octave scales (Croner and Kaplan, 1995). Two separate filter sets were used: one with slightly smaller filter sizes (0.12, 0.24, 0.48, 0.96, and 1.92 degrees) for CE and a range of larger filter sizes (0.16, 0.32, 0.64, 1.28, and 2.56 degrees) for SC (Ghebreab et al., 2009). Following the LGN suppressive field approach (Bonin et al., 2005), all filter responses were rectified and divisively normalized to account for nonlinear neuronal properties.

Step 2: Scale selection. Per parameter (CE or SC), one filter response for each image location was selected from their specific filter set using minimum reliable scale selection (Elder and Zucker, 1998). In this MIN approach, the smallest filter size that yields an above-threshold response is preferred over other filter sizes. Filter-specific noise thresholds were determined in a separate image set (Corel database) (Ghebreab et al., 2009). This step was again performed separately for the three color-opponent layers.

Step 3: Pooling responses. Applying the selected filters per image location results in two contrast magnitude maps: one highlighting detailed (from the set of smaller filter sizes, for CE) and the other more coarse edges (from the set of larger filter sizes, for SC). For the pooling step, a different amount of visual space was taken into account for each parameter. For CE, the central 1.5 degrees of the visual field was used, whereas for SC, 5 degrees of visual field was used. Again, these settings were chosen because they yielded the best model fits for regression analyses in separate, independent EEG datasets (Ghebreab et al., 2009, Scholte et al., 2009; S.G. et al., unpublished observations). Finally, parameter values were averaged across color-opponent layers resulting in a single CE and SC value per image (Fig. 1*C*).

Other image statistics For comparison of the results with previous findings of sensitivity of human observers to spatial frequency distributions (e.g., Kaping et al., 2007), Fourier amplitude statistics were computed using the procedure described by Oliva and Torralba (2001) (i.e., fitting a line to the rotationally averaged power spectrum). This procedure yields one Fourier intercept (Fi) and Fourier slope (Fs) value per image (Groen et al., 2012a).

Experimental design

Experimental procedure. Subjects completed two EEG recording sessions on 2 consecutive days: in each session, 1600 images were shown. Every image was repeated once. Images were presented on a 19 inch Ilyama CRT monitor (1024 × 768 pixels, frame rate 60 Hz). Subjects were seated 90 cm from the monitor such that stimuli subtended ~14 × 10° of visual angle. On each trial, one image was randomly selected and presented in the center of the screen on a gray background for 100 ms, on average every 1500 ms (range 1000–2000 ms). Subjects were instructed to indicate, as quickly as possible, whether the image was man-made ("made by humans") or natural ("not made by humans") using two button boxes, one for each index finger. Response mappings were counterbalanced across subjects. To familiarize subjects with stimulus presentation, 30 practice images (never used in the main experiment) were presented without feedback before the first session. Each session was divided in 4 runs, which were subdivided in 5 self-paced mini-blocks: subjects were encouraged to take breaks between blocks. Stimuli were presented using the Presentation software (www.neurobs.com).

Behavioral data analysis. Trials at which the subject failed to respond within 200–1200 ms after stimulus onset were discarded from analysis (median rejection rate 0.1%, minimum 0%, maximum 4%). For each of the 1600 images, we computed the following indices: (1) naturalness rating (i.e., the average behavioral responses across subjects and repetitions), with 0 indicating that none of the participants rated the scene as natural and 1 indicating that all participants rated the scene as natural; (2) subject-specific naturalness rating (i.e., the average response across the two repetitions), with 0 indicating a man-made response and 1 indicating a natural response; (3) average reaction time (RT) across subjects and repetitions; and (4) subject-specific RT. These indices were correlated (Spearman's ρ) with CE and SC, as well as the Fourier slope and intercept values; the resulting p values were FDR-corrected at $\alpha = 0.05$ for the total number of correlations computed. Finally, we also computed for each image a "reliability index" by comparing the subject-specific naturalness ratings across the two presentations of the image. If the rating was the same across repetitions, reliability was coded as 1, whereas it was coded as 0 if the rating was different. Averaging these numbers across subjects yielded an estimate of reliability for each image.

EEG data acquisition and preprocessing. EEG recordings were made with a Biosemi 64-channel Active Two EEG system (Biosemi Instrumentation; www.biosemi.com). Recording set-up and preprocessing were identical to procedures described previously (Groen et al., 2012a,b). We used caps with an extended 10–20 layout, modified with two additional occipital electrodes (I1 and I2, while removing electrodes F5 and F6). Eye movements were monitored with electro-oculograms (EOGs). Recording was followed by offline referencing to external electrodes placed on the earlobes. Preprocessing occurred in Brain Vision Analyzer and consisted of: a high-pass filter at 0.1 Hz (12 dB/octave); a low-pass filter at 30 Hz (24 dB/octave); two notch filters at 50 and 60 Hz; automatic removal of deflections >300 mV; epoch segmentation in −100 to 500 ms from stimulus onset; ocular correction using the EOG electrodes (Gratton et al., 1983); baseline correction between −100 and 0 ms; automated artifact rejection (maximal voltage steps 50 μV, minimal/maximal amplitudes −75/75 μV, lowest activity 0.50 μV); and, finally, conversion to Current Source Density responses (Perrin, 1989). Median rejection rate was 203 of 3200 trials (mean 7%, minimum 1%, maximum 20%). We also removed, per subject, trials that were excluded based on behavior (see above), which increased the median rejection rate to 8%; overall, 9% of the total amount of data was removed. No trial or electrode averaging was performed: preprocessing thus resulted in an sERP specific to each subject, electrode, and individual image presentation.

Regression on single-trial ERPs. To test whether differences between sERPs were modulated by differences in CE and SC, the sERPs were read into MATLAB (MathWorks), where we conducted linear regression of sERP amplitude on image parameters (Fig. 1*D*) using the Statistics Toolbox. For each subject, each electrode, and each time point, the two parameters were entered as predictor variables, and the z-scored sERP amplitudes as observations in the regression model. The two presentations of each image were modeled as two separate predictor columns, to examine effects of repetition; the values in each column were z-scored

independently (Rousselet et al., 2011). This analysis thus results in a measure of model fit (explained variance or R^2) for each subject, electrode, and time point separately. In addition, we read out from the regression results the β-coefficients assigned to each predictor column, which reflect the regression weights between the image parameters and sERP amplitude (again, separately for each subject, electrode, and time point). To assess the significance of these coefficients, we tested the subject-averaged coefficient against zero using t tests (separate for each time point and electrode): a significant test result indicates that the association between a predictor and evoked neural activity is reliably larger than zero at a specific time point and electrode. To correct for multiple comparisons (time points, electrodes and parameters), p values were FDR-corrected at $\alpha = 0.05$. The rationale of this approach is quite similar to conventional fMRI analysis (GLM); we compute the β-coefficients for each predictor in our model and threshold using multiple-comparison corrected t values. By doing this for every electrode ("whole-scalp analysis"), we created spatial maps that indicate which electrodes contain significant β-coefficients.

Regression results for other image statistics. For comparison of the regression results for CE and SC with parameters obtained from the Fourier transform, we ran the regression analyses described above (Fig 1*D*) while replacing the predictor columns with Fi and Fs. In addition, we examined the dependence between the two models by running a third regression analysis in which we entered all four parameters (CE, SC, Fi, and Fs) together as predictors, yielding a "full" model. By subtracting the R^2 values for the first two models from the full model, we obtained a measure of "unique explained variance" by each model (Groen et al., 2012b). These values were again obtained separately per subject, electrode, and time point.

