Inter-area correlations in the ventral visual pathway reflect feature integration

Jeremy Freeman
Center for Neural Science, New York University,
New York, NY, USA

Tobias H. Donner
Department of Psychology, University of Amsterdam,
Amsterdam, The Netherlands

David J. Heeger
Center for Neural Science, New York University,
New York, NY, USA, &
Department of Psychology, New York University,
New York, NY, USA

During object perception, the brain integrates simple features into representations of complex objects. A perceptual phenomenon known as visual crowding selectively interferes with this process. Here, we use crowding to characterize a neural correlate of feature integration. Cortical activity was measured with functional magnetic resonance imaging, simultaneously in multiple areas of the ventral visual pathway (V1–V4 and the visual word form area, VWFA, which responds preferentially to familiar letters), while human subjects viewed crowded and uncrowded letters. Temporal correlations between cortical areas were lower for crowded letters than for uncrowded letters, especially between V1 and VWFA. These differences in correlation were retinotopically specific, and persisted when attention was diverted from the letters. But correlation differences were not evident when we substituted the letters with grating patches that were not crowded under our stimulus conditions. We conclude that inter-area correlations reflect feature integration and are disrupted by crowding. We propose that crowding may perturb the transformations between neural representations along the ventral pathway that underlie the integration of features into objects.

Keywords: crowding, object recognition, ventral stream, visual word form area, inter-area correlations, fMRI

Citation: Freeman, J., Donner, T. H., & Heeger, D. J. (2011). Inter-area correlations in the ventral visual pathway reflect feature integration. *Journal of Vision, 11*(4):15, 1–23, http://www.journalofvision.org/content/11/4/15, doi:10.1167/11.4.15.

Introduction

Object representations in the primate brain are constructed through a series of transformations along the ventral visual pathway (DiCarlo & Cox, 2007; Hubel, 1982; Riesenhuber & Poggio, 1999). This process likely involves recurrent interactions across multiple levels of the cortical visual hierarchy (Lamme & Roelfsema, 2000; Roelfsema, 2006). Neurophysiological and functional magnetic resonance imaging (fMRI) studies of object recognition have characterized neural responses in individual cortical areas in the ventral pathway (Grill-Spector & Malach, 2004; Hung, Kreiman, Poggio, & DiCarlo, 2005; Logothetis & Sheinberg, 1996; Treisman & Kanwisher, 1998; Ungerleider & Haxby, 1994). For example, fMRI studies have shown that an area called the visual word form area (VWFA) in human occipitotemporal cortex responds preferentially to letters in a familiar alphabet (Baker et al., 2007; Cohen & Dehaene, 2004). However, as a result of primarily characterizing just one area at a time, we have a poor understanding of how object representations emerge from the interactions between cortical areas.

A perceptual phenomenon known as crowding provides a tool for specifically manipulating the recognizability of objects without changing their detectability (Levi, 2008; Pelli & Tillman, 2008). Crowding occurs when closely spaced but non-overlapping flankers hinder the identification of a peripherally viewed target object, such as a letter (Figure 1A; Bouma, 1970; Pelli, Palomares, & Majaj, 2004). During crowding, the local features of the target remain the same and the target remains visible, but it becomes unrecognizable, appearing as a dynamically changing texture of elementary features that lack identity (Balas, Nakano, & Rosenholtz, 2009; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Pelli, Palomares, & Majaj, 2004). Crowding operationally distinguishes feature detection (unaffected by crowding) from object identification (disturbed by crowding), which has been shown to depend on “feature integration” (the process by which information is combined from multiple features to construct object representations; Levi, 2008; Martelli, Majaj, & Pelli, 2005; Nandy & Tjan, 2007; Pelli, Burns,

Received August 17, 2010; published April 26, 2011 ISSN 1534-7362 © ARVO

Farell, & Moore-Page, 2006; Pelli, Farell, & Moore, 2003; Pelli et al., 2004; Pelli & Tillman, 2008; Treisman & Schmidt, 1982; Treisman & Gelade, 1980; Wolfe & Cave, 1999). Thus, crowding is an ideal tool for specifically probing the neural basis of feature integration (Levi, 2008; Pelli et al., 2004; Pelli & Tillman, 2008).

We used letter crowding to selectively disrupt letter recognizability and interfere with feature integration, and we characterized the effect of crowding on interactions between early visual cortex and VWFA. We used fMRI to measure the effect of crowding on correlations between intrinsic activity fluctuations in several different visual cortical areas. We also characterized the effect of crowding on mean stimulus response amplitudes in early visual areas and VWFA. Temporal correlations between early visual areas (including V1) and VWFA were lower when the targets were crowded than when they were uncrowded. These differences in correlation were retinotopically specific to the locations of the peripheral target letters, and persisted when attention was diverted away from the target letters. But correlation differences were not evident when the letter stimuli were substituted with grating patches that were not crowded even though they were presented under the same stimulus conditions as the letters. Crowding reduced mean response amplitudes for both letters and gratings within some visual areas (V2, V3, and V4), but the effect of crowding on temporal correlations between V1 and VWFA occurred without evidence of a change in mean response amplitudes.

We conclude that crowding affects coherent, intrinsic activity fluctuations between high-level, category-selective and earlier, feature-selective neural populations of the ventral visual pathway. We propose that feature integration depends on the stability of the series of transformations linking neuronal representations across these different cortical areas and that crowding dynamically perturbs these transformations.

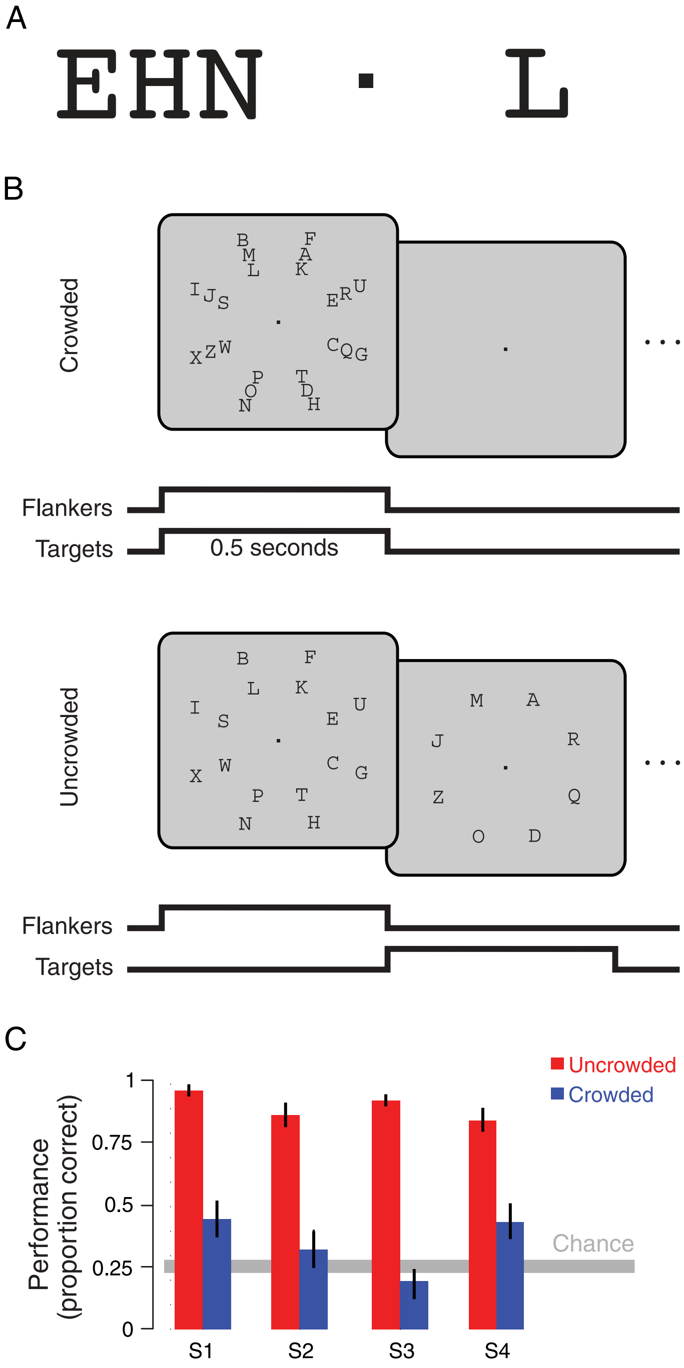

Figure 1. Stimulus and task. (A) Crowding demonstration. Fixate on the square and identify the single letter on the right. It is easy. Now fixate and try to identify the middle letter in the triplet on the left. It is hard. It is even hard to tell whether or not the letter on the right is among the letters on the left. The flankers crowd the target, spoiling recognition. (B) Manipulating crowding. Displaying tightly spaced targets and flankers simultaneously produces crowding, making the targets hard to identify (crowded). Displaying tightly spaced targets and flankers in alternation relieves crowding, making the targets easy to identify (uncrowded). Targets and flankers were presented at 1 Hz, either simultaneously or in alternation. Both uncrowded and crowded trials began with a 0.5 s frame of simultaneous targets and flankers followed by a 0.5 s blank frame (not shown in the figure) to equate stimulus onset between conditions. All subsequent frames were as shown in the figure (simultaneous for crowded, alternating for uncrowded). The full duration of each trial (time from start of trial to end of last frame) in the fMRI experiments ranged from 15 to 21 s (durations were randomly selected from 15, 16.5, 18, 19.5, and 21 s). (C) Behavioral effect of crowding. In a separate psychophysical experiment, subjects viewed a 1 s stimulus presentation, with targets and flankers presented either in alternation (red, uncrowded) or simultaneously (blue, crowded). Observers responded by selecting from one of four possible choices, one of which was the target (chance = 0.25, gray bar). Crowding reduced letter identification performance by a factor of two or more ($p < 0.01$, two-sided t-test, $n = 4$ subjects). Error bars are 68% confidence intervals estimated from the bootstrapped distribution of proportion correct.

Methods

Subjects

Data were acquired from four healthy subjects with normal or corrected-to-normal vision (all male, age range: 22–34 years). Two subjects were authors. Experiments were conducted with the written consent of each subject and in accordance with the safety guidelines for fMRI research, as approved by the University Committee on Activities Involving Human Subjects at New York University. Each subject participated in several scanning sessions: one to obtain a high-resolution anatomical volume, one to define retinotopically organized cortical visual areas (V1, V2, V3, V4, and parafovea), one to define the visual word form area region of interest (Figure 2), and four sessions in the main experiment. In addition, all four subjects participated in four sessions each for the grating control experiment, and three subjects participated

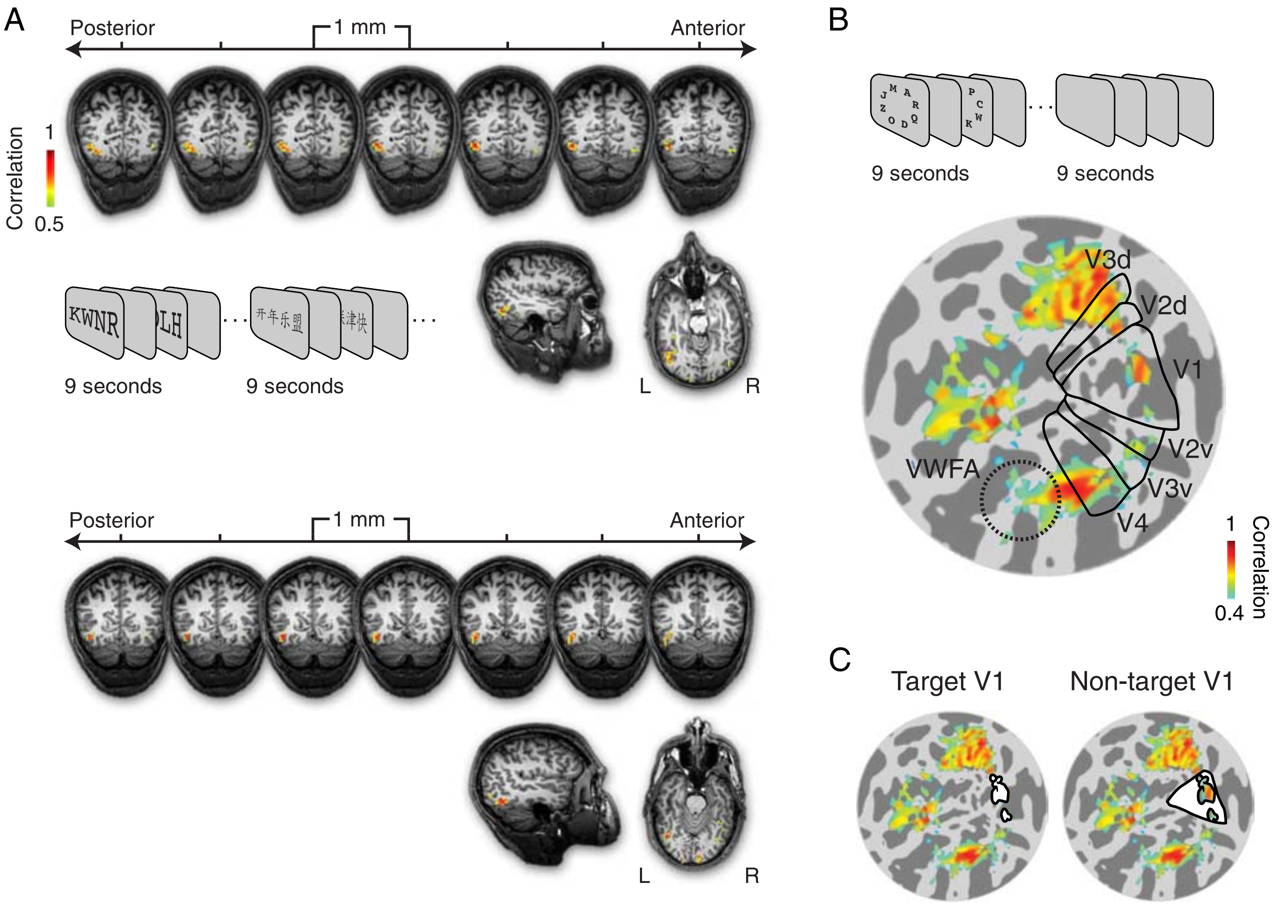

Figure 2. Regions of interest. (A) The visual word form area (VWFA) was localized, separately for each subject, by presenting alternating blocks of English and Chinese characters. For two sample subjects, responses are shown on seven representative coronal anatomical slices (1-mm slice thickness). Sample sagittal and axial slices are shown in the lower right. Color indicates the correlation between the measured time series and the best-fitting sinusoid with period equal to that of the stimulus alternations ($r > 0.5$, in phase with blocks of English letters, see Methods section). The robust left-lateralized activity was consistent across subjects and was used to define the VWFA in each subject. (B) Target-specific subregions of visual areas V1–V4. The borders of visual areas V1, V2, V3, and V4 are indicated with black lines on a flattened representation of the left hemisphere occipital lobe of a sample subject. For reference, the approximate location of the VWFA is indicated with a dotted circle, but this area was defined in the volume, not the flat map, so this does not correspond exactly to the region of interest used in our analysis. Dark gray indicates sulci and light gray indicates gyri. Color indicates correlation with the best-fitting sinusoid ($r > 0.4$, in phase with blocks of target letters). Letters were presented in the periphery, and thus evoked activity in corresponding subregions of each visual area. (C) Left: "Target V1" was defined by restricting V1 to those voxels showing responses in phase with blocks of target letters and exceeding the correlation threshold of 0.4. The region of interest is filled in white, with a black outline. Target subregions for other areas (V2–V4) were defined similarly (not shown). Right: In a control analysis to test for retinotopic specificity, a "Non-target V1" subregion was defined by taking the entire retinotopically defined V1 (shown in (B)) and excluding the "Target V1" subregion.