Linear discriminant component analysis. To identify decision-related activity in the EEG signal, we used linear discrimination analysis as developed by Philiastides and colleagues (Parra et al., 2002; Philiastides and Sajda, 2006; Philiastides et al., 2006). Linear discriminant analysis performs logistic regression of binary data on multivariate EEG data to identify spatial weighting vectors (w) across electrodes that maximally discriminate between conditions of interest (e.g., a face or car stimulus). Here, we used as conditions of interest the ratings made by the subjects during naturalness rating, by entering the trial-specific behavioral responses (man-made or natural) as the binary data. The analysis outcome is a "discriminating component" y, which is specific to activity correlated with one condition while minimizing activity correlated with both task conditions (Philiastides et al., 2006); in our case, we isolated activity specific to the decision that a scene was natural.

We used an online available regularized logistic regression algorithm (Conroy and Sajda, 2012) that allows for fast estimation of the discriminating components. We tested various regularization terms ($\lambda = 1e^x$ with $x = -6{:}1{:}2$), which yielded similar results to those reported here for $\lambda = 1e^{-1}$. We computed y for a number of temporal windows (window size $\delta = 20$ ms), to estimate the temporal evolution of discriminant activity over the course of the ERP. Per time window, discriminator performance (Az) across all trials was quantified using the area under the receiver-operator characteristic curve, with a leave-one-out cross-validation approach (Duda et al., 2001). Following Blank et al. (2013), significance of performance was evaluated by bootstrapping the Az values. We used 100 bootstraps per time window (epoch of 600 ms/$\delta = 40$ windows) and subject ($n = 14$), resulting in 42,000 Az values in total. The overall distribution of these values was used to determine the Az value leading to a significance level of $p = 0.05$. Finally, the components were projected back on the scalp using a forward model that multiplies the w-vector with average ERP amplitude in a specific time window (Parra et al., 2002).

Relating discriminant component activity to image statistics. After obtaining an estimate of discriminant component activity at each trial and time window, the y values were correlated (Spearman's ρ) with the CE and SC values of the image presented at that trial, for each subject separately. To assess the significance of these correlations, we again tested the average correlation across subjects against zero using separate t tests for each time window; the resulting p values were again corrected for multiple comparisons using an FDR-correction at $\alpha = 0.05$.

Results

Behavior

Fourteen subjects rated 1600 scene images as either man-made or natural while EEG was recorded. We first examined how behavioral ratings were distributed across the entire set of images. Next, we tested how these results were related to differences in CE and SC estimated from modeled responses to local contrast.

Naturalness rating and RT for individual images

On average, subjects rated 49.7% of the trials as man-made and 50.3% as natural (SD = 3.3% for both man-made and natural). Reaction times for man-made versus natural responses did not differ significantly (mean $RT_{man\text{-}made}$ = 523 ms, SD = 67 ms; mean $RT_{natural}$ = 527 ms, SD = 79 ms; median $RT_{man\text{-}made}$ = 497 ms, SD = 64 ms, median $RT_{natural}$ = 504 ms, SD = 76 ms; all $t_{(13)} < 1$, all $p > 0.56$), showing that subjects did not have a bias toward one particular response. There was, however, considerable variability in ratings across trials, both within subjects and across subjects.

Within subjects, on average 10% of the trials were rated differently on the second presentation than the first (minimum 3%, maximum 18%, SD 4%). For these trials, reaction times were longer than for trials that were rated the same (mean RT_{same} = 520 ms, SD = 70 ms, mean $RT_{different}$ = 571 ms, SD = 90 ms; paired t test $t_{(13)} = 6.1$, $p = 0.00003$; median RT_{same} = 505 ms, SD = 68 ms, median $RT_{different}$ = 557 ms, SD = 93 ms; $t_{(13)} = 5.5$, $p = 0.0001$). Across subjects, the scenes could be subdivided evenly in three bins based on differences in naturalness rating (Fig. 2*A*): scenes that had a rating <0.1 (indicating they were rated as natural by <10% of the subjects), scenes with a rating >0.9, and scenes for which the rating was intermediate (0.1–0.9 ratings). Average RT for the intermediately rated images was higher compared with consistently rated images (repeated-measures ANOVA, $F_{(2,13)} = 42.6$, $p = 0.0001$); this bin also more often contained scenes for which ratings differed between the repetitions (repeated-measures ANOVA on within-subject consistency, $F_{(2,13)} = 138.6$, $p = 0.0001$; Fig. 2*B*).

In sum, although subject's behavior was on average similar for the two scene categories, the increase in variability in rating both within and across subjects for a specific subset of the images shows that some scenes were experienced as more "ambiguous" than others. This is not surprising as the stimulus set was purposely composed to span a wide range of natural images (see Materials and Methods). Next, we examined whether this pattern of results could be explained based on differences in image statistics between the scenes.

CE and SC predict behavioral performance

The CE and SC values for each image in the stimulus set are shown in Figure 2*C*, with color-coding indicating the average naturalness rating per image. Binning images on either CE or SC (Fig. 2*C*, side histograms) shows that naturalness rating is related to differences in SC: the higher the SC value of the scene, the higher the number of subjects that rated the scene as natural. A higher CE value is associated with shorter RTs. Figure 2*C* also shows that the SC values give rise to a continuous space, rather than two discrete categories. Interestingly, variability in image ratings is related to the SC value of a scene: images with intermediate ratings are found at intermediate SC values (Fig. 2*D*; independent-samples Kruskal–Wallis test, $\chi^2(2) = 131.3$, $p =$ 3e-29), whereas there is no such effect for CE ($\chi^2(2) = 2.4$, $p =$ 0.29). As can be seen in Figure 2*E*, images with intermediate