							No. of trials	No. of trials
		V1	V2	V3	V4	VWFA	(uncrowded)	(crowded)
Subject 1	Letters	396	754	951	684	1,104	125	125
	Gratings	147	543	936	364	1,104	125	125
	Diverted attention	671	751	831	271	1,104	70	70
Subject 2	Letters	1,020	1,638	1,938	807	2,005	135	135
	Gratings	493	1,051	1,499	523	2,005	145	145
	Diverted attention	815	1,110	1,222	302	2,005	70	70
Subject 3	Letters	921	1,075	1,164	694	2,268	145	145
	Gratings	460	889	591	627	2,268	135	135
	Diverted attention	595	890	573	658	2,268	70	70
Subject 4	Letters	864	1,049	1,332	1,749	1,731	125	125
	Gratings	505	681	921	1,141	1,731	125	125

Table 1. ROI volumes and number of experimental trials. ROI volumes (mm^3) were averaged across repeated scanning sessions (3–4), for each subject and each experiment. Target-specific subregions in V1, V2, V3, and V4 were defined in each session using the average of 1–2 localizer runs, separately for letters and gratings. The VWFA was defined once for each subject in an independent localizer session (see Methods section). The final two columns list number of uncrowded (sequential target and flankers) and crowded (simultaneous) trials, per subject and condition.

in two sessions for the attentional control experiment. Table 1 lists region of interest (ROI) sizes and the number of trials for each subject and experiment.

Main experiment

Crowding occurs when non-overlapping flankers jumble the appearance of a target object, making it hard to identify (Bouma, 1970; Pelli et al., 2004; Figure 1A). Crowding does not affect the detectability of objects, including letters (Pelli et al., 2004). Stimuli in our experiments were arrays of black letters on a gray background. Letters were arranged in eight radial arms, each consisting of three letters; the center letter was the "target" and the letters inside and outside were the "flankers" (Figure 1B). Letters were capital letters displayed in the Courier font (see samples in Figures 1A and 1B). Each letter subtended approximately 1°. The spacing between targets and flankers was 1.5° center to center, with no overlap between targets and flankers. The eccentricity of the targets was 8°; at this eccentricity, the spacing between targets and flankers was well below that required to produce crowding (Pelli et al., 2004; Toet & Levi, 1992). Uncrowded letters at this size and eccentricity were readily recognizable (see Figure 1C and Psychophysics experiments section).

Stimuli were generated on a Macintosh PowerPC computer running MATLAB with MGL (available at http://gru.brain.riken.jp/doku.php). Stimuli for fMRI experiments were displayed via an LCD projector onto a back-projection screen in the bore of the magnet. Subjects were supine and viewed the projected stimuli through an angled mirror (maximum eccentricity of 11°). Stimuli for psychophysical experiments (see below) were displayed on 32 × 22 cm LCD screen. For the fMRI experiment, gray background luminance of the background was 526 cd/m^2. At maximum contrast, the minimum luminance of the stimulus was 31 cd/m^2, and the maximum luminance was 1083 cd/m^2. Gray background luminance for the psychophysical display was 69 cd/m^2 (minimum was 1.8 cd/m^2; maximum was 345 cd/m^2).

Crowding was manipulated by presenting targets and flankers either simultaneously or sequentially (Figure 1B). Specifically, in the crowded condition, flankers and targets were presented simultaneously for 0.5 s alternating with 0.5 s of fixation. In the uncrowded condition, flankers were presented alone for 0.5 s alternating with targets alone for 0.5 s. At each small location within the visual field, the overall visual stimulation (integrated over time) was the same for both simultaneous and sequential presentations.

Each experimental run consisted of five uncrowded trials and five crowded trials, plus one trial at the beginning that was later discarded (see below). Thus, each subject completed an equal number of crowded and uncrowded trials. At the start of each trial, targets and flankers were displayed simultaneously for 0.5 s followed by a 0.5 s blank period. This was used as the start of both crowded and uncrowded trials, so as to equate the onset between crowding conditions. Subsequent frames of each trial presented target and flanker letters at 1 Hz (simultaneous for crowded and sequential for uncrowded) for a duration of between 15 and 21 s, randomly selected from 15, 16.5, 18, 19.5, and 21 s. At the start of each trial, eight target letters were randomly selected from the 26 letters in the English alphabet. These eight letters remained the targets for the entire trial. On each subsequent flanker presentation, 16 new letters were randomly selected as flankers from the remaining 18 letters. Thus, the targets were the same throughout the trials, but the flankers changed each time they were presented, and were always different from the targets. Flanker and target selections were balanced across the crowded and uncrowded trials. Each trial was followed by an inter-stimulus interval with a randomly chosen duration (15, 16.5, 18, 19.5, or 21 s).

For the entire duration of each experimental run, subjects performed a near-threshold contrast discrimination task at fixation. In each trial of this contrast discrimination task, the fixation cross changed contrast twice, and subjects indicated with a button press which contrast presentation (first or second) was darker. A “1 up 2 down” staircase procedure was used to maintain performance at approximately 70% correct. Each trial of this contrast decrement task lasted 2.9 s, such that the contrast decrements were out of phase with the main experimental stimulus presentations.

Grating control experiment

The trial sequence, task, and experimental protocol were identical to the main fMRI experiment, but instead of letters, stimuli consisted of oriented grating patches (100% contrast). Each grating patch was a sinusoidal grating with a spatial frequency of 2.5 cycles/deg windowed by a two-dimensional Gaussian (size, 2° × 2°; standard deviation, 1/4°). In each arm of the display, there were three identical grating patches with the same orientation. Thus, the targets and flankers had the same orientation. As in the main experiment, the eccentricity of the target was 8°, the target-to-flanker spacing was 1.5° center to center, and the gratings did not overlap. The orientation of the three gratings in each arm was randomly chosen from 26 orientations equally spaced between 0° and 180°, and grating orientations were constant throughout each trial. The main experiment and the grating control experiment were performed in separate scanning sessions on different days.

Psychophysics experiments

Three psychophysics experiments were performed outside the scanner to measure the effects of flankers on letter and grating identification.

The first was a psychophysics experiment to confirm the effectiveness of the crowding manipulation with letters in the main fMRI experiment (Figure 1C). For each of 80 trials, subjects viewed one stimulus presentation (either 0.5 s of simultaneous targets and flankers followed by 0.5 s of fixation or 0.5 s of flankers followed by 0.5 s of targets), with a spatial arrangement identical to that in the main experiment. Immediately before the letter display, a line cue at fixation pointed to one of the radial arms. After the letter display, four possible responses were presented at fixation. Subjects indicated with key press which of the four choices was the target. Chance performance was 25%.

Second, subjects performed a psychophysical experiment to assess the effect of simultaneous presentation on coarse grating identification. For each of 160 trials, subjects viewed one stimulus presentation (either 0.5 s of simultaneous targets and flanker gratings followed by 0.5 s of fixation or 0.5 s of flanker gratings followed by 0.5 s of targets), with a spatial arrangement identical to that in the grating fMRI experiment. In each arm of the display, the grating in the center (the target) was randomly chosen to be either the same orientation as the flankers or rotated 90° (i.e., orthogonal to the flankers). Immediately before the grating display, a line cue at fixation pointed to one of the radial arms. After the grating display, observers indicated (two-alternative forced choice, 2AFC) with a button press whether the center grating in the cued arm was either parallel to the flankers or orthogonal. Chance performance was 50%.

Third, we performed an additional experiment with letters, but using a 2AFC task that was more comparable to that in the grating experiment. Trial structure and stimulus arrangement were identical to the psychophysics experiment with letters described above. However, in each arm of the display, the two flanker letters were identical to one another, and the letter in the center (the target) was randomly chosen to be either the same as the flankers or mirror-reversed. Letters were restricted to those that are different when mirror-reversed. Observers indicated with a button press whether the center letter in the cued arm was normal or mirror-reversed. Chance performance was 50%.

Diverted attention control experiment

The trial sequence and peripheral stimuli were nearly identical to the main experiment, with only two differences. First, subjects performed a highly demanding rapid serial visual presentation (RSVP) task at fixation. For the duration of each trial (between 15 and 21 s), subjects were required to count the number of Xs in a stream of rapidly presented letters (V, N, Z, K, W, A, M, Y, L, and X). Presentation rate was either 4 Hz or 5 Hz, chosen individually for each subject to ensure that performance was below ceiling. Letters subtended approximately 1°, in the Helvetica font. Between one and four Xs were shown during a trial. After each trial, the letters were replaced by a fixation point, cueing subjects to respond by pressing one of four keys corresponding to the number of target Xs detected (one to four). The second difference between this experiment and the main experiment is that during each trial, Xs were occasionally substituted for the peripheral target letters. These were included only to match the dual-task psychophysics experiment (see below). The Xs were presented at two random locations among the peripheral targets, between one and four times during each trial.

Subjects performed a dual-task psychophysics experiment outside the scanner to confirm that their letter identification performance in the periphery was indistinguishable from chance for both crowded and uncrowded letters while they performed the central letter identification task. In each block of trials, subjects viewed letters in the periphery and letters at fixation, and they were asked to attend primarily to either the peripheral letters or the letters at fixation. Just as in the diverted attention fMRI experiment, peripheral letters were shown at 1 Hz either sequentially (uncrowded) or simultaneously (crowded), and letters were shown rapidly at fixation (4–5 Hz, chosen for each subject, see below). Subjects counted the number of Xs appearing at fixation (fixation task) and the number of frames containing Xs among the peripheral letters (peripheral task). These "X-frames" occurred between one and four times during each trial; chance performance was 0.25. For each block of trials, one of the two tasks was designated the primary task. Subjects were instructed to first respond to the primary task, and then respond to the secondary task. Presentation rate for letters at fixation was chosen so that fixation task performance would be near 80% when the fixation task was primary (4 Hz for S1 and S3, 5 Hz for S2). In addition, for each subject, the difficulty of the peripheral task was adjusted so that performance on the peripheral task was approximately 80% when the letters were uncrowded and the peripheral task was primary. To adjust the difficulty of the peripheral task, we adjusted the number of peripheral target letters (among the eight) that were substituted with Xs on each "X-frame" (two Xs for S1 and S3, five Xs for S2). Each subject completed 8 blocks of 10 trials each (5 uncrowded trials and 5 crowded trials). The peripheral task was the primary task in 4 blocks, and the fixation task was the primary task in the other 4 blocks (randomly ordered).

MRI data acquisition

MRI data were acquired with a 3-T Siemens Allegra (Erlangen, Germany) scanner equipped with a four-channel phased-array surface coil positioned at the back of the head (NM-011 transmit head coil and NMSC-021 receive coil, Nova Medical, Wakefield, MA). We measured blood oxygen level-dependent changes in MRI signal intensity using a standard echoplanar imaging sequence with the following parameters: repetition time, 1.5 s; echo time, 30 ms; flip angle, 75°; 64 × 64 matrix size; voxel size, 3 × 3 × 3 mm; 27 slices oriented roughly perpendicular to the calcarine sulcus. At the beginning of each session, we acquired anatomical (T1-weighted) images in the same slices as functional images but with twice the in-plane resolution (MPRAGE, voxel size, 1.5 × 1.5 × 3 mm).

A high-resolution anatomical volume was acquired for each subject in separate scanning sessions (MPRAGE, voxel size of 1 × 1 × 1 mm). This volume was used for cortical surface extraction and flattening and for registration across scanning sessions.

Preprocessing

The fMRI data were preprocessed as follows. Data from the beginning of each functional run were discarded (12 time points for localizer runs, corresponding to the first cycle; 24 time points in main experimental runs, corresponding to the first full trial) to minimize the effect of transient magnetic saturation and allow the hemodynamic response to reach steady-state baseline. Head movement within and across scans was compensated using standard procedures (Nestares & Heeger, 2000). The time series from each voxel was divided by its mean to convert from arbitrary intensity units to percent modulation and high-pass filtered (cutoff = 0.01 Hz) to remove low-frequency noise and drift (Smith et al., 1999). Our results did not differ qualitatively when using a lower cutoff of 0.005 Hz.

The anatomical volume, acquired in each scanning session, was aligned to a high-resolution T1-weighted anatomy volume (acquired in a different scanning session) by an automated robust image registration algorithm (Nestares & Heeger, 2000). The resulting alignment parameters were used to resample the functional data from each scanning session into the image space of the high-resolution anatomy and to transform ROIs from the coordinates of the high-resolution anatomy to the image slice coordinates of each scanning session. This allowed us to coregister the data and extract time series from corresponding ROIs across scanning sessions.

Inter-area correlation analysis

Mean stimulus-evoked responses were computed using deconvolution (Dale, 1999), separately for each subject, each ROI (see below for descriptions of how the ROIs were defined), each condition (simultaneous vs. sequential), and each trial duration. We formally express the estimation of mean stimulus-evoked responses as the deconvolution procedure (Dale, 1999). However, our trial durations (including stimulus and inter-stimulus interval) were all of longer duration than the estimated response time courses. Therefore, the analysis was equivalent to averaging responses across all repeats of each trial type (i.e., a trial-triggered average). There were 10 different trial types (simultaneous vs. sequential × 5 durations) in each experiment (main experiment, grating control experiment, diverted attention control experiment). fMRI response time courses were averaged across voxels in each ROI. Each subject participated in multiple scanning sessions, so for each subject and experiment, response time courses from each ROI were concatenated across scans and scanning sessions. Mean responses were then estimated by solving an equation of the form $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$. The

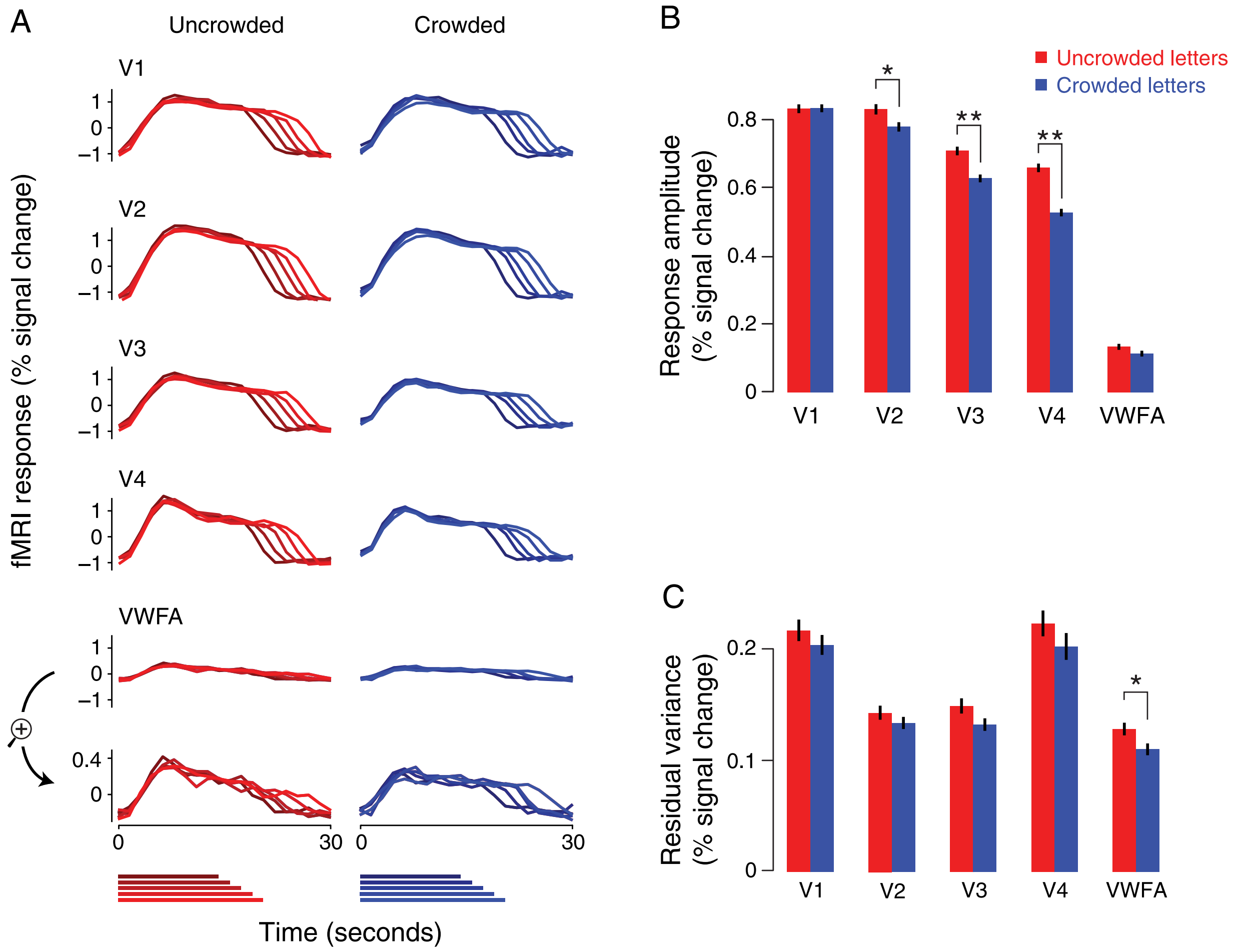

Figure 3. Responses to crowded and uncrowded letters. (A) Stimulus-evoked response time courses for a representative subject estimated using deconvolution, separately for uncrowded (left, red) and crowded (right, blue) trials. Each panel plots the fMRI responses as a function of time from the start of the trial. Responses were estimated separately for each of five trial durations. Darker colors correspond to shorter durations. Stimulus durations are shown at the bottom as horizontal colored bars. Longer stimulus durations evoked responses that were sustained for correspondingly longer time. Responses in VWFA were weaker than in other ROIs, so they are replotted in the bottom panel at a magnified scale; although weaker, the VWFA responses still followed the stimulus durations. (B) Mean fMRI response amplitudes for each ROI, averaged over the same temporal period used in the correlation analysis (see Figure 4A). Error bars are 68% confidence intervals estimated from the bootstrapped distribution of mean response amplitudes. Asterisks indicate statistically significant differences in response amplitudes (*$p < 0.05$ and **$p < 0.001$, two-sided permutation test). (C) Variance of residual responses (see Methods section). Error bars are 68% confidence intervals estimated from the bootstrapped distribution of variances. Asterisks indicate statistically significant differences in residual variances (*$p < 0.05$, two-sided permutation test).

vector $\boldsymbol{y}$ was an $n \times 1$ vector containing the measured response time course concatenated across scans and sessions; n indicates the total number of time points, including all trials from all scans and sessions. The vector $\boldsymbol{x}$ was a $10j$ vector containing the j time points of the estimated mean responses for each of the 10 trial types, and $\mathbf{A}$ was the design matrix that characterized the trial sequence. The design matrix $\mathbf{A}$ had n rows and $10j$ columns. The first column contained the value 1 at indices corresponding to the onset of the first trial type (e.g., uncrowded, 15 s) and 0 elsewhere. The second column contained a value of 1 at indices corresponding to the second time point of those trials, and so on for j columns. The next j columns were similarly organized but corresponded to the second trial type, and so on for the 10 trial types. Mean response time courses $\hat{\boldsymbol{x}}$ were estimated using ordinary least squares (i.e., regression). We used $j = 20$ parameter estimates, corresponding to 30 s (see Figure 3A for representative mean response time courses, $\hat{\boldsymbol{x}}$).

Two complementary procedures were adopted for computing inter-area correlations. The results of both analyses are presented (see Results section). The first step in both analyses was to extract time series epochs corresponding to individual trials. These epochs were time-shifted with respect to the trial to account for hemodynamic delay and minimize any residual effects of onset and offset transients. The extracted epochs began 6 s after stimulus onset and ended 3 s after stimulus offset

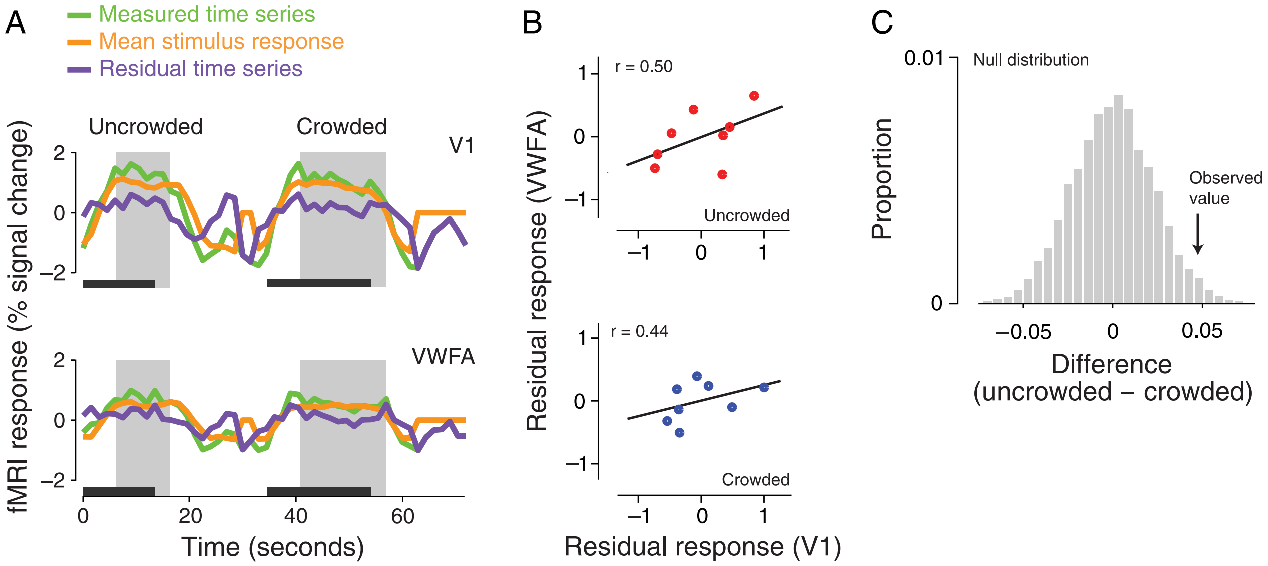

Figure 4. Inter-area correlation analysis. (A) Sample fMRI responses as a function of time for two ROIs in a single subject (top: V1; bottom: VWFA). Two full trials are shown (left: uncrowded; right: crowded). Three curves are plotted: green, measured time series; orange, mean stimulus-evoked response (estimated using deconvolution separately for crowded and uncrowded trials, different trial durations, ROIs, and subjects); purple, residual time series after removing the mean stimulus-evoked response. Dark gray bars indicate the duration of each trial. Shaded gray rectangles indicate the epoch of residual time series that was extracted for correlation analysis, shifted relative to the trial to account for hemodynamic delay. (B) Within-trial inter-area correlation analysis. Each data point represents a single time point of the residual responses in each of the two ROIs during a single representative trial of the experiment (top: uncrowded trial; bottom: crowded trial). A best-fitting regression line is drawn only for illustration, because successive time points may exhibit serial dependence and hence may not be independent. (C) Statistical significance of the effect of crowding on inter-area correlation. A non-parametric permutation test was performed by randomly shuffling the “uncrowded” and “crowded” trial labels and then recomputing the correlations (see Methods section). Repeating this 10,000 times produced a distribution of correlation differences under the null hypothesis. The proportion of the null distribution greater than the observed difference provided a one-sided *p*-value.

(Figure 4A). Each of these extracted epochs was between 12 and 18 s (depending on trial duration).

In the “within-trial” analysis for each pair of ROIs, we first removed the mean stimulus-evoked responses from the entire time course (as described in the following paragraph). We then computed correlation coefficients between residual time series epochs corresponding to each individual trial and averaged the resulting correlation coefficients separately for crowded and uncrowded trials. This procedure was similar to the coherence analysis of electrophysiological data (Mitra & Pesaran, 1999); it was possible, in spite of the sluggish hemodynamics underlying the fMRI responses, because of the long inter-stimulus intervals and trial durations in our experimental protocol. The effect of crowding was quantified as the difference between the two correlations (uncrowded minus crowded), for each pair of ROIs. This analysis isolated within-trial dynamics and thus eliminated any effects due to between-trial variability, but it did so with minimal preprocessing of the data, beyond the initial removal of the mean stimulus-evoked responses. (To further minimize preprocessing, we performed a version of the within-trial analysis in which correlations were computed directly on the time series without removing the mean stimulus-evoked responses and also obtained similar results, see Results section). However, the within-trial analysis had limited statistical power because it relied on high-variance individual trial correlation estimates that were each based on a small number of samples.

To remove the driving effect of the stimulus, and measure correlations reflecting intrinsic interactions between cortical areas, correlations were computed using the residuals that remained after removing the mean stimulus-evoked responses. For each ROI, a residual time series was computed by subtracting a predicted response time series. Specifically, after computing the parameter estimates $\hat{\boldsymbol{x}}$, a predicted time series $\hat{\boldsymbol{y}}$ was computed by multiplying the design matrix by the parameter estimates, that is, $\hat{\boldsymbol{y}} = \mathbf{A}\hat{\boldsymbol{x}}$. The residual time series was then computed by subtraction, $\boldsymbol{r} = \boldsymbol{y} - \hat{\boldsymbol{y}}$. This procedure isolated correlations reflecting intrinsic cortical interactions. We expect these intrinsic interactions to be modulated by the stimulus condition (e.g., simultaneous versus sequential, and letters versus gratings) but not driven directly by the external stimulus.