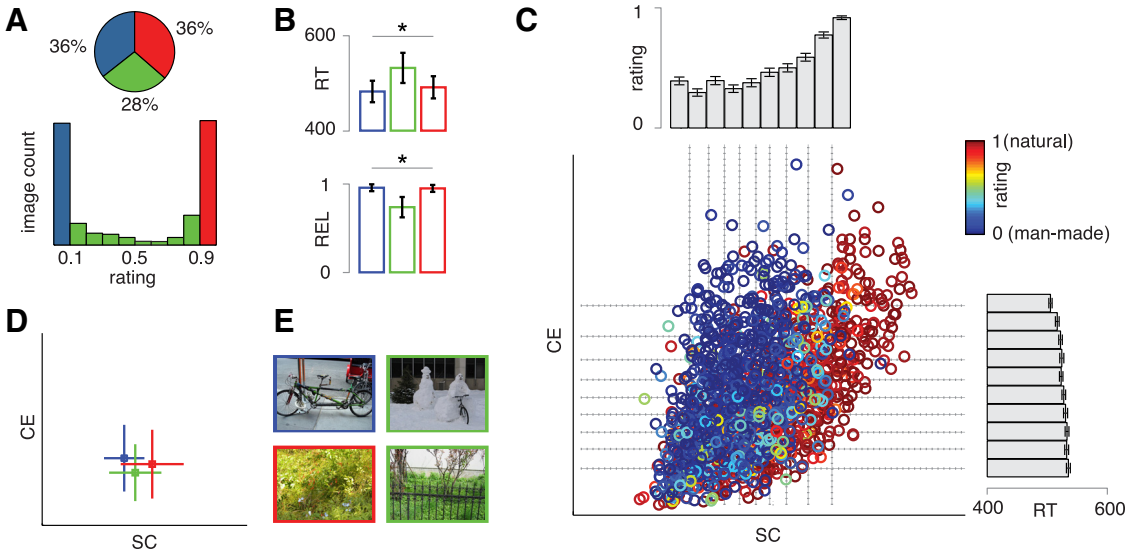

Figure 2. Behavioral results. ***A***, Distribution of naturalness ratings for the entire stimulus set. Approximately two-thirds of the scenes had a high man-made (<0.1, blue bin) or high natural rating (>0.9, red bin), whereas the remaining images had intermediate ratings (green bin, 0.1–0.9). ***B***, RTs (top) and within-subject repeat reliability (REL; bottom) for the three different bins of trials. *Significant main effects of bin ($p < 0.0001$). Error bars indicate SD. ***C***, CE and SC values plotted for all 1600 images, color-coded by naturalness rating. The histograms display 10% bins based on either CE or SC, containing the average rating obtained after sorting on SC (top histogram, *x*-axis), or the average RT after sorting on CE (side histogram, *y*-axis). Error bars indicate SEM. ***D***, Intermediately rated images (***A***, green bin) are found at intermediate SC values, whereas images with high man-made (blue) or natural (red) ratings have more extreme SC values. Squares represent medians; lines represent SDs. ***E***, Example images from each bin. Whereas the highly man-made or natural rated images (left) contain either exclusively man-made or exclusively natural elements, one intermediate image (top right) contains a building and a bicycle (man-made) as well as snowy objects and a bush (natural); another (bottom right) has shrubbery (natural) and a fence and wall (man-made).

Table 1. Correlations of image parameters with naturalness rating and RTs[a]

	Naturalness rating			RT		
Image parameter	ρ	*p*	% of subjects (FDR-corrected)	ρ	*p*	% of subjects (FDR-corrected)
CE	−0.08	0.0007	71%	−0.2	2.9e-16	86%
SC	0.39	2.3e-58	100%	−0.01	NS	43%
Fs	−0.19	1.3e-15	100%	−0.18	2.4e-13	78%
Fi	−0.36	2.5e-50	100%	−0.09	6.7e-5	71%

[a]Reported are Spearman correlation values (ρ), corresponding significance values (*p*), and the percentage of subjects whose individual rating or RT was significantly correlated with the image parameters. NS, Not significant.

ratings and SC values typically contain patches with both man-made and natural elements. This suggests that SC in particular is diagnostic of scene naturalness.

We confirmed these observations by correlating the CE and SC values directly with the behavioral measures (Table 1). SC correlated with naturalness rating, but not with RT, whereas CE correlated with RT, and to a lesser extent with rating. For comparison, we also computed correlations with image statistics derived from the power spectra of the scenes (Oliva and Torralba, 2001), Fi and Fs (see Materials and Methods). These correlations are in a similar range as those with CE and SC (Table 1). For this model, however, the two parameters are less well dissociable, as they both correlate significantly with accuracy as well as RT.

The behavioral results support our proposal that CE and SC, which are computed by pooling of local contrast, capture global scene information. They predict human categorization of scene naturalness to a similar degree as statistics obtained from a global image transformation (Fourier parameters). In particular, there is a relation between SC and perceived naturalness, whereas CE seems to influence the speed of processing.

EEG

sERPs were extracted from the continuous EEG data after which we performed linear regression of sERP amplitude on image statistics. We first examined how CE and SC affected evoked activity to individual images, again comparing them with the Fourier parameters. Next, we tested how the EEG activity itself was related to the behavioral ratings, and whether this relation was dependent on CE and/or SC.

Explained variance by CE and SC

Regression analysis of sERP amplitude on CE and SC values revealed a strong relation between these parameters and evoked neural activity. For all subjects, explained variance was maximal at either Oz or Iz (the occipital mid-line electrodes), ranging between $R^2 = 0.16$ and 0.46 at 109–137 ms (for all subjects, $p <$ 1e-10, FDR-corrected), whereas mean explained variance over subjects was highest at Oz (maximal $R^2 = 0.27$, 117 ms after stimulus onset; Fig. 3*A*). Explained variance at this time point also extended to other electrodes (Fig. 3*B*); maximal R^2 values were reliable in all 14 subjects at several occipital (I1, O1, O2) and parietal electrodes (P1, P3, P4, P6, P8, P10). These results replicate earlier findings (Scholte et al., 2009) of ERP sensitivity to statistics derived from local contrast in natural images from ~100 ms after stimulus onset. They show that differences between individual scenes in evoked neural activity are reliably correlated with differences in CE and SC of the scenes.

Regression weights for individual parameters

Given the dissociable effects of CE and SC on behavior, we next tested whether these parameters also differentially influenced sERP amplitude by examining their individual β-coefficients (see Materials and Methods). At electrode Oz (Fig. 3*C*), the β-coefficients for CE were largest during an extended early interval (94–156 ms). In contrast, the largest β-coefficients for SC were found at a nonoverlapping, later interval (160–195 ms). Regression weights associated with each predictor were highly reliable across repetitions. The difference between the CE and SC