A complementary “concatenated” analysis, suggested by an anonymous reviewer, eliminated the effect of between-trial variability by projecting (regressing) out the mean responses separately on each trial rather than on the entire time course. This projection was performed on each trial epoch. The residual time series epochs corresponding to uncrowded and crowded trials were then separately concatenated, and correlation coefficients were computed on the concatenated time series. This version of the analysis provided greater statistical power than the within-trial analysis (because the correlations were computed on much longer time series) but still eliminated between-trial variability. However, it was farther removed from the data because it required an additional projection operation applied to each trial.

A large component of the correlations between neural population signals in different cortical areas is independent of stimulus and task parameters (Fox & Raichle, 2007; Leopold, Murayama, & Logothetis, 2003; Vincent et al., 2007). This component of correlated activity seems, at least in part, driven by common input from subcortical structures (Leopold et al., 2003). By calculating the difference in correlation induced by a manipulation of feature integration (a cortical process linking specific stimulus representations), we isolated the component that reflected the cortical interactions underlying feature integration from the stimulus-independent component of correlated fMRI activity.

In an alternate version of the analysis, we assessed inter-area coupling by computing covariance rather than correlation. The analysis of covariances was otherwise identical to the analysis of correlation.

Our inter-area correlation analyses differed from the standard procedure adopted in the literature, in which the mean response is removed from the entire time course (as described above), residual time courses are concatenated across trials, and correlations are computed on the concatenated time courses (Friston, 2002). This procedure is limited in its specificity because effects can potentially reflect both slow inter-trial dynamics (i.e., trial-to-trial covariation in neural response amplitudes) as well as fast within-trial dynamics (i.e., covariation in activity fluctuations during each trial). In contrast, our analyses were designed to isolate correlations reflecting within-trial dynamics.

Statistics

Most of our statistical analyses used a bootstrapping procedure to compute confidence intervals and a nonparametric two-sample permutation test to compute *p*-values (Efron & Tibshirani, 1993).

A permutation test was used to determine the statistical significance of differences in correlation coefficients between the residual responses to the crowded (simultaneous presentation) and uncrowded (sequential presentation) trials. The trial type labels (crowded/uncrowded) were permuted, and correlation coefficient differences were recomputed from the residual time series epochs. Correlations were computed either first separately within each trial and then averaged across trials (within-trial analysis) or after concatenating time series across trials (concatenated analysis). Repeating this 10,000 times produced an empirical distribution of the correlation coefficient differences under the null hypothesis that there was no effect of crowding (see Figure 4C for a sample null distribution). The observed values of the correlation coefficients were then compared to the null distribution. For a one-sided test, a *p*-value was computed as the proportion of samples in the null distribution greater than (if the observed value was positive) or less than (if it was negative) the observed value. For a two-sided test, a *p*-value was computed as the proportion of samples in the null distribution greater than or equal in magnitude to the observed value (ignoring sign). This test was computed separately for each ROI pair.

A bootstrapping procedure provided confidence intervals for the differences in correlation coefficients, as plotted in Figures 7A and 7C. In this procedure, residual responses corresponding to individual trials (e.g., residual time series epochs) were resampled with replacement, keeping the labels intact. After each resampling, the differences in correlation coefficients were recomputed. Repeating this 10,000 times produced empirical distributions of the correlation differences. The 16th and 84th percentiles of this distribution provided an estimated 68% confidence interval (equivalent to ±1 standard deviation if it was normally distributed).

A bootstrapping procedure was also used to compare the magnitude of the crowding effect (difference in correlation between the responses to simultaneous and sequential presentations) between letters and grating patches. In this procedure, responses corresponding to individual trials from all four conditions—sequential letters, simultaneous letters, sequential gratings, and simultaneous gratings—were resampled with replacement, keeping the labels intact. After each resampling, the differences in correlation coefficients (sequential minus simultaneous) were recomputed separately for letters and gratings, and the difference of these differences (letters minus gratings) was then computed. Repeating this 10,000 times produced empirical distributions of the difference in effect size between letters and gratings. For each pair of areas, if the 5th percentile of this distribution exceeded 0, it implies that the effect of simultaneous versus sequential presentation on correlations for letters was significantly larger than for grating patches at the 0.05 (one-sided) significance level.

We performed two additional analyses to characterize the mean and variance of the stimulus-evoked responses in each ROI during the main experiment. We estimated a single response amplitude for each trial by averaging the measured time series across those time points that were contained in the same epoch that was extracted for the correlation analysis (see Figure 4A). These response amplitudes were then averaged across trials and across subjects. A two-sided permutation test was used to compare mean response amplitudes between uncrowded and crowded trials. We computed the variance of the residual responses, after removing the mean stimulus-evoked responses from the entire time course (see above). The variance was computed over the residual time series obtained from each trial epoch, separately for crowded and uncrowded trials. Trial variances were combined across subjects, and a permutation test was used to test for statistically significant differences in variance between the crowded and uncrowded trials. We also used a bootstrapping procedure to estimate 68% confidence intervals for both mean response amplitudes and residual variances.

Defining retinotopic visual area ROIs

Standard traveling wave methods were used to identify meridian representations corresponding to the borders between retinotopically organized visual areas V1, V2, V3, and V4 (Larsson & Heeger, 2006; Wandell, Dumoulin, & Brewer, 2007; see Larsson & Heeger, 2006 for details). The definition of human visual area V4 is controversial (Brewer, Liu, Wade, & Wandell, 2005; Hansen, Kay, & Gallant, 2007; Tootell & Hadjikhani, 2001); we followed the conventions used by Wandell et al. (2007). Note that area human visual area V4 has alternatively been called hV4 by some laboratories (Wandell et al., 2007), but we use the term V4 throughout this paper. A control parafoveal ROI was defined for each subject using the retinotopic mapping data. Specifically, regions V1–V4 were combined, and then restricted to those voxels responding preferentially to expanding ring stimuli shown between 0.5° and 2°. Thus, our parafoveal ROI included the near fovea of the retinotopic areas V1–V4 only, and not other areas.

Target-specific subregions of visual areas V1–V4 were identified by having subjects complete one or two runs of a localizer experiment at the beginning of each session. Different localizers were used for letters and gratings, each of which used a periodic block alternation protocol (20 cycles of 18 s). For letters, alternation was between 9 s blocks of random letters displayed at the eight target locations (flickering at 2 Hz: letters for 0.25 s, fixation for 0.25 s) and 9 s blocks of fixation only. Each letter on each frame was randomly chosen from the 26 English letters. For gratings, alternation was between 9 s blocks of randomly oriented gratings displayed at the eight target locations (flickering at 4 Hz: gratings for 0.125 s, fixation for 0.125 s) and 9 s blocks of fixation only. Each grating on each frame was randomly chosen from 26 orientations equally spaced between 0° and 180°.

Time series from each voxel were averaged across runs and fit with a sinusoid with period equal to that of the stimulus alternations. Each visual area V1–V4 was restricted to those voxels with responses that modulated in phase with the target presentations (0 to π) and with a correlation greater than 0.4 (see Figure 2C for an example). Repeating the analysis with different thresholds (between 0.2 and 0.5) produced qualitatively similar results. The sizes of the resulting ROIs are listed in Table 1. For subject 1, in one session, some voxels in V1/V2 target areas, as defined by the localizer, overlapped with a large vein along the posterior midline, as identified by high T1-weighted image intensity in the in-plane anatomy, low T2*-weighted intensity in the functional images, and high variance (Olman, Inati, & Heeger, 2007). These voxels were removed from the corresponding ROIs for that session.

Defining the visual word form area

Our experiments studied the effect of crowding on familiar letter stimuli, so we measured fMRI responses in an area outside of the retinotopic cortex that responds preferentially to familiar letters. An area in left occipitotemporal cortex responds strongly to letters in a familiar alphabet and plays a putative role in letter identification and word recognition (Baker et al., 2007; Cohen & Dehaene, 2004). We identified this area by alternating random strings of English vs. Chinese characters and refer to the responsive cortex as "visual word from area" (VWFA), based on a procedure introduced by Baker et al. (2007). We are aware that other groups have identified the VWFA based on responses to visual words (Cohen & Dehaene, 2004). It is possible that there are distinct, non-overlapping areas in the occipitotemporal cortex responsive to familiar letters and words. However, we adopted the localizer used by Baker et al. because our experimental manipulations focused on letter perception.

We presented subjects with alternating blocks of English and Chinese random character strings (20 cycles of 18 s; each cycle showed 9 s of English characters followed by 9 s of Chinese characters). Letter strings were flickered at 1 Hz (letters for 0.5 s, fixation for 0.5 s), and new random strings of either 4 or 6 characters were generated on each frame. English characters were capital letters in the font Courier: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z. Chinese characters were in the TrueType font Hanzi-Kaishu (available at http://www.azfonts.net/load_font/hanzikaishu.html): 际 家 加 東 节 姐 届 津 京 举 开 快 来 乐 联 临 美 盟 民 明 南 你 年 您 亲 请. English and Chinese characters both subtended approximately 1.5°, and letter strings were thus matched for overall spatial envelope. Subjects performed 7–8 runs of the localizer. Data were analyzed by averaging across runs and fitting a sinusoid, as was done for the target localizer (above), except that here a region of interest was defined in each subject as a contiguous left-lateralized cluster of voxels in the region of the occipitotemporal sulcus that modulated in phase with the English letter presentations (0 to π) and with a correlation greater than 0.5 (see Figure 2A). The robust left-lateralized activity was consistent across subjects and was thus used to define VWFA in each subject.

Results

We used fMRI to examine the effects of crowding (Figure 1A) on correlations between responses in multiple visual cortical areas. Subjects viewed target letters at eight locations in the periphery of the visual field (8° eccentricity), each of which was paired with closely spaced flanker letters (Figure 1B). Letters are useful as examples of objects because they are a basic level stimulus category consisting of features that are integrated into a single object, without any intermediate parts (Kleinschmidt, Büchel, Hutton, Friston, & Frackowiak, 2002; Pelli et al., 2006). In psychophysical studies, crowding is typically manipulated by increasing the spacing between flankers and targets (Bouma, 1970; Pelli et al., 2004). Narrow spacing produces crowding, and wide spacing alleviates it. For fMRI measurements, however, manipulating spacing would introduce important confounds: the different stimulus configurations would elicit responses in different retinotopic neural subpopulations, and for a given retinotopic region, the content of visual stimulation integrated over time would differ between the two configurations.

Therefore, we instead used a temporal manipulation to control crowding, analogous to previous studies of surround suppression and selective attention in cluttered scenes (Kastner et al., 2001; Zenger-Landolt & Heeger, 2003). In the main experiment, letters were presented at 1 Hz for a duration of between 15 and 21 s separated by between 15 and 21 s of no stimulation. Target and flanker letters were displayed in alternation in the uncrowded trials and simultaneously in the crowded trials (Figure 1B). At each small location within the visual field, the overall visual stimulation (integrated over time) was the same for both simultaneous and sequential presentations, but crowding only occurred for the simultaneous presentation. In a separate psychophysics experiment, we confirmed that with this temporal manipulation letter identification performance was impaired by a factor of two or more when target and flanker letters were presented simultaneously (Figure 1C; $p < 0.01$, $n = 4$ subjects, paired t-test). During fMRI, subjects performed either a near-threshold contrast discrimination task (Main experiment section and Grating control experiment section) or a highly demanding letter identification task (Diverted attention control experiment section) at fixation.

Stimulus-driven responses

Mean stimulus-driven responses were estimated using deconvolution (see Methods section), separately for crowded and uncrowded trials of each stimulus duration, for each visual cortical area from each subject. Following Baker et al. (2007), we identified the visual word form area in each subject by measuring cortical responses to alternating blocks of English and Chinese characters (Figure 2A). We acknowledge that there may be other nearby (or overlapping) cortical regions preferentially responding to visual word stimuli (Cohen & Dehaene, 2004), but we refer to the region we identified in the occipitotemporal cortex as the "visual word form area" (VWFA; see Methods section for further discussion). The location of the identified area was consistent across subjects (Talairach coordinates: −43 ± 4, −70 ± 7, −12 ± 10; mean ± standard deviation across $n = 4$ subjects).

Retinotopic mapping was used to identify visual cortical areas V1, V2, V3, and V4 and to identify subregions of each visual area corresponding to the locations of the target letters (Figures 2B and 2C, see Methods section). The time courses of the responses in each of these visual areas reflected the different stimulus durations (Figure 3A). Response amplitudes were larger in areas V1–V4 than they were in VWFA (Figure 3B), but VWFA responses were clearly above baseline, and rescaling VWFA responses showed that they too reliably reflected the stimulus durations (Figure 3A).

Hence, although the VWFA was defined using foveally presented letters, it responded to the peripherally presented letters as well. Two factors may account for the relatively small VWFA responses to peripheral letters. First, both letters and faces show an eccentricity bias with stronger foveal representations (Hasson, Levy, Behrmann, Hendler, & Malach, 2002). Second, subjects were not attending nor explicitly identifying (naming) the letters but were instead performing a non-letter-based task at fixation (see Methods section). VWFA response amplitudes were significantly larger when subjects performed a letter-based task at fixation (Figure 7D; see Correlation differences were not caused by attention section).

Response amplitudes in V2, V3, and V4 were significantly larger for uncrowded than crowded letters (Figure 3B; V2: $p = 0.013$; V3: $p < 0.0001$; V4: $p < 0.0001$; two-sided permutation test, $n = 4$ subjects). However, in V1 and in VWFA we did not find evidence that response amplitudes were significantly different for crowded versus uncrowded letters (V1: $p = 0.94$; VWFA: $p = 0.11$). The lack of evidence for an effect of crowding on mean responses in VWFA may seem surprising, given that the responses of such high-level category-selective areas have often been suggested to reflect object perception (Beck, Rees, Frith, & Lavie, 2001; Dehaene et al., 2001; Grill-Spector, Knouf, & Kanwisher, 2004; Grill-Spector, Kushnir, Hendler, & Malach, 2000; Summerfield, Egner, Mangels, & Hirsch, 2006; Tong, Nakayama, Vaughan, & Kanwisher, 1998).

Correlations between visual areas were lower for crowded letters

The primary goal of our study was to test whether crowding affected the dynamic interactions between cortical areas, as measured by the correlation between their fMRI time series (Friston, 2002; Haynes, Driver, & Rees, 2005; Haynes, Tregellas, & Rees, 2005). We report the results of two complementary procedures for characterizing inter-area correlations. Both were designed to specifically isolate within-trial dynamics (see Methods section), unlike the more standard procedures adopted in the literature (Friston, 2002), which conflate slow between-trial and fast within-trial dynamics. One of our procedures ("concatenated") had more statistical power but involved additional steps of preprocessing. The other procedure ("within-trial") involved less preprocessing of the data but had less statistical power. The driving effects of a stimulus can induce correlations that do not reflect intrinsic neural interactions within the brain (Gerstein & Perkel, 1969). To eliminate this effect of the stimulus-driven responses on the correlations, we removed the stimulus-driven responses from the response time courses of each trial within each visual area, separately for crowded and uncrowded trials of each duration (Figure 4). The two complementary analysis procedures differed in how they removed the stimulus-driven responses (see Methods section).

Correlations were larger for uncrowded than for crowded letters (Figure 5B). Correlation differences were particularly large between early retinotopic areas and VWFA (V1–VWFA: $p = 0.018$; V2–VWFA: $p = 0.016$; V3–VWFA: $p = 0.033$; V4–VWFA: $p = 0.016$; one-sided permutation test, within-trial analysis, combining data across $n = 4$ subjects, see Methods section). These differences were also highly significant in the concatenated version of the analysis (V1–VWFA: $p < 0.001$; V2–VWFA: $p < 0.001$; V3–VWFA: $p = 0.001$; V4–VWFA: $p < 0.001$; one-sided permutation test, see Methods section). We will focus on these effects including the VWFA for most of the Results section, but see below (No evidence of correlation differences for grating patches section) for a discussion of effects among early visual areas.

The change in inter-area correlations due to crowding was approximately 10%, relative to the "baseline" level of correlation in the uncrowded condition. The mean difference in correlation across pairs between V1–V4 and VWFA was 0.045 compared to a baseline correlation of 0.51 (for the uncrowded condition). This magnitude of correlation difference is expected, because correlations between fMRI time series are likely driven by multiple sources of variability, many of which are unrelated to neuronal stimulus processing. Non-neural sources of variability may include instrumental noise, head motion (and motion correction algorithms), respiration and heartbeat, and the mechanisms of neurovascular coupling. In addition, the coherent fluctuations of neural activity driving correlations between activity in distant cortical areas seem to include strong components unrelated to stimulus processing, such as common input from subcortical structures (Leopold et al., 2003). Thus, any crowding-related changes in inter-area correlation are expected to be small relative to the baseline level of correlation.

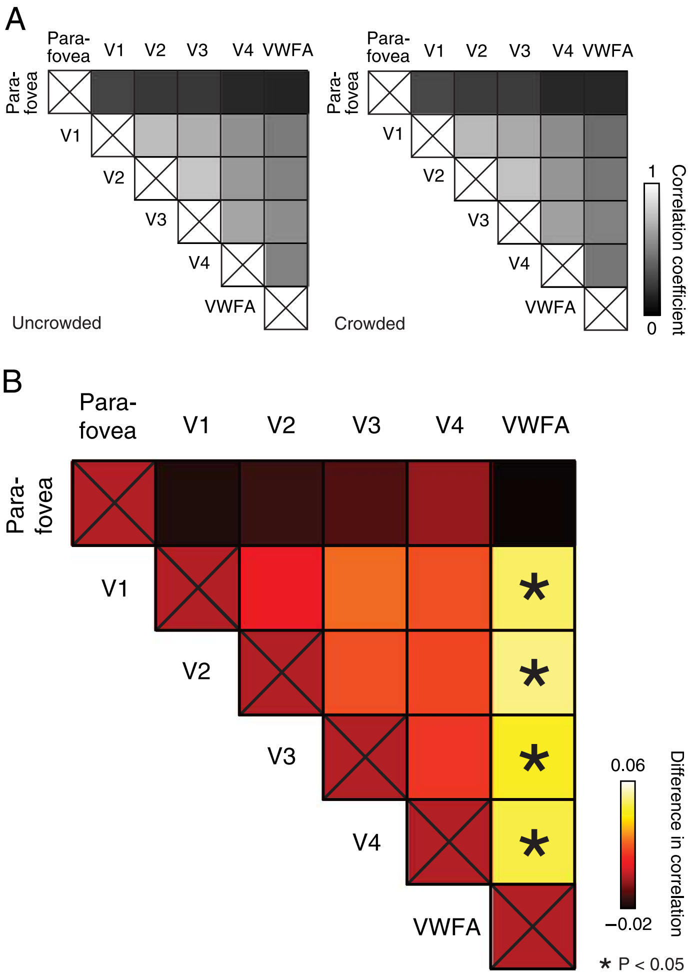

Figure 5. Crowding changed the response correlations between pairs of visual areas. (A) Correlation coefficients for (left) uncrowded trials and (right) crowded trials. Each square indicates the correlation between residual time courses (within-trial analysis combining trials across $n = 4$ subjects), for a pair of ROIs (see Methods section). (B) Differences in correlations (uncrowded − crowded). Asterisks represent ROI pairs showing a statistically significant difference in correlation (*$p < 0.05$, one-sided permutation test, within-trial analysis combining trials across $n = 4$ subjects). The statistical significance of the V1–VWFA, V2–VWFA, V3–VWFA, and V4–VWFA correlation differences was confirmed in a complementary analysis using paired *t*-tests, i.e., treating inter-subject variability as a random factor (see Methods section and Results section).

The magnitude of the change in inter-area interactions was twice as large when quantified in terms of covariance rather than correlation. As discussed below (see Correlation differences were not driven by differences in response amplitude or variability section), the effects of crowding on correlation did not reflect differences in within-area response variability but instead reflected robust changes in the covariance in activity between cortical areas. In fact, differences in variability weakened the effect of crowding on correlations. Inter-area covariances were 20% smaller for crowded than for uncrowded letters (mean difference in covariance across pairs between V1–V4 and VWFA was 0.02 compared to a baseline covariance of 0.09 in the uncrowded condition). Covariance differences were also highly significant in both versions (within-trial and concatenated) of the analysis ($p < 0.01$ for all VWFA pairs).

Correlation differences were retinotopically specific to the target-related subregions of each visual area (Haynes, Driver et al., 2005); they were not detectable in a separately defined parafoveal subregion of visual areas V1–V4 (Figure 5B, top row; parafovea–VWFA: $p = 0.82$, within-trial analysis; $p = 0.82$, concatenated analysis). As an additional test of retinotopic specificity, we defined a control region of interest by taking the entire retinotopically defined V1 and excluding only the target-related subregion (Figure 2C), thus yielding a non-target subregion of V1. There was no evidence for correlation differences for the pairing of this non-target subregion of V1 with VWFA ($p = 0.86$, within-trial analysis; $p = 0.65$, concatenated analysis). These tests confirm that the effect was retinotopically specific to an area near the target subregion, ruling out confounds related to global changes in arousal or attention. However, our conclusions do not require that our target subregions exclusively responded to the targets and not at all to the flankers. First, in our experimental design, the overall visual stimulation (integrated over time) within the target ROI was the same for both simultaneous and sequential presentations, so any difference within that ROI specifically reflects simultaneous presentation. Second, we expected crowding to be weaker or absent for the flankers than the target. Thus, the effect of crowding on inter-area correlations should be stronger if the ROI better isolated the target. Evidence suggests that localizers like the one used here can isolate neural representations of stimuli as small as our target letters, at conventional scanning resolutions (Donner, Sagi, Bonneh, & Heeger, 2008). However, because our ROI might have included the flankers, we interpret the effect of crowding that we found in the target ROI as a lower bound on the possible magnitude of the change in inter-area correlations that might be obtained with an even better localizer.

To ensure that these results were not driven by only a subset of the subjects, we performed a complementary analysis, treating inter-subject variability as a random factor (after first performing the within-trial analysis on each subject separately). This variant of the analysis revealed statistically significant differences in correlation coefficients (uncrowded correlations > crowded correlations) for pairings between V1, V2, and V4 and the VWFA (V1–VWFA: $p = 0.0007$; V2–VWFA: $p = 0.0028$; V3–VWFA: $p = 0.12$; V4–VWFA: $p = 0.025$; one-sided paired *t*-test, $n = 4$ subjects) but not for the parafoveal subregion (parafovea–VWFA: $p = 0.76$).

In a complementary version of the within-trial analysis, correlations were computed without first removing the mean stimulus-driven responses. This analysis was performed on the raw time courses, and thus minimized the preprocessing applied to the data. The correlation differences were similar to what was observed with the residual responses and again significant ($p < 0.05$) for all VWFA pairs. However, we focused on the results of the primary two analyses that isolated within-trial neural dynamics as opposed to trial-to-trial fluctuations in activity.

The effect of crowding on inter-area correlations was particularly robust, across all of these analyses, for the pairing of V1 and VWFA, which were the two visual areas most distant (both physically and in terms of the visual cortical hierarchy) among those studied. Because crowding selectively interferes with feature integration (see Introduction section for references), and because correlations in activity between visual areas were lower for crowded than for uncrowded letters, we conjecture that the inter-area correlations reflect some aspect of the feature integration process. This conjecture is supported by various control experiments and analyses presented below.

Correlation differences were not driven by differences in response amplitude or variability

Crowding affected the correlation between activity in VWFA and V1, not their mean response amplitudes (Figure 3B). Although we removed the stimulus-driven responses before computing correlations, a potential concern is that larger correlations between areas in the uncrowded trials might have reflected larger responses and signal-to-noise ratio to uncrowded letters than to crowded letters (as evident in V2–V4, see Stimulus-driven responses section above). However, as discussed above, in V1 and VWFA we did not find evidence that response amplitudes were significantly different for crowded versus uncrowded letters, even though this pair of areas showed one of the largest differences in correlation. This confirms that the differences in inter-area correlation did not simply reflect differences in mean response amplitudes (and thus signal-to-noise ratio).

The effect of crowding on correlations did not reflect changes in response variability (Arieli, Sterkin, Grinvald, & Aertsen, 1996; Faisal, Selen, & Wolpert, 2008), which may correlate with perceptual performance (Donner et al., 2007; Ress, Backus, & Heeger, 2000). Residual response variances were either indistinguishable between crowded and uncrowded letters or larger for uncrowded letters (Figure 3C; V1: $p = 0.33$; V2: $p = 0.27$; V3: $p = 0.061$; V4: $p = 0.23$; VWFA: $p = 0.023$; two-sided permutation test, computing variances separately within each trial, then combining trials across $n = 4$ subjects). Larger variances for uncrowded letters without correspondingly larger covariances would have yielded lower correlation coefficients, whereas we observed the opposite. Thus, our results do not reflect differences in within-area response variability but instead reflect robust changes in the covariance in activity between cortical areas.

No evidence of correlation differences for grating patches

Letters were used as the stimuli in the main experiment to characterize inter-area correlations related to crowding and feature integration; letters consist of features that are integrated into complex wholes during recognition, and consequently, letter recognition suffers from crowding (Pelli et al., 2004). We used a temporal manipulation rather than a spatial manipulation so as to equate stimulus content (integrated over time) between the crowded and uncrowded conditions. Perhaps, however, the significant effects of crowding on inter-area correlations were due to confounding aspects of the temporal manipulation rather than crowding per se. For example, the temporal manipulation likely produced apparent motion during sequential presentation of targets and surround suppression during simultaneous presentation. The two conditions also differed with respect to the number and timing of stimulus onsets.

We performed a control experiment with simple grating patch stimuli to rule out all of these possible confounds. We hypothesize that the automatic task-independent neural processing of stimuli depends on their complexity, with only more complex stimuli engaging feature integration. Letters contain multiple features that must be integrated during recognition, whereas grating patches contain a single feature that can be detected by a single orientation-tuned channel. As a result, we should be able to identify a behavioral difference in the effect of crowding on the two kinds of stimuli. Furthermore, if the effects on neural processing observed in the main experiment (smaller inter-area correlations for crowded than for uncrowded letters) reflected an interference in feature integration due to crowding, then there should be no such differences in correlations for the gratings. If the results of the main experiment were instead due to any other aspect of the temporal manipulation, correlation differences should also be present for grating patches.

In the control experiments with gratings, subjects viewed triplets of parallel grating patches with identical orientations (Gaussian-windowed two-dimensional sinusoids at 100% contrast) instead of triplets of letters, but all other aspects of the experiment and analysis were identical to the main experiment. We performed a two-alternative forced-choice psychophysical experiment to confirm that coarse grating orientation was readily identifiable for simultaneous and sequential presentations using these stimuli (Figure 6A).

We found no evidence for a difference in coarse orientation discrimination between the two kinds of presentation ($p = 0.2863$, $n = 3$ subjects, two-sided paired -test). To obtain psychophysical data for letters in a more comparable task, we performed another psychophysical experiment using letters in which observers judged whether the center letter in a triplet of identical letters was mirror-reversed or not. In this experiment, performance was significantly lower for simultaneous presentation ($p = 0.015$, paired t-test, $n = 3$ subjects; Figure 6B). Thus, simultaneous presentation impaired identification for letters but not gratings, with a comparable stimulus manipulation and 2AFC task.