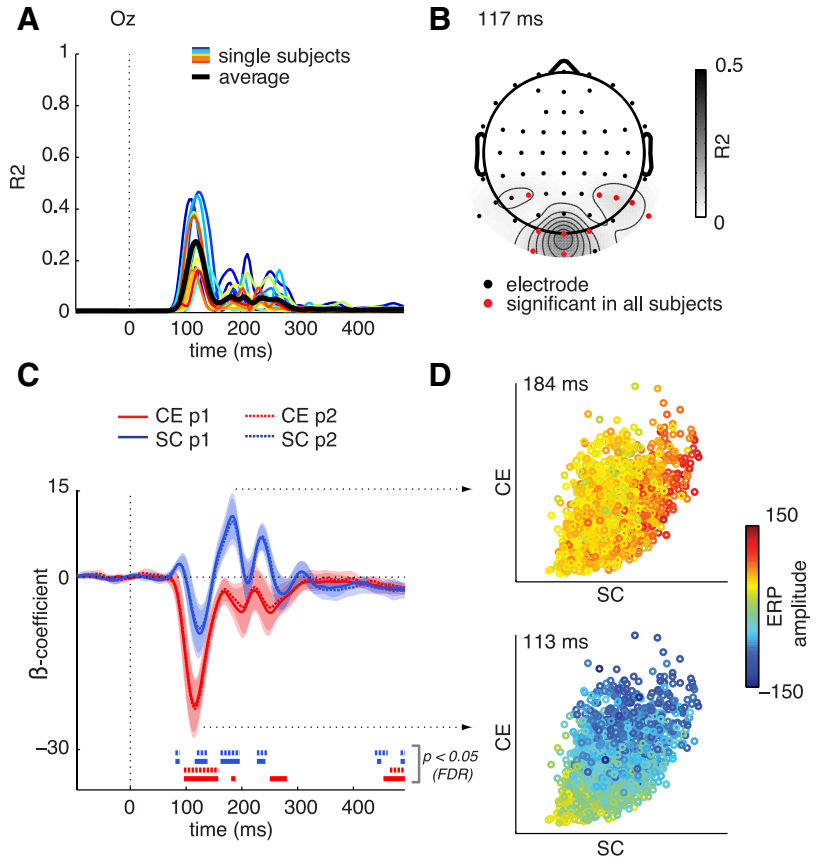

Figure 3. EEG regression results at occipital electrode Oz. ***A***, Explained variance for single subjects (colored lines) and mean R^2 across subjects (thick black line) over time. ***B***, Distribution of mean R^2 values across the scalp at the time point of maximal mean R^2 (117 ms). Black dots represent electrode locations; for red colored electrodes, R^2 values were significant in all participants (FDR-corrected). ***C***, Regression weights (β-coefficients) for each predictor column of the regression model, revealing several significant time windows (bottom bars): β-coefficients for CE are maximal early in time, whereas β-coefficients for SC reach a maximum later in time. The shading indicates 95% confidence intervals as obtained from a *t* test against zero. ***D***, Subject- and repeat-averaged ERP amplitude (color-coded) for each image plotted against its CE and SC values at the time points of maximal regression weights.

effects is illustrated in Figure 3*D*, where sERP amplitude for each image relative to its SC and CE value is shown for the two time points with largest β-coefficients (113 and 184 ms, respectively). At the first maximum, differences in amplitude are mostly aligned with the CE-axis of the space. At the second maximum, the amplitudes align with the SC-axis instead. Significant but smaller coefficients for SC were also found at two shorter and earlier intervals (78–85 ms and 117–136 ms) and one later interval (226–242 ms). Neural sensitivity to different image parameters thus differs substantially over the time course of visual processing. Interestingly, these results again demonstrate dissociable effects of CE and SC, but now on evoked neural activity.

Finally, for both CE and SC, regression coefficients were also significant toward the end of the ERP epoch, most likely reflecting spill-over of effects at other electrodes. These might be electrodes near motor cortex, as this interval is in the RT range (minimal average RT was 436 ms, maximal average RT was 647 ms); the β-coefficients at Oz are very small at this time point, indicating low sERP amplitudes. In the next section, we test to what extent each image parameter affected responses at other electrodes.

Whole-scalp regression weights

The distribution of regression coefficients across the whole scalp mirrors the effects observed at electrode Oz. Generally, the β-coefficients of CE are largest early in time, at occipital electrodes overlying early visual cortex. For SC, on the other hand, the β-coefficients are largest later in time and more sustained at perioccipital electrodes that are near mid- or higher-level visual areas (Fig. 4*A*). Coefficients of CE are virtually absent after 150 ms, whereas the coefficients of SC remain significant for an extended period across multiple electrodes. This difference is illustrated in Figure 4*B*, where average sERP amplitude is shown for bins of trials that are sorted based on either CE or SC. At occipital electrodes O1 and O2, the modulation by CE is clearly visible in the early interval, whereas a sustained modulation by SC is visible at perioccipital visual electrodes PO7 and PO8. The distinct effects of CE and SC are thus even more evident at the whole-scalp level, with differences in CE giving rise to a transient, early modulation, whereas SC differences lead to more widespread effects.

Comparison with Fourier parameters

The regression results demonstrate strong effects of CE and SC on evoked neural activity during categorization of naturalness. However, naturalness ratings did not only correlate with these image statistics estimated from local contrast but also correlated significantly with Fourier parameters obtained from a global Fourier transform. Do these latter parameters affect evoked activity in a similar way as CE and SC? To test this, we repeated the EEG regression analyses using the Fi and Fs parameters.

Explained variance of the Fourier parameters has a similar time course as for CE and SC, but with, on average, lower values (Fig. 5*A*): the average maximal value was 17% (range 11%–29% for single subjects) at electrode Oz at 117 ms, the same electrode and time point as for CE and SC. The spatial extent of the explained variance is also similar to that observed for CE and SC (Fig. 5*B*). To compare the two models directly, we estimated the unique explained variance by each pair of parameters by subtracting the R^2 values obtained for each model from a full model containing all parameters (see Materials and Methods). Unique explained variance for CE and SC reached a maximum of 11% at 113 ms, whereas the values were much lower for the Fourier parameters (2% at 184 ms; Fig. 5*C*). For both models, the maxima of unique explained variance were again at Oz, and the whole-scalp results (Fig. 5*D*) show that the unique explained variance for Fourier was indeed minimal across all electrodes, ruling out the possibility that Fourier parameters influence neural activity at other brain sites than CE and SC do. In previous findings with naturalistic image categories (Groen et al., 2012a), the Fourier parameters also explained ~10% variance less than two parameters derived from a Weibull fit to the contrast distribution (of which CE and SC are approximations, see Materials and Methods), as well as very little unique variance.

These results suggest that modulations of single-trial evoked activity during naturalness categorization are also correlated with differences in spatial frequency content (Fi and Fs). However, the effects were weaker compared with those observed for CE and SC, and the Fourier parameters did not explain substantial variance above and beyond the variance explained by CE and SC. This suggests that information that is contained in the Fourier parameters is also captured by CE and SC and that the latter may provide a more plausible description of evoked neural activity during naturalness categorization.

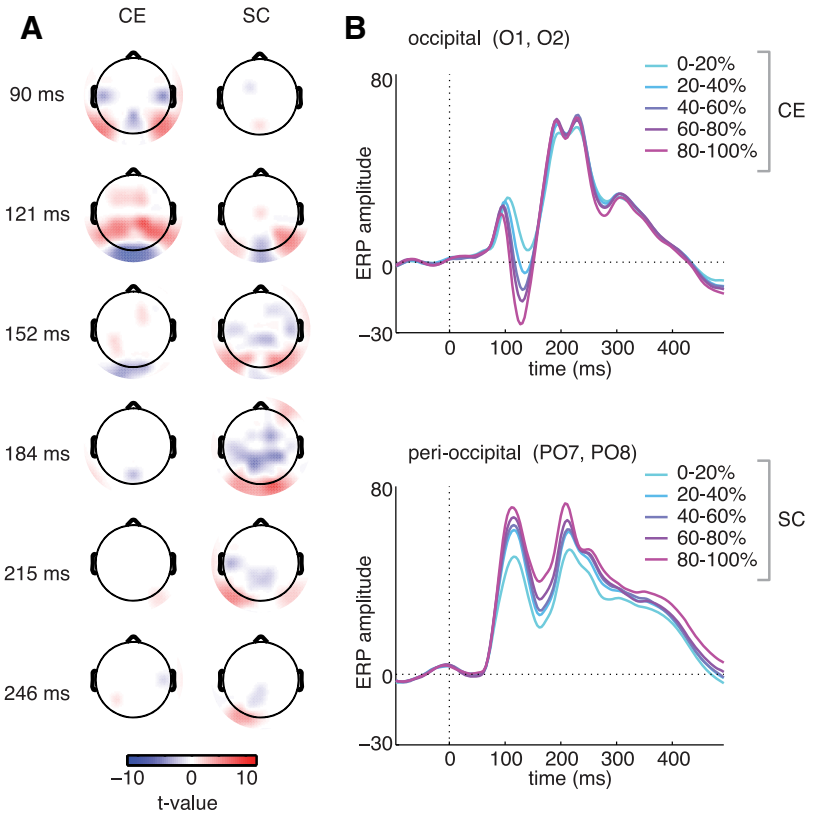

Figure 4. Whole-scalp regression results. ***A***, Topographical representations of *t* values (FDR-corrected and thresholded; corrected $t = 3.22$ for $\alpha = 0.05$) for β-coefficients at each electrode, reflecting reliable effects of CE and SC on sERP amplitude. The two parameters have different effects: CE influences ERP amplitude early in time, mostly at occipital electrodes, whereas SC gives rise to a sustained correlation later in time (>150 ms) associated with activity surrounding mid-level visual, peri-occipital electrodes. ***B***, Subject-averaged ERP amplitude of trials that were sorted and binned in 20% bins based on either CE or SC values, displayed for two different electrode poolings: occipital (O1, O2) and perioccipital (PO7 and PO8).