In the fMRI experiment with gratings, inter-area correlations were indistinguishable for simultaneous versus sequential presentation of grating patches (Figure 7A). As reported above, letter crowding had a statistically significant effect on the correlations for all four pairings between V1–V4 and VWFA, but this was not the case for the gratings (V1–VWFA: $p = 0.89$; V2–VWFA: $p = 0.58$; V3–VWFA: $p = 0.66$; V4–VWFA: $p = 0.16$, one-sided permutation test, within-trial analysis, combining across $n = 4$ subjects). Nor did we observe a significant effect of simultaneous versus sequential presentation for the gratings in the concatenated version of the analysis (V1–VWFA: $p = 0.92$; V2–VWFA: $p = 0.89$; V3–VWFA: $p = 0.90$; V4–VWFA: $p = 0.58$, one-sided permutation test, concatenating across $n = 4$ subjects). We additionally found that the effect of crowding on correlations (i.e., the difference in correlation between simultaneous and sequential presentations) was statistically significantly larger for letters than it was for the grating patches for pairs V1–VWFA and V2–VWFA (the bottom 5th percentile of the bootstrapped distribution of the effect size difference between letters and gratings exceeded 0, within-trial analysis, see Methods section for details). These effect size differences were significant in all four pairs for the concatenated version of the analysis. Thus, correlations were lower during simultaneous presentation of targets and flankers than during sequential presentation when subjects viewed letters but not when they viewed gratings.

Simultaneous presentation evoked smaller mean response amplitudes in areas V2, V3, and V4 for gratings (Figure 7B; V1: $p = 0.63$; V2: $p = 0.023$; V3: $p = 0.020$; V4: $p = 0.0001$; VWFA: $p = 0.95$, two-sided permutation test), just as it did for letters. The smaller response amplitudes induced by simultaneous presentation, for both stimuli, were likely due to other aspects of the manipulation, such as surround suppression (Zenger-Landolt & Heeger, 2003), whereas the correlation differences, which occurred for letters and not gratings, reflected the effect of crowding on feature integration. Previous studies have found different results concerning surround suppression in V1 as measured by fMRI. Studies have consistently found evidence for surround suppression beyond V1 (Kastner et al., 2001; Zenger-Landolt & Heeger, 2003), but two studies found no evidence for surround suppression in V1 (Fang & He, 2008; Kastner et al., 2001), whereas one did (Zenger-Landolt & Heeger, 2003). This inconsistency is likely due to the fact that Zenger-Landolt and Heeger (2003) designed their stimulus to maximize the effect of surround suppression, which was achieved through the use

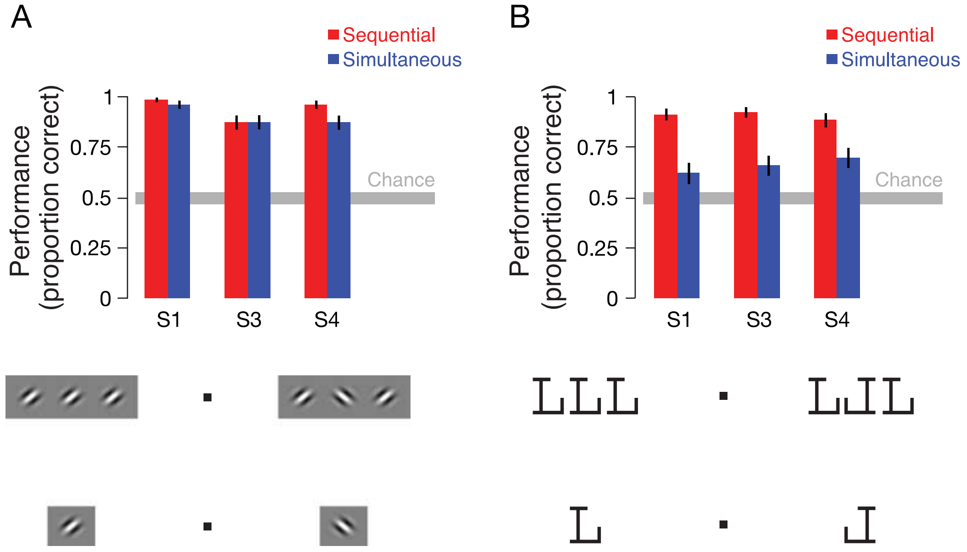

Figure 6. Grating and mirrored letter psychophysics. (A) A target grating was presented between two oriented flanker gratings. The two flankers shared the same orientation. Subjects judged whether the orientation of the target grating orientation was parallel or orthogonal to that of the flankers (chance = 0.5, gray bar). Targets and flankers were presented either simultaneously or sequentially. The demos below the graph show that orientation discrimination is easy whether or not there are flankers. There was no evidence for a difference in orientation discrimination performance for simultaneous and sequential presentations (p = 0.29, paired t-test, n = 3 subjects). Error bars are 68% confidence intervals estimated from the bootstrapped distribution of proportion correct. (B) A target letter was presented between two flanker letters. The two flanker letters shared the same identity. Subjects judged whether the target letter was mirror-reversed or not. The demos below the graph show that this is difficult when there are flankers but easy when the flankers are taken away. Performance was significantly better for sequential than for simultaneous presentation (p = 0.015).

of a large surround, much larger than a pair of flanking letters. The target stimuli were also larger than the letters in the present study and closer to the fovea. We suspect that these features of Zenger-Landolt and Heeger's study produced surround suppression in V1 that was stronger than that observed in the other studies.

Response amplitudes evoked by the gratings in VWFA were small but statistically greater than zero (Figure 7B, $p < 0.0001$, two-tailed t-test comparing mean responses to 0, combining trials across n = 4 subjects), which is consistent with the hypothesized functional specialization of this area. A possible concern is that the different results with letters versus gratings were due to this difference in the VWFA mean response amplitudes.

However, several pieces of additional evidence suggest that the differences in mean response amplitudes alone do not explain the inter-area correlation results. First, in all experiments, we removed the stimulus-driven responses before computing correlations. Second, we show below (Correlation differences were not caused by attention section) that the effects of letter crowding on inter-area correlations were reliably present in two experiments, one with small VWFA responses and one with large responses, indicating that the smaller inter-area correlations with crowding were robust to changes in the mean VWFA response. Third, although the differences in correlation (for sequential versus simultaneous presentation) were larger for letters than for gratings, the raw correlation values (i.e., for sequential presentation only) were comparable for the two stimuli. In particular, the average correlation between V1–V4 and VWFA for sequential presentation was 0.51 for letters and 0.49 for gratings. Fourth, gratings evoked large response amplitudes in early visual areas, but the inter-area correlations between these areas supported the same conclusion. The response amplitudes in V1–V4 were large for both gratings and letters, whether presented simultaneously or sequentially. Grating-evoked responses were about 25% smaller than letter-evoked responses in these early visual areas (compare Figures 3B and 7B), but about ten times larger than the grating-evoked responses in VWFA (Figure 7B). Inter-area correlations were significantly lower for letter crowding in all three pairings between V1 and V2–V4 (V1–V2: p = 0.030; V1–V3: p = 0.0006; V1–V4: p = 0.016, one-sided permutation test, concatenated analysis, combining across n = 4 subjects). For gratings, however, there was no evidence for a difference in inter-area correlations between simultaneous and sequential presentations (V1–V2: p = 0.37; V1–V3: p = 0.46; V1–V4: p = 0.64, concatenated analysis). The correlation difference (i.e., the effect size) was statistically significantly larger for letters than for gratings in all three pairings (bottom 5th percentile of the bootstrapped distribution of the effect size difference between letters and gratings exceeded 0, concatenated analysis, n = 4). The correlation differences (for letters) among early areas were not as robust as those between early visual areas and the VWFA; for the within-trial analysis (which had less statistical power compared to the concatenated analysis, see Methods section), the differences in inter-area correlations among early areas were larger for letters than for gratings, but neither were statistically significant (Figure 5). Even so, these results suggest that letter crowding affected inter-area correlations among early areas, but there was no evidence of this with gratings, even though mean response amplitudes to gratings in early areas were comparable to those for letters (Figures 3B and 7B). All of these pieces of evidence together suggest that the differences between letters and gratings were not a trivial consequence of the small response amplitudes to gratings in the VWFA.

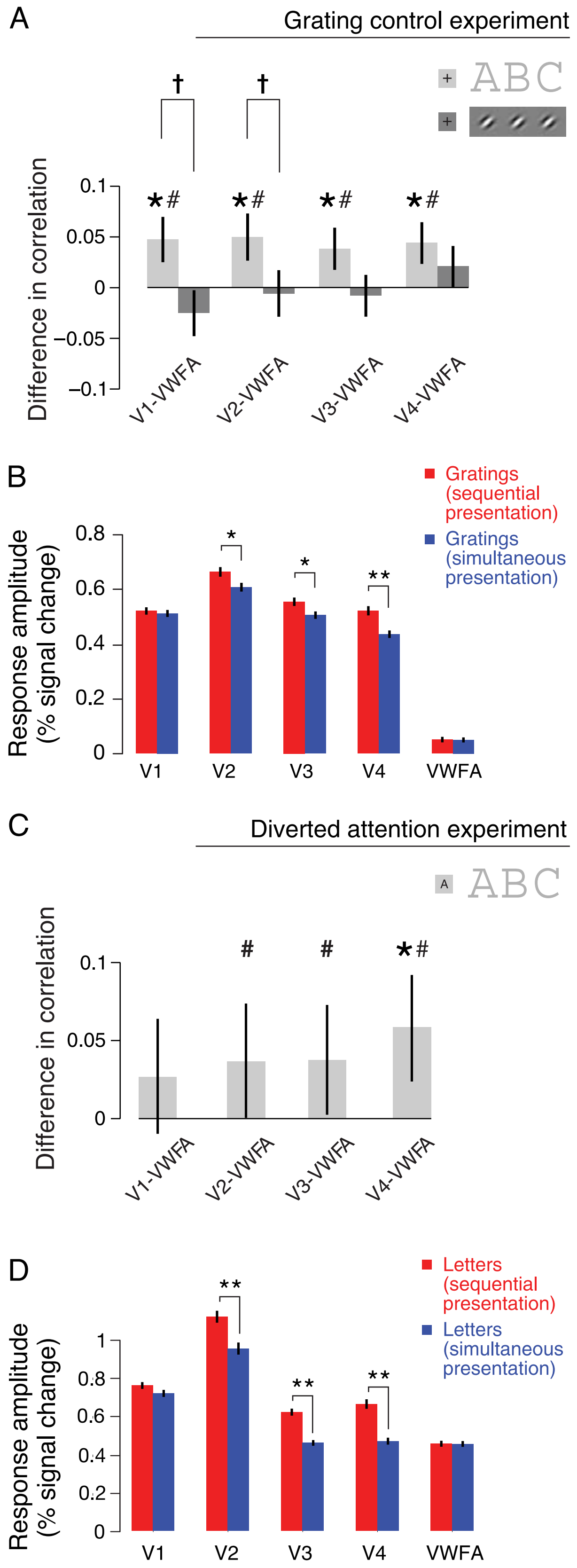

Figure 7. Grating and diverted attention control experiments. (A) Inter-area correlations were indistinguishable for simultaneous versus sequential presentation of grating patches, which contain only a single feature, unlike letters, which contain multiple features that must be integrated. Differences in correlations (for sequential versus simultaneous presentations) are shown for the letter experiment (light gray bars) and the control experiment with gratings (dark gray bars), for the four ROI pairs that included VWFA. The total number of trials (530) was identical for letters and gratings, ensuring equal statistical power in the two experiments. Error bars are 68% confidence intervals estimated from the bootstrapped distribution of correlation differences (within-trial analysis). Asterisks represent ROI pairs showing a statistically significant difference in correlation (*$p < 0.05$, one-sided permutation test, within-trial analysis, combined across $n = 4$ subjects, analyzed separately for the main letter experiment and the grating control experiment). Number signs represent ROI pairs showing a statistically significant difference in correlation in a complementary (concatenated, see Methods section) version of the analysis ($^{\#}p <$ 0.05, one-sided permutation test, concatenating across $n = 4$ subjects). Daggers represent ROI pairs showing a statistically significantly larger difference in correlation for letters than for grating patches ($^{\dagger}p <$ 0.05, one-sided bootstrap test, within-trial analysis, combined across $n = 4$ subjects). (B) Mean fMRI response amplitudes in the grating control experiment. Error bars are 68% confidence intervals estimated from the bootstrapped distribution of mean response amplitudes. Asterisks indicate statistically significant differences in response amplitudes (*$p <$ 0.05 and **$p < 0.001$, one-sided permutation test, $n = 4$). (C) Inter-area correlation differences for letters persisted when attention was diverted. Differences in correlations (uncrowded − crowded) are shown for the diverted attention control experiment for the four ROI pairs that included VWFA ($n = 3$ subjects). Same conventions as in (A). (D) Mean fMRI response amplitudes in the diverted attention control experiment. Same conventions as in (B).

Correlation differences were not caused by attention

When characterizing the neural correlates of specific perceptual processes, it is important to dissociate them from the neural correlates of attention (Huk, Ress, & Heeger, 2001; Koch & Tsuchiya, 2007; Lamme, 2003). Attention strongly modulates the amplitudes (Kanwisher

& Wojciulik, 2000; Ress et al., 2000) and inter-area correlations (Haynes, Driver et al., 2005) of fMRI responses throughout human visual cortex. Despite performing a contrast decrement task at fixation in the main experiment, subjects' residual attention may have been drawn more to the uncrowded letters than to the crowded letters in the periphery. Any fluctuations in attention during the stimulus presentations would have evoked correlated modulations in cortical activity across multiple visual areas. For example, if subjects tended to shift attention back and forth between fixation and the uncrowded letters but not the crowded letters, then this could have resulted in higher inter-area correlations for uncrowded than for crowded letters.

To rule out any such attentional confound, we performed a control experiment in which subjects performed a highly demanding letter identification task at fixation. Subjects counted Xs in a stream of rapidly presented letters (4 or 5 Hz) for the duration of each trial. Meanwhile, target and flanker letters were presented in the periphery, either sequentially or simultaneously, as in the main experiment. Psychophysical measurements using a dual-task protocol outside the scanner established that subjects had no residual attentional resources to allocate differentially to the peripheral letters (Figure 8).

This demonstrated that the fixation task was highly demanding. If subjects still attended differentially to the crowded and uncrowded stimuli during scanning, then this would have produced a difference in performance for the letter identification task at fixation. Performance in the task was, however, indistinguishable between the crowded and uncrowded trials during fMRI (two-sided permutation test; $p > 0.5$ for each subject), confirming that subjects' attention was fully diverted from the letter stimuli.

With attention fully diverted, inter-area correlations for all four pairings with VWFA were lower for crowded than for uncrowded letters (Figure 7C). This effect was significant for V4–VWFA in the within-trial analysis (V1–VWFA, $p = 0.24$; V2–VWFA, $p = 0.156$; V3–VWFA, $p = 0.14$; V4–VWFA, $p = 0.047$, one-sided permutation test, combining across $n = 3$ subjects) and significant for V2–VWFA, V3–VWFA, and V4–VWFA in the concatenated analysis (V1–VWFA, $p = 0.12$; V2–VWFA, $p = 0.015$; V3–VWFA, $p = 0.029$; V4–VWFA,

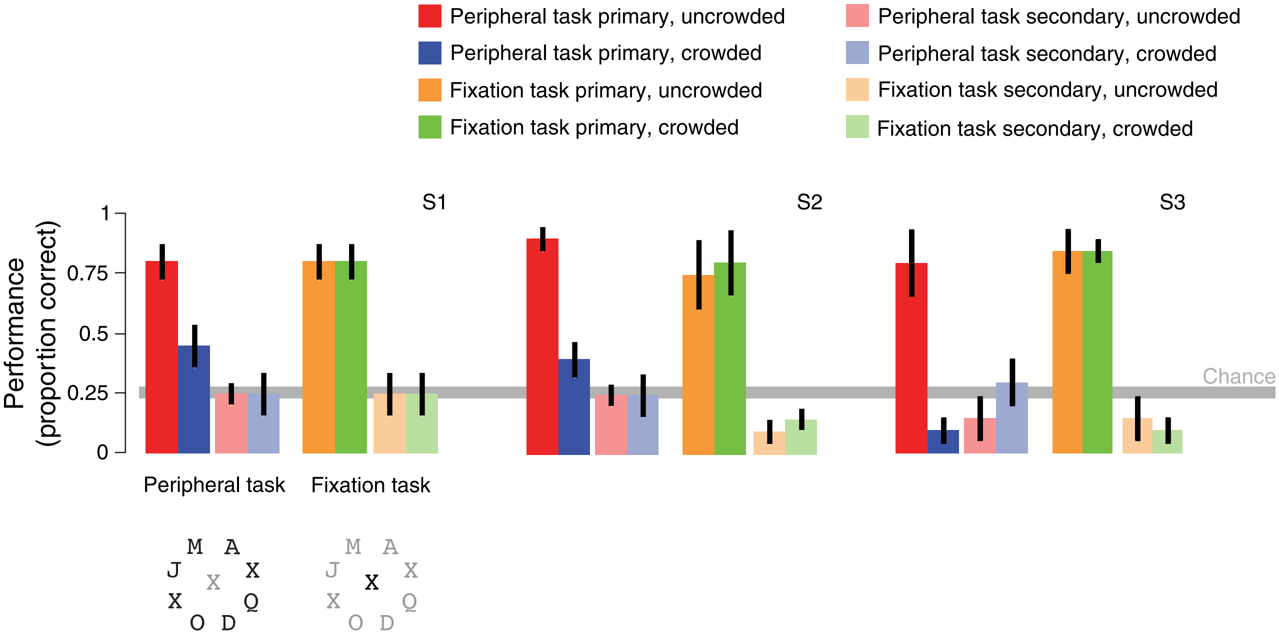

Figure 8. Diverted attention psychophysics. Three subjects performed a dual-task experiment outside the scanner. Subjects counted the number of frames containing Xs among the peripheral letters (*Peripheral task*) and the number of Xs appearing at fixation (*Fixation task*). For each block of trials, one of the two tasks was designated the primary task. The peripheral letters were either crowded (simultaneous presentation) or uncrowded (sequential presentation). Subjects were instructed to first respond to the primary task, and then respond to the secondary task. Chance performance was 0.25. For the peripheral task, conditions were uncrowded and primary (dark red), crowded and primary (dark blue), uncrowded and secondary (light red), and crowded and secondary (light blue). Similarly, for the fixation task, conditions were uncrowded and primary (dark yellow), crowded and primary (dark green), uncrowded and primary (light yellow), and crowded and primary (light green). For the fixation task, subjects performed well when it was primary and poorly when it was secondary, but there was no effect of crowding in either case. For the peripheral task, when primary, subjects performed well for uncrowded letters and performance was significantly lower for crowded letters ($p = 0.036$, two-sided paired t-test, $n = 3$). However, when the peripheral task was secondary (and subjects' attention was instead diverted to the fixation task), peripheral performance was indistinguishable from chance for both crowded and uncrowded letters ($p > 0.1$, two-sided one-sample t-test comparing performance to chance, 0.25, $n = 3$), confirming that the fixation task effectively diverted attention from the peripheral stimuli.

$p = 0.01$; one-sided permutation test, concatenating across $n = 3$ subjects). As in the main experiment, there was no evidence for a correlation difference in the parafoveal subregion (parafovea–VWFA, $p = 0.49$, within-trial analysis; $p = 0.39$, concatenated analysis).

Because of the letter identification task at fixation, mean response amplitudes in VWFA were almost twice as large as they were in the main experiment ($p = 0.0001$, two-sided permutation test) and comparable to responses in early visual areas (Figure 7D). However, as before, response amplitudes in VWFA were indistinguishable for crowded and uncrowded letters ($p = 0.96$, two-sided permutation test). Therefore, correlations between activity in VWFA and early visual areas reflected crowding of the peripheral letters, regardless of the response amplitudes. The large responses to the foveal stimuli were constant across crowding conditions, whereas the effect of crowding on correlations isolated a component of the signal reflecting the cortical representation of the peripheral letters.

Discussion

The central conclusion from our results is that feature integration is reflected in response correlations between object-selective cortical areas and earlier, feature-selective areas such as V1. We found that: (1) correlations between responses in early retinotopic areas and the VWFA were lower for crowded than for uncrowded letters; (2) correlation differences were retinotopically specific to the peripheral targets, so they were not caused by spatially non-specific modulations of correlated activity such as arousal; (3) correlation differences were not observed when the letters were replaced with grating patches, ruling out low-level confounds (such as surround suppression) due to simultaneous versus sequential stimulus presentations but unrelated to crowding and feature integration; and (4) correlation differences persisted (and were of similar magnitude) when attention was diverted from the letter stimuli, implying that they were not caused by modulation of attention.

We have used crowding as a tool to identify a neural correlate of feature integration. Several different accounts of crowding have been proposed (see Levi, 2008; Pelli & Tillman, 2008 for reviews), but our experiments were not designed to distinguish between models of crowding. Rather, we interpret our results in the context of a set of assumptions and in terms of current computational models of feature integration and object recognition. These assumptions and models are discussed in detail in the following paragraphs, but we list them here to preview the logic. (1) Objects are encoded through a largely automatic task-independent process as proposed in several hierarchical models of object processing in the ventral stream. (2) This hierarchical process performs feature integration automatically for complex objects (e.g., letters) that contain multiple features. (3) Feature integration does not automatically occur for simple objects (e.g., gratings) that contain only one feature, but feature integration might nonetheless be required depending on the task (e.g., fine orientation discrimination). (4) Crowding causes inappropriate or improper feature integration that disrupts recognition without affecting stimulus visibility. (5) This effect of crowding is unstable such that the percept and the corresponding neural representation vary dynamically over time. (6) The unstable neural processing reduces intrinsic correlations between the neuronal activities in different visual cortical areas. (7) This reduction in correlated neuronal activity is reflected as lower inter-area correlations in fMRI responses. Our conclusions follow from these assumptions. In both the fMRI and the psychophysics experiments, we found different results for letters and gratings. For letters, behavioral performance and inter-area correlations were both affected by crowding, but for gratings, neither was affected by crowding. Given that crowding specifically interferes with feature integration, we infer from our results with letters that inter-area correlations reflect feature integration. The fact that crowding did not affect behavioral performance or inter-area correlations for gratings is consistent with our hypothesis that the effects of crowding on the automatic processing of a stimulus depends on the complexity of the stimulus.

We assume that objects are encoded through a largely automatic task-independent process, as proposed in several hierarchical models of object processing in the ventral stream (Riesenhuber & Poggio, 1999). Following this automatic encoding process, the neural representation of the stimulus must also be decoded through read-out mechanisms, which likely depend on the task. In our psychophysical experiments, with both letters and gratings, 2AFC discrimination performance assessed the neural representations available for read-out. In our fMRI experiments, on the other hand, we had subjects perform a fixation task in the scanner so as to avoid confounds related to performance or arousal. We assume that this isolated the automatic component of the stimulus processing in the visual cortex, thereby measuring the effects of crowding on that processing regardless of the task or read-out.

We further hypothesize that the automatic encoding of objects depends on the complexity of the stimulus. Letters are complex, meaning that they consist of multiple elementary features. We assume that such complex, but familiar, stimuli engage automatic feature integration mechanisms that are independent of task and attention. We assume, on the other hand, that the automatic processing of simple stimuli like gratings does not necessarily involve feature integration. We found that performance in a coarse orientation discrimination task (0° vs. 90° rotation) was not affected by crowding, when using the same stimulus manipulation that yielded a strong crowding effect for a 2AFC letter identification task (Figure 6). Pelli et al. (2004) also found no effects of crowding when observers were asked to identify the coarse orientation of a target grating (±45°) presented between two oriented flanker gratings. They concluded, like us, that such a coarse identification task might be more akin to detection, unaffected by crowding because it does not rely on feature integration, and can instead be performed by reading out from a single orientation-tuned channel. Other experiments have identified crowding effects for coarse (i.e., ±45°) orientation discrimination tasks (He, Cavanagh, & Intriligator, 1996; Petrov, Popple, & Mckee, 2007), though they used different kinds of flankers and presentation parameters: He et al. (1996) used four flanker gratings instead of two and presented stimuli at nearly twice the eccentricity used here, and Petrov et al. (2007) used a plaid flanker with higher contrast than the target. There are also reports of crowding effects on fine orientation discrimination, as assessed by measuring orientation discrimination thresholds (Levi & Carney, 2009; Wilkinson, Wilson, & Ellemberg, 1997). Fine orientation discrimination likely requires combining information from multiple orientation-selective channels, which is a form of feature integration, and thus we hypothesize that performance on these tasks might suffer from crowding due to inappropriate feature integration. The key fact for interpreting our present results is that, under conditions well matched to our fMRI experiment, we found crowding effects for letters but not gratings. These psychophysical results are consistent with our fMRI findings: a decrease in inter-area correlations for letters when presented simultaneously with flankers, but no effect of simultaneous presentation on inter-area correlations for gratings.

Crowding specifically interferes with feature integration (i.e., the transformation of simple feature representations into the complex representations underlying object recognition) without affecting stimulus visibility (Pelli et al., 2004). Many previous fMRI studies of object perception have manipulated visibility to identify neural correlates of object perception (Beck et al., 2001; Dehaene et al., 2001; Grill-Spector et al., 2004, 2000; Summerfield et al., 2006; Tong et al., 1998). In contrast, our use of crowding allowed us to establish a more specific link between feature integration and neural activity. In interpreting our results, we assume that the automatic, task-independent processing of the stimuli depends on the spacing between target and flankers, as predicted by "bottom-up" theories of crowding (Levi, 2008; Pelli et al., 2004; though we acknowledge that this is not the only theory of crowding). When object features are spaced too closely together, there is inappropriate feature integration, which impairs recognizability but not detectability.

Why might inappropriate feature integration reduce inter-area correlations? Most models of object recognition describe feature integration as a cascade of transformations along the ventral visual pathway, possibly with feedback. In the simplest case (Riesenhuber & Poggio, 1999), a large set of linear filters extract features, and each successive cortical area integrates features by a non-linear pooling operation. In response to a variety of target objects, the details of this transformation will be different for each target. Any given fixed target, however, will cause a particular set of filters to be the most active, thereby inducing a stable transformation from the responses of a population of neurons in an early stage of the ventral visual pathway to that in a higher level area. This will be evident as a high temporal correlation between neural responses at early and higher levels, which could, in turn, produce a high correlation as measured with fMRI (see below). During crowding, the features, albeit physically constant, may be integrated differently from one moment to the next (see below). If the transformation from feature to object representations varies dynamically, it will disrupt the correlation between early and higher areas. For example, the presence of multiple, closely spaced objects may cause non-linear pooling to dynamically select inappropriate feature combinations. Thus, the transformation from early to higher areas will become unstable, resulting in lower inter-area correlations (as we observed) and the percept of an unstable jumble of features.

The phenomenological instability of the crowded percept has been noted before (Pelli et al., 2004), and although this dynamic aspect of crowding has not been fully characterized experimentally, it can be readily observed in the demo in Figure 1A. While looking at the triplet of letters on the left, you will perceive several plausible identities of the middle letter from moment to moment. Sometimes it appears to be an E, sometimes an H, sometimes an N, and sometimes it is something not present (like a K or an F or a jumble of letter-like features). Our results suggest that psychophysical studies of the temporal dynamics of crowding may provide a tool for probing the mechanisms underlying object recognition.