Role of image statistics in perceptual decision-making?

The regression results suggest that CE and SC play different roles in visual processing of natural images, with CE giving rise to early, transient effects, whereas SC differences lead to sustained effects on evoked activity. Interestingly, the modulations by SC extend into time intervals associated with mid-level stages of visual processing (beyond 200 ms) that are sensitive to top-down influences (Luck et al., 2000; Scholte et al., 2006) and possibly involved in perceptual decision-making (Philiastides et al., 2006). Task relevance of the scene parameters may thus play a role in the differential effects of CE and SC that were observed here. Specifically, the late modulation by SC could reflect an influence of SC on the naturalness decision.

To test how CE and SC affected decision-related neural activity, we used linear discriminant analysis to identify discriminating components in the EEG (Parra et al., 2002; Philiastides and Sajda, 2006; Philiastides et al., 2006; Conroy and Sajda, 2012; Blank et al., 2013). This is essentially again a single-trial regression, but of behavior onto the EEG: observations correspond to the (subject-specific) man-made or natural ratings, and the predictor variables consist of the sERP amplitudes at each electrode. This analysis yields two measures: overall discrimination accuracy, summarized in parameter Az and trial-specific discriminant components *y*, reflecting evidence toward a natural versus a man-made rating at a given trial (see Materials and Methods). Both measures were determined for consecutive time windows (see Materials and Methods), to examine the development of discriminant activity over time. With this analysis, we aimed to address two questions: (1) From what point in time can we reliably predict whether a given trial will be rated as man-made or natural? (2) To what extent is the strength of evidence toward the man-made or the natural rating modulated by CE and SC?

Discriminating man-made versus natural ratings based on EEG

First, response discrimination accuracy (Az) based on EEG was significantly above chance starting from a time window between 80 and 100 ms (Fig. 6*A*). It reached a local maximum between 180 and 200 ms (but remained significant) and started to rise again from 260 ms onwards. Projections of the discriminant activity back on the scalp (insets in Fig. 6*A*) show that, for the first two of these windows, activity was located at occipital/perioccipital sites, whereas the activity in the third window was more distributed across the scalp. Importantly, the early maximum in discrimination accuracy was found at the same moment in time at which SC most strongly affected sERP amplitude (Fig. 3*C*).

Second, discriminant components (*y*) were significantly correlated with SC, from 120 ms onwards (maximal mean $\rho = 0.18$, 180–200 ms, $t_{(13)} = 9.1$, $p = 5.1\text{e-}7$; Fig. 6*B*). For CE, correlations were significant for just one early window (80–100 ms, mean $\rho = -0.06$, $t_{(13)} = -3.7$, $p = 0.002$), and again much later in time, from 360 ms onwards. Note that, however, for this early window, discrimination accuracy is not yet significant (Fig 6*A*). Fig. 6*C* shows discriminant activity for each image sorted based on either CE or SC. These component maps reveal that images with high SC values have strong evidence toward the natural response (red), whereas images with low SC values are mapped toward man-made responses (blue). No such effects are visible when sorting on CE.

These results show that activity discriminating between man-made versus natural responses is present in the EEG from as early as 100 ms. Importantly, the strength of this activity at the single-trial level is related to the SC, but not CE, value of the image. Overall, the EEG results reveal strong neural sensitivity to image statistics derived from receptive field model responses to local contrast (CE and SC). Neural sensitivity was stronger for CE and SC than for spatial frequency parameters derived by means of a

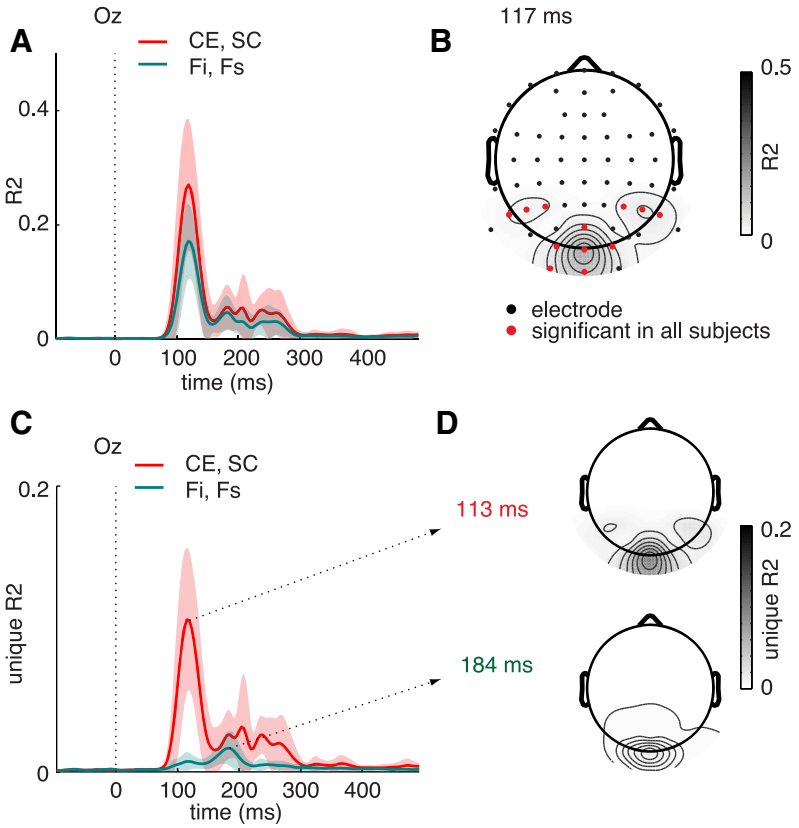

Figure 5. Comparison of results with regression analysis on Fourier statistics. ***A***, Explained variance averaged across single subjects for regression of single-trial ERP amplitude on Fi and Fs (green). For comparison, the average R^2 for CE and SC is plotted again as well (red). Shading indicates SD across subjects. ***B***, Distribution of average R^2 values for the Fourier parameters across the scalp at the time point of maximal R^2 (117 ms). Reliable electrodes across subjects (red) were as follows: I1, I2, O1, O2, Oz, P3, P4, P5, P6, P7, P8, and POz. ***C***, Unique explained variance for CE and SC (red) compared with Fi and Fs (green), averaged over subjects. Shading indicates SD across subjects. ***D***, Scalp plots of unique R^2 values for each model at their respective maximal time points: for CE and SC, this was at 117 ms (top topographical plot) and at 184 ms for the Fourier parameters (bottom).

global Fourier transformation. We found dissociable effects of CE and SC on neural activity, both in time (early and late intervals) and space (occipital vs perioccipital electrodes). This suggests that CE mainly affects activity in early visual areas involved in encoding of the stimulus, whereas SC also modulates subsequent decision-related activity, which likely involves more visual areas and processing time. The finding that SC, but not CE, is correlated with EEG components that maximally discriminate between the subject's behavioral ratings further supports this dissociation.

Discussion

A computational substrate for scene gist perception

How does the visual system estimate scene gist? We find that, for at least one global scene property (naturalness), single-trial differences in behavior and neural activity are related to differences in image statistics derived by integrating local contrast responses. The modulation of neural activity by these statistics as soon as ~100 ms after stimulus onset confirms that this information is available early in processing. However, the sustained modulations at later time points reveal a shift in neural sensitivity from CE to SC, suggesting a transformation toward coding of more relevant information for estimating naturalness, which appears to be carried by SC.

These results verify our previous observations of extensive ERP sensitivity to image statistics. We showed that differences in statistics explain a large amount of variance in ERP amplitude (Scholte et al., 2009), to the extent that they can be used to classify which natural image was viewed (Ghebreab et al., 2009). Perceived similarity of textures and naturalistic images is also related to CE and SC (Groen et al., 2012a,b). Here, we extend these results to scene gist, by linking single-trial differences in global property categorization and neural activity to differences in image statistics.

Model comparison with spatial frequency distributions

The role of spatial frequency in visual processing has been studied at multiple levels, ranging from local receptive field tuning (De Valois and De Valois, 1990) to the entire image: for example, the "spectral signature" of a scene, reflecting the decay of the 2D power spectrum, can be used to computationally discriminate global scene properties (Torralba and Oliva, 2003). Whereas the power spectrum reflects the distribution of overall energy across spatial scales ("amplitude"), another type of information is local alignment of spatial frequencies ("phase coherence"), which can be quantified from the shapes of contrast distributions (Tadmor and Tolhurst, 2000). It is currently debated which of these two sources (amplitude vs phase) is more important for scene discrimination (McCotter et al., 2005; Einhäuser et al., 2006; Loschky et al., 2007; Loschky and Larson, 2008).

Our model extracts phase information based on filters modeled after receptive fields (Scholte et al., 2009). However, phase information is inferred from contrast computed at and selected locally from multiple spatial scales (Elder and Zucker, 1998; Ghebreab et al., 2009), and the information is then spatially integrated across the entire scene. This procedure thus likely captures both phase and some amplitude information, and we have shown previously that our model outperforms separate descriptions of amplitude and phase (Groen et al., 2012a). Here, we also observed that CE and SC explained the neural data better than power spectra alone. This is also consistent with the observation that coarse localization of spectral signatures (Torralba and Oliva, 2003) improves computational discrimination of global properties. Our results thus agree with the idea that both phase and amplitude play a role in scene perception (Gaspar and Rousselet, 2009; Joubert et al., 2009).

It is important to emphasize, however, that CE and SC are derived in a very different way compared with traditional measures of phase and amplitude. The latter are obtained by performing a global Fourier transformation, whereas our model integrates locally filtered information using averaging and division, which can easily be implemented in a spiking neural network (e.g., using "integrate-and-fire" rules). We thus also attribute the fact that CE and SC provided a better description of neural activity to the physiological plausibility

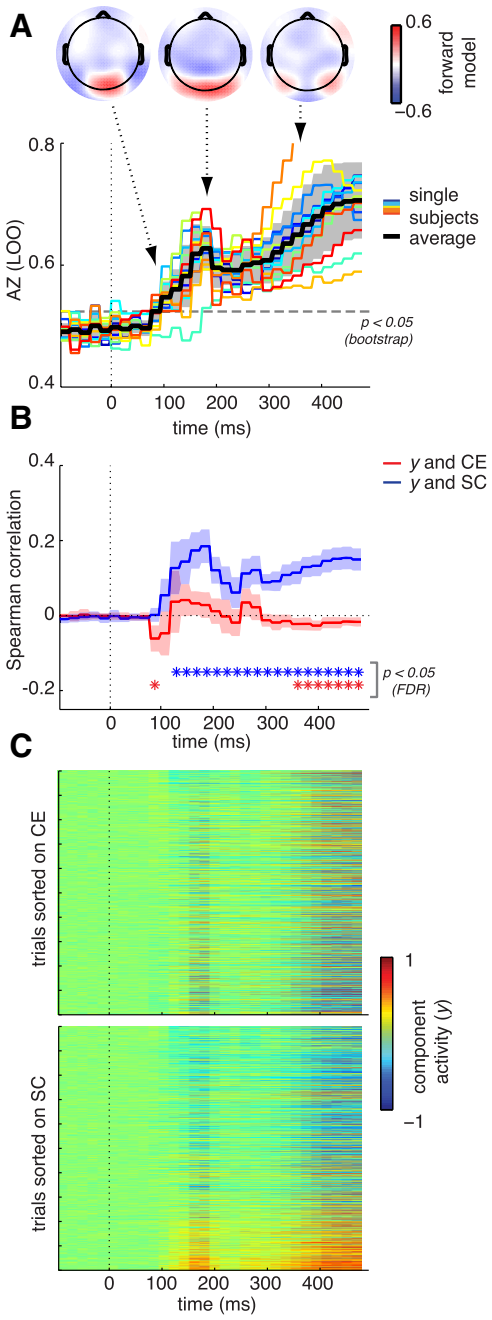

Figure 6. Discriminant analysis results. ***A***, Discriminator performance (AZ) determined using a leave-one-out procedure (LOO). Colored lines reflect single-subject results; the black line represents the average across subjects. Grayscale shading indicates SDs. The dashed gray line represents the Az value leading to a significance level of $p = 0.05$ (obtained from a bootstrap test). Insets, Scalp distributions of discriminating component activity for three moments in time: the first time window of significant discriminator performance (100–120 ms), the local maximum (180–200 ms), and the sustained effects from 250 ms onwards (distribution shown for the 320–340 ms window). ***B***, Correlation of discriminant component value (*y*) with CE (red) and SC (blue). Confidence intervals and *p* values (FDR-corrected for the number of time points and parameters) were obtained by testing the average correlation across subjects against zero. *Time windows with significant correlations. ***C***, Discriminating component maps displaying the value of discriminating component amplitude *y* for each image and time window, averaged over the two single-trial presentations and sorted on either CE (top) or SC (bottom). Higher component activity (red) indicates relatively more evidence for a "natural" response.

of our model. We propose that our model provides a biologically realistic computational substrate from which image statistics can be derived: local contrast, which could be represented by the population response in visual areas, such as LGN or V1.

The role of image statistics in visual perception

The early, transient correlations of sERP amplitude with CE are consistent with previous reports of early ERP sensitivity to image statistics of natural scenes, demonstrating modulation of early visual ERP components (e.g., C1 and P1) by energy at different spatial frequencies (Hansen et al., 2011, 2012). During object recognition, early ERP sensitivity to luminance differences (Martinovic et al., 2011), power spectra (Johnson and Olshausen, 2003), and phase scrambling (Rousselet et al., 2008a; Bieniek et al., 2012) have also been reported. In face categorization, ERP sensitivity to phase scrambling has been found to extend into later time ranges (up to 300 ms) (Rousselet et al., 2008b), thus overlapping in time with our present SC effects. In addition, ERP sensitivity to "geometric similarity" between faces has also been found in this time range (Kahn et al., 2010). Here, we show that modulations in this time range were related to the behavioral categorization on each trial, suggesting that they might not reflect pure stimulus encoding, but also neural processing leading up to the decision outcome.