The neurophysiological basis of the correlations of fMRI time series between cortical areas remains to be determined. Evidence suggests that slow covariation of the amplitude envelopes of voltage fluctuations in different cortical areas drives such fMRI correlations (Leopold et al., 2003; Shmuel & Leopold, 2008). Phase synchronization between these neuronal fluctuations on a faster time scale (Siegel, Donner, Oostenveld, Fries, & Engel, 2008; Womelsdorf et al., 2007) may also play a role, in particular when the modulations in inter-area coupling are spatially specific, as in our case. Regardless of the neurophysiological origin, the sluggishness of the hemodynamic response precludes inferences about the directionality of the underlying neuronal interactions from fMRI measurements. Thus, the relative contribution of feedforward and feedback interactions (Lamme & Roelfsema, 2000) to object recognition remains to be determined in future studies. The current results suggest that it might be fruitful to exploit our experimental approach to studying the neural mechanisms of object recognition in electrophysiological measurements of neuronal coherence between different stages of cortical visual processing.

Our results do not imply that inter-area correlations are directly driving feature integration. They instead show that inter-area correlations are a reliable and specific marker of feature integration that is selectively affected by crowding, and our proposal above provides a plausible interpretation of this marker. Nor do our results imply that decreases in correlation reflect a failure of integration. Based on psychophysical measurements, some have attributed crowding to excessive integration of features over large regions in the periphery (Pelli et al., 2004; although there are other theories of crowding). We propose that features are being integrated during crowding, but excess integration combines the wrong features and fails repeatedly to produce a consistent object percept, and this dynamic and unstable integration process leads to decreased inter-area correlations.

The grating control experiment allows us to rule out potentially confounding aspects of the temporal manipulation (simultaneous versus sequential presentation) that we used to induce crowding. As noted above, we found differences between letters and gratings: for letters, behavioral performance and inter-area correlations were both affected by crowding, but we found no evidence of this for gratings, neither in behavioral performance nor in the fMRI responses. Simultaneous versus sequential presentations differ in many respects including differences in apparent motion, surround suppression, and the timing and number of visual onsets (Kastner et al., 2001; Zenger-Landolt & Heeger, 2003). All of these factors were present both for the gratings and letters. Thus, the results for gratings rule out any such confounds. Furthermore, the differences between gratings and letters are consistent with our hypothesis (outlined above) that the automatic neural processing of stimuli in the visual cortex and the behavioral read-out of the resulting representation depend on both the presentation (simultaneous or sequential) and the complexity (single feature or multiple features) of the stimulus.

Response amplitudes were lower for simultaneous versus sequential presentations, with both letters and gratings, which could, in principle, reflect surround suppression or crowding or a combination of the two. Several previous fMRI studies have also shown that presenting target stimuli in the presence of flankers reduces responses to the targets in human visual cortex (Bi, Cai, Zhou, & Fang, 2009; Kastner et al., 2001; Zenger-Landolt & Heeger, 2003). One study also showed that flanking gratings reduce orientation-specific adaptation of fMRI responses in V2/V3, which was attributed to crowding (Bi et al., 2009). However, a possible alternative interpretation of that result is that surround suppression induced by the flankers weakened the effectiveness of the adaptor, yielding weaker adaptation. Psychophysically, there are many similarities between surround suppression and crowding, and it is difficult to definitively distinguish them, though properties like inward–outward anisotropy may be unique to crowding (Petrov et al., 2007). Using fMRI, we found that unlike the response amplitude differences, inter-area correlation differences were only present for letters. If the automatic processing of stimuli does indeed depend on complexity, then, for our experimental parameters, the response amplitudes of individual visual areas do not discriminate between surround suppression and crowding, whereas inter-area correlations specifically reflect crowding and feature integration.

Our results reinforce the notion that mean response amplitudes within a functionally specialized area may not provide information about the representation in that area. We did not find evidence that response amplitudes in the VWFA were affected by crowding, though they were affected by both the stimulus and the task. Response amplitudes in VWFA were larger for peripheral letters than for peripheral gratings, and response amplitudes were larger when subjects were also viewing and performing a letter-based fixation task than when they performed a contrast discrimination fixation task. In all cases, however, VWFA response amplitudes were indistinguishable between simultaneous and sequential stimulus presentations. In areas V2, V3, and V4, mean response amplitudes were lower for simultaneous presentation, but we observed this difference for both letters and gratings, suggesting that these mean response differences might be due to surround suppression (Zenger-Landolt & Heeger, 2003). Thus, our results challenge the simple notion of functional specialization in which the perception of an object category depends largely on the mean response of a single object-selective area, like the VWFA (Beck et al., 2001; Dehaene et al., 2001; Grill-Spector et al., 2004, 2000; Kanwisher, McDermott, & Chun, 1997). Instead, perception and recognition are likely determined by the relative responses of different neuronal subpopulations interacting across multiple stages of visual processing.

We used crowding of letter stimuli to link feature integration to correlations between early visual areas and the VWFA, but we are not suggesting that crowding in general necessarily involves the VWFA. Rather, we hypothesize that the inter-area correlation differences reported here will generalize to other objects and more complex "features." Psychophysically, crowding has been observed for many non-letter objects, including faces (Martelli et al., 2005; Pelli & Tillman, 2008), and we expect that repeating our experiment with different object classes would yield analogous results but in different brain areas that selectively respond to each object class.

A potential concern is that eye movements might account for the inter-area correlation differences between crowded and uncrowded trials. Although we did not measure eye movements, this explanation seems unlikely for several reasons. First, differential eye movements should have yielded differential performance on the central fixation task, but we found that performance was indistinguishable between crowding conditions. Second, differential eye movements ought to have produced different mean response amplitudes in V1, but response amplitudes were indistinguishable between crowding conditions in V1. Finally, differential eye movements should have yielded responses, and possibly inter-area correlation differences, in the non-target subregions, but we did not observe these effects.

"Top-down" attention is known to boost neural activity across the visual pathway and increase the correlation between responses in these areas. If levels of sustained attention differed between crowded and uncrowded trials, this may have caused a difference in inter-area correlations (Friston, 2002; Haynes, Driver et al., 2005). We controlled attention by using a demanding letter identification task at the center of fixation to divert attention from the peripheral letters, which made it impossible to reliably identify them. When attention was diverted, inter-area correlations were still significantly lower for crowded than for uncrowded letters. This may seem surprising in light of our psychophysical data. Crowding impaired performance when letters were attended, but when attention was diverted, performance was indistinguishable from chance for both crowded and uncrowded letters (Figures 1C and 8). Taken together, this suggests that the neural transformations leading to the representation of highly familiar objects such as letters are largely automatic and dependent on interactions local to the visual cortex (i.e., independent of the state of top-down attention), even though attention is required to make these object representations available for perceptual decision-making and behavioral report.

Acknowledgments

This work was supported by NIH Grant R0I-EY016752 (D.J.H.), by a German Academy of Science Leopoldina Fellowship BMBF-LPD 9901/8-136 (T.H.D.), and by an NSF Graduate Student Fellowship (J.F.). We thank Justin Gardner, Denis Pelli, and Elisha Merriam for helpful comments.

Commercial relationships: none.

Corresponding author: Jeremy Freeman.
Email: freeman@cns.nyu.edu.
Address: Center for Neural Science, New York University, 4–6 Washington Place, New York, NY 10003, USA.

References

Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science, 273,* 1868–1871.

Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 9087–9092.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12):13, 1–18, http://www.journalofvision.org/content/9/12/13, doi:10.1167/9.12.13. [PubMed] [Article]

Beck, D. M., Rees, G., Frith, C. D., & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nature Neuroscience, 4,* 645–650.

Bi, T., Cai, P., Zhou, T., & Fang, F. (2009). The effect of crowding on orientation-selective adaptation in human early visual cortex. *Journal of Vision, 9*(11):13, 1–10, http://www.journalofvision.org/content/9/11/13, doi:10.1167/9.11.13. [PubMed] [Article]

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177–178.

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience, 8,* 1102–1109.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage, 22,* 466–476.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8,* 109–114.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4,* 752–758.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11,* 333–341.

Donner, T. H., Sagi, D., Bonneh, Y. S., & Heeger, D. J. (2008). Opposite neural signatures of motion-induced blindness in human dorsal and ventral visual cortex. *Journal of Neuroscience, 28,* 10298–10310.

Donner, T. H., Siegel, M., Oostenveld, R., Fries, P., Bauer, M., & Engel, A. K. (2007). Population activity in the human dorsal pathway predicts the accuracy of visual motion detection. *Journal of Neurophysiology, 98,* 345–359.

Efron, & Tibshirani. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9,* 292–303.

Fang, F., & He, S. (2008). Crowding alters the spatial distribution of attention modulation in human primary visual cortex. *Journal of Vision, 8*(9):6, 1–9, http://www.journalofvision.org/content/8/9/6, doi:10.1167/8.9.6. [PubMed] [Article]

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience, 8,* 700–711.

Friston, K. (2002). Beyond phrenology: What can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience, 25,* 221–250.

Gerstein, G. L., & Perkel, D. H. (1969). Simultaneously recorded trains of action potentials: Analysis and functional interpretation. *Science, 164,* 828–830.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience, 7,* 555–562.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3,* 837–843.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27,* 649–677.

Hansen, K. A., Kay, K. N., & Gallant, J. L. (2007). Topographic organization in and near human visual area V4. *Journal of Neuroscience, 27,* 11896–11911.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron, 34,* 479–490.

Haynes, J.-D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron, 46,* 811–821.

Haynes, J.-D., Tregellas, J., & Rees, G. (2005). Attentional integration between anatomically distinct stimulus representations in early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 14925–14930.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383,* 3343–3347.

Hubel, D. H. (1982). Exploration of the primary visual cortex, 1955–78. *Nature, 299,* 515–524.

Huk, A. C., Ress, D., & Heeger, D. J. (2001). Neuronal basis of the motion aftereffect reconsidered. *Neuron, 32,* 161–172.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience, 1,* 91–100.

Kastner, S., Weerd, P. D., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology, 86,* 1398–1411.

Kleinschmidt, A., Büchel, C., Hutton, C., Friston, K. J., & Frackowiak, R. S. J. (2002). The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron, 34,* 659–666.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11,* 16–22.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *Journal of Neuroscience, 26,* 13128–13142.

Leopold, D. A., Murayama, Y., & Logothetis, N. K. (2003). Very slow activity fluctuations in monkey visual cortex: Implications for functional brain imaging. *Cerebral Cortex, 13,* 422–433.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48,* 635–654.

Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better. *Current Biology, 19,* 1988–1993.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621.

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*(1):6, 58–70, http://www.journalofvision.org/content/5/1/6, doi:10.1167/5.1.6. [PubMed] [Article]

Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysics Journal, 76,* 691–708.

Nandy, A. S., & Tjan, B. S. (2007). The nature of letter crowding as revealed by first- and second-order classification images. *Journal of Vision, 7*(2):5, 1–26, http://www.journalofvision.org/content/7/2/5, doi:10.1167/7.2.5. [PubMed] [Article]

Nestares, O., & Heeger, D. J. (2000). Robust multiresolution alignment of MRI brain volumes. *Magnetic Resonance in Medicine, 43,* 705–715.

Olman, C. A., Inati, S., & Heeger, D. J. (2007). The effect of large veins on spatial localization with GE BOLD at 3 T: Displacement, not blurring. *NeuroImage, 34,* 1126–1135.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4,* 739–744.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46,* 4646–4674.

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature, 423,* 752–756.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12):12, 1136–1169, http://www.journalofvision.org/content/4/12/12, doi:10.1167/4.12.12. [PubMed] [Article]

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11,* 1129–1135.

Petrov, Y., Popple, A. V., & Mckee, S. P. (2007). Crowding and surround suppression: Not to be confused. *Journal of Vision, 7*(2):12, 1–9, http://www.journalofvision.org/content/7/2/12, doi:10.1167/7.2.12. [PubMed] [Article]

Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nature Neuroscience, 3,* 940–945.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience, 29,* 203–227.

Shmuel, A., & Leopold, D. A. (2008). Neuronal correlates of spontaneous fluctuations in fMRI signals in monkey visual cortex: Implications for functional connectivity at rest. *Human Brain Mapping, 29,* 751–761.

Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., & Engel, A. K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron, 60,* 709–719.

Smith, A. M., Lewis, B. K., Ruttimann, U. E., Ye, F. Q., Sinnwell, T. M., Yang, Y., et al. (1999). Investigation of low frequency drift in fMRI signal. *NeuroImage, 9,* 526–533.

Summerfield, C., Egner, T., Mangels, J., & Hirsch, J. (2006). Mistaking a house for a face: Neural correlates of misperception in healthy humans. *Cerebral Cortex, 16,* 500–508.

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32,* 1349–1357.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21,* 753–759.

Tootell, R. B., & Hadjikhani, N. (2001). Where is 'dorsal V4' in human visual cortex? Retinotopic, topographic and functional evidence. *Cerebral Cortex, 11,* 298–311.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A. M., & Kanwisher, N. G. (1998). Perceiving visually presented objects: Recognition, awareness, and modularity. *Current Opinion in Neurobiology, 8,* 218–226.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4,* 157–165.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Essen, D. C. V., et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature, 447,* 83–86.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron, 56,* 366–383.

Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 14,* 2057–2068.

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24,* 111–125.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., et al. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science, 316,* 1609–1612.

Zenger-Landolt, B., & Heeger, D. J. (2003). Response suppression in v1 agrees with psychophysics of surround masking. *Journal of Neuroscience, 23,* 6884–6893.