Supporting this idea, the maximal timing of the late SC effects is close to a discriminating component reported in a set of papers that used a face/car categorization task (Philiastides and Sajda, 2006; Philiastides et al., 2006; Ratcliff et al., 2009). The authors proposed that the D200 reflects an intermediate stage between early sensory processing and accumulation of decision-related information, signaling the "availability of diagnostic information." Another study, however, argued that the modulation of this component was purely related to the addition of phase noise to the stimuli (Bankó et al., 2011). Both claims could be in accordance with our findings, as SC is also sensitive to the addition of phase noise, which leads to a more Gaussian contrast distribution and thus to higher fragmentation of the scene. For the purpose of our task, however, this information could be useful (i.e., the brain does not need to discard this information as noise but may use it as input for the decision). In that case, the similarity in timing with the D200 component supports the notion that SC contains available information for global property categorization.

Indeed, modulation of neural activity by seemingly simple, low-level image properties, such as contrast, does not necessarily imply that these properties are irrelevant for image recognition: if they are relatively consistent across categories, they may have become part of a template used by the visual system to classify incoming information (Johnson and Olshausen, 2003). The CE and SC parameters reflect variations in the distribution of local contrast: they are thus derived from information that is generally considered to be "low-level." However, because their computation requires integration of this information across the scene, they pick up on global "high-level" scene information (global energy and scene fragmentation). It is not unlikely that, over the course of evolution and/or development, the visual system has adapted and developed templates that are sensitive to such variations in integrated low-level information if they are diagnostic of relevant global properties.

Naturalness as a visual primitive?

Previous work has suggested that the man-made/natural distinction is fundamental in scene perception. Categorization of man-made versus natural scenes occurs faster than basic-level categories (sea, mountain, city) (Joubert et al., 2007; Greene and Oliva, 2009b; Loschky and Larson, 2010). Basic-level categories from the same superordinate (e.g., sea vs mountain) level are also more easily confused than those from different levels (e.g., sea vs city) (Rousselet et al., 2005). Within global properties, categori-

zation may occur hierarchically, starting with man-made versus natural (Kadar and Ben-Shahar, 2012). The relation between SC and scene naturalness, as well as the influence of SC on evoked neural activity during naturalness categorization, suggests that the SC of the scene may drive this early primary distinction.

The fact that humans can quickly decide about naturalness, however, does not imply that the brain computes it automatically. Would the brain be interested in determining the naturalness of visual input in everyday viewing? SC varies with scene fragmentation, signaling the relative presence of "chaos" versus "order," rather than an absolute distinction between man-made and natural. However, there is a relation because natural scenes are more likely to be chaotic because of the presence of foliage or other structure that has not been organized by humans (as urban environments are).

We thus suggest that the primacy of man-made versus natural reflects early sensitivity to the fragmentation level of visual input. Estimating the fragmentation of incoming visual information may be a useful step in rapid scene processing, for example to allocate attention (when presented with a highly chaotic scene, or alternatively, with a highly coherent scene containing a rapidly approaching object) or cognitive control mechanisms. Future experiments can establish to what extent image properties, such as CE and SC, predict recruitment of attention or control networks.

In conclusion, together, these results suggest that natural image statistics, derived in a physiologically plausible manner, affect the perception of at least one global property: scene naturalness. The results revealed strong neural sensitivity to image statistics when subjects categorized this global property, with decision-related activity specifically being modulated by SC. We propose that, during scene categorization, the brain extracts diagnostic image statistics from pooled responses in early visual areas.

References

Baddeley R (1997) The correlational structure of natural images and the calibration of spatial representations. Cogn Sci 21:351–372. CrossRef

Bankó EM, Gál V, Körtvélyes J, Kovács G, Vidnyánsky Z (2011) Dissociating the effect of noise on sensory processing and overall decision difficulty. J Neurosci 31:2663–2674. CrossRef Medline

Bieniek MM, Pernet CR, Rousselet GA (2012) Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: evidence from parametric, test-retest, single-subject analyses. J Vis 12:1–24. CrossRef Medline

Blank H, Biele G, Heekeren HR, Philiastides MG (2013) Temporal characteristics of the influence of punishment on perceptual decision making in the human brain. J Neurosci 33:3939–3952. CrossRef Medline

Bonin V, Mante V, Carandini M (2005) The suppressive field of neurons in lateral geniculate nucleus. J Neurosci 25:10844–10856. CrossRef Medline

Conroy B, Sajda P (2012) Fast, exact model selection and permutation testing for L2-regularized logistic regression. In: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics, pp 246–254. La Palma, Canary Islands.

Croner LJ, Kaplan E (1995) Receptive fields of P and M ganglion cells across the primate retina. Vision Res 35:7–24. CrossRef Medline

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. IEEE Conf Comput Vis Pattern Recognit 248–255.

De Valois RL, De Valois KK (1990) Spatial vision. New York: Oxford UP.

Duda R, Hart P, Stork D (2001) Pattern classification. New York: Wiley.

Einhäuser W, Rutishauser U, Frady EP, Nadler S, König P, Koch C (2006) The relation of phase noise and luminance contrast to overt attention in complex visual stimuli. J Vis 6:1148–1158. CrossRef Medline

Elder JH, Zucker SW (1998) Local scale control for edge detection and blur estimation. IEEE Trans Pattern Anal Mach Intell 20:699–716. CrossRef

Fei-Fei L, Iyer A, Koch C, Perona P (2007) What do we perceive in a glance of a real-world scene? J Vis 7:10. CrossRef Medline

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am 4:2379–2394. CrossRef Medline

Gaspar CM, Rousselet GA (2009) How do amplitude spectra influence rapid animal detection? Vision Res 49:3001–3012. CrossRef Medline

Geusebroek JM, Smeulders AWM (2002) A physical explanation for natural image statistics. In: Paper presented at the 2nd International Workshop on Texture Analysis, Heriot-Watt University.

Ghebreab S, Smeulders AWM, Scholte HS, Lamme VAF (2009) A biologically plausible model for rapid natural image identification. In: Advances in Neural Information Processing Systems 22, pp 629–637.

Goffaux V, Jacques C, Mouraux A, Oliva A, Schyns PPG, Rossion B (2005) Diagnostic colours contribute to the early stages of scene categorization: behavioural and neurophysiological evidence. Vis Cogn 12:878–892. CrossRef

Graham N (1979) Does the brain perform a Fourier analysis of the visual scene? Trends Neurosci 2:207–208. CrossRef

Gratton G, Coles MG, Donchin E (1983) A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 55:468–484. CrossRef Medline

Greene MR, Oliva A (2009a) Recognition of natural scenes from global properties: seeing the forest without representing the trees. Cogn Psychol 58:137–176. CrossRef Medline

Greene MR, Oliva A (2009b) The briefest of glances: the time course of natural scene understanding. Psychol Sci 20:464–472. CrossRef Medline

Groen IIA, Ghebreab S, Lamme VAF, Scholte HS (2012a) Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. PLoS Comput Biol 8:e1002726. CrossRef Medline

Groen IIA, Ghebreab S, Lamme VAF, Scholte HS (2012b) Low-level contrast statistics are diagnostic of invariance of natural textures. Front Comput Neurosci 6:34. CrossRef Medline

Hansen BC, Jacques T, Johnson AP, Ellemberg D (2011) From spatial frequency contrast to edge preponderance: the differential modulation of early visual evoked potentials by natural scene stimuli. Vis Neurosci 28:221–237. CrossRef Medline

Hansen BC, Johnson AP, Ellemberg D (2012) Different spatial frequency bands selectively signal for natural image statistics in the early visual system. J Neurophysiol 108:2160–2172. CrossRef Medline

Hegdé J (2008) Time course of visual perception: coarse-to-fine processing and beyond. Prog Neurobiol 84:405–439. CrossRef Medline

Hochstein S, Ahissar M (2002) View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36:791–804. CrossRef Medline

Intraub H (1981) Rapid conceptual identification of sequentially presented pictures. J Exp Psychol Hum Percept Perform 7:604–610. CrossRef

Jegou H, Douze M, Schmid C (2008) Hamming embedding and weak geometric consistency for large scale image search. In: Proceedings of the 10th European conference on Computer Vision, pp 304–317.

Johnson JS, Olshausen BA (2003) Timecourse of neural signatures of object recognition. J Vis 3:499–512. CrossRef Medline

Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M (2007) Processing scene context: fast categorization and object interference. Vision Res 47:3286–3297. CrossRef Medline

Joubert OR, Rousselet GA, Fabre-Thorpe M, Fize D (2009) Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. J Vis. 9:2 1–16. CrossRef Medline

Kadar I, Ben-Shahar O (2012) A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. J Vis 12:1–17. CrossRef Medline

Kahn DA, Harris AM, Wolk DA, Aguirre GK (2010) Temporally distinct neural coding of perceptual similarity and prototype bias. J Vis 10:1–12. CrossRef Medline

Kaping D, Tzvetanov T, Treue S (2007) Adaptation to statistical properties of visual scenes biases rapid categorization. Vis Cogn 15:12–19. CrossRef

Koenderink JJ, van de Grind WA, Bouman MA (1972) Opponent color coding: a mechanistic model and a new metric for color space. Kybernetik 10:78–98. CrossRef Medline

Loschky LC, Larson AM (2008) Localized information is necessary for scene categorization, including the Natural/Man-made distinction. J Vis 8:1–9. CrossRef Medline

Loschky LC, Larson AM (2010) The natural/man-made distinction is made before basic-level distinctions in scene gist processing. Vis Cogn 18:513–536. CrossRef

Loschky LC, Sethi A, Simons DJ, Pydimarri TN, Ochs D, Corbeille JL (2007) The importance of information localization in scene gist recognition. J Exp Psychol Hum Percept Perform 33:1431–1450. CrossRef Medline

Luck SJ, Woodman GF, Vogel EK (2000) Event-related potential studies of attention. Trends Cogn Sci 4:432–440. CrossRef Medline

Martinovic J, Mordal J, Wuerger SM (2011) Event-related potentials reveal an early advantage for luminance contours in the processing of objects. J Vis 11:1–15. CrossRef Medline

McCotter M, Gosselin F, Sowden P, Schyns PG (2005) The use of visual information in natural scenes. Vis Cogn 12:938–953. CrossRef

Oliva A (2005) Gist of the scene. In: Encyclopedia of neurobiology of attention (Itti L, Rees G, Tsotsos JK, eds.), pp 251–256. San Diego: Elsevier.

Oliva A, Schyns PG (2000) Diagnostic colors mediate scene recognition. Cogn Psychol 41:176–210. CrossRef Medline

Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 42:145–175. CrossRef

Oliva A, Torralba AB, Guerin-Dugue A, Herault J (1999) Global semantic classification of scenes using power spectrum templates. In: Proceedings of the Challenge of Image Retrieval, Electronic Workshops in Computing Series. Newcastle: Springer.

Olmos A, Kingdom FA (2004) A biologically inspired algorithm for the recovery of shading and reflectance images. Perception 33:1463–1473. CrossRef Medline

Opelt A, Pinz A, Fussenegger M, Auer P (2006) Generic object recognition with boosting. IEEE Trans Pattern Anal Mach Intell 28:416–431. CrossRef Medline

Parra L, Alvino C, Tang A, Pearlmutter B, Yeung N, Osman A, Sajda P (2002) Linear spatial integration for single-trial detection in encephalography. Neuroimage 230:223–230. Medline

Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. Electroencephalogr Clin Neurophysiol 72:184–187. CrossRef Medline

Philiastides MG, Sajda P (2006) Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 16:509–518. CrossRef Medline

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. CrossRef Medline

Potter MC (1975) Meaning in visual search. Science 187:965–966. CrossRef Medline

Ratcliff R, Philiastides MG, Sajda P (2009) Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad Sci U S A 106:6539–6544. CrossRef Medline

Renninger LW, Malik J (2004) When is scene identification just texture recognition? Vision Res 44:2301–2311. CrossRef Medline

Rousselet GA, Joubert OR, Fabre-Thorpe M (2005) How long to get to the "gist" of real-world natural scenes? Vis Cogn 12:852–877. CrossRef

Rousselet GA, Husk JS, Bennett PJ, Sekuler AB (2008a) Time course and robustness of ERP object and face differences. J Vis 8:1–18. CrossRef Medline

Rousselet GA, Pernet CR, Bennett PJ, Sekuler AB (2008b) Parametric study of EEG sensitivity to phase noise during face processing. BMC Neurosci 9:98. CrossRef Medline

Rousselet GA, Gaspar CM, Wieczorek KP, Pernet CR (2011) Modeling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). Front Psychol 2:1–19. CrossRef Medline

Scholte HS, Witteveen SC, Spekreijse H, Lamme VA (2006) The influence of inattention on the neural correlates of scene segmentation. Brain Res 1076:106–115. CrossRef Medline

Scholte HS, Ghebreab S, Waldorp L, Smeulders AWM, Lamme VAF (2009) Brain responses strongly correlate with Weibull image statistics when processing natural images. J Vis 9:1–15. CrossRef Medline

Schyns PG, Oliva A (1994) From blobs to boundary edges: evidence for time and spatial-scale-dependent scene recognition. Psychol Sci 5:195–200. CrossRef

Simoncelli EP (1999) Modeling the joint statistics of images in the wavelet domain. Proc SPIE D:188–195.

Tadmor Y, Tolhurst DJ (2000) Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. Vision Res 40:3145–3157. CrossRef Medline

Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520–522. CrossRef Medline

Torralba A (2003) Contextual priming for object detection. Int J Comput Vis 53:169–191. CrossRef

Torralba A, Oliva A (2003) Statistics of natural image categories. Network 14:391–412. Medline

Walther DB, Caddigan E, Fei-Fei L, Beck DM (2009) Natural scene categories revealed in distributed patterns of activity in the human brain. J Neurosci 29:10573–10581. CrossRef Medline

Wichmann FA, Braun DI, Gegenfurtner KR (2006) Phase noise and the classification of natural images. Vision Res 46:1520–1529. CrossRef Medline

Wolfe JM, Võ ML, Evans KK, Greene MR (2011) Visual search in scenes involves selective and nonselective pathways. Trends Cogn Sci 15:77–84. CrossRef Medline

Zhu SC, Mumford D (1997) Prior learning and Gibbs reaction-diffusion. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp 1236–1250.