Published in final edited form as:
J Neurophysiol. 2006 September ; 96(3): 1556–1568. doi:10.1152/jn.01367.2005.

Uncertainty and Invariance in the Human Visual Cortex

Bosco S. Tjan[1], **Vaia Lestou**[2], and **Zoe Kourtzi**[3,4]

1*Department of Psychology and Neuroscience Graduate Program, University of Southern California, Los Angeles, California*

2*Department of Psychology, University of Glasgow, Glasgow, United Kingdom*

3*Max Planck Institute for Biological Cybernetics, Tuebingen, Germany*

4*School of Psychology, University of Birmingham, Edgbaston, Birmingham, United Kingdom*

Abstract

The way in which input noise perturbs the behavior of a system depends on the internal processing structure of the system. In visual psychophysics, there is a long tradition of using external noise methods (i.e., adding noise to visual stimuli) as tools for system identification. Here, we demonstrate that external noise affects processing of visual scenes at different cortical areas along the human ventral visual pathway, from retinotopic regions to higher occipitotemporal areas implicated in visual shape processing. We found that when the contrast of the stimulus was held constant, the further away from the retinal input a cortical area was the more its activity, as measured with functional magnetic resonance imaging (fMRI), depended on the signal-to-noise ratio (SNR) of the visual stimulus. A similar pattern of results was observed when trials with correct and incorrect responses were analyzed separately. We interpret these findings by extending signal detection theory to fMRI data analysis. This approach reveals the sequential ordering of decision stages in the cortex by exploiting the relation between fMRI response and stimulus SNR. In particular, our findings provide novel evidence that occipitotemporal areas in the ventral visual pathway form a cascade of decision stages with increasing degree of signal uncertainty and feature invariance.

INTRODUCTION

There has been a long tradition in engineering to use noise to perturb an unknown system to decipher the system's internal mechanisms. Such external noise methods for system identification have found a wide range of applications in visual psychophysics. They have been used to identify the causes of performance limitations (Burgess and Colborne 1988; Legge et al. 1987; Lu and Dosher 1999; Pelli 1990; Tanner and Birdsall 1958), infer properties of features and representations for high-level pattern-recognition tasks (Liu et al. 1995; Majaj et al. 2002; Pelli et al. 2003; Solomon and Pelli 1994; Tjan et al. 1995), determine how multiple cues are combined to generate a percept (Ernst and Banks 2002; Knill 1998; Landy et al. 1995), identify the sources of capacity limits in visual search (Eckstein and Whiting 1996), and distinguish the different enhancement mechanisms due to attention (Lu and Dosher 1998) and perceptual learning (Chung et al. 2005; Dosher and Lu 1999; Gold et al. 1999). Remarkably, all these studies provided detailed information about the visual system's functional architecture using only limited means to quantify the system's external behavior (e.g., accuracy and speed) and without any access to the system's internal states. Blood-

Address for reprint requests and other correspondence: B. Tjan, Dept. of Psychology, University of Southern California, Los Angeles, CA 90007 (E-mail: E-mail: btjan@usc.edu).

oxygenation-level-dependent (BOLD) functional MRI (fMRI) allows noninvasive access to a system's internal states. There exists an attractive possibility that external noise methods can be combined with fMRI to identify the neural mechanisms at different cortical regions in the human brain. Although the effects of stimulus noise (Boynton et al. 1999; Dakin et al. 2002; Rainer et al. 2001; Ress and Heeger 2003; Rees et al. 2000; Zenger-Landolt and Heeger 2003) and noise-like image scrambling (Grill-Spector et al. 1998) on the BOLD response have been previously studied, the absence of a general theoretical framework that relates stimulus signal-to-noise ratio (SNR) to BOLD activation has so far prevented any broad application of the external noise methods in fMRI studies.

The goal of our study was twofold: to measure the effect of stimulus noise on the analysis of visual scenes along the human visual pathway, from retinotopic regions to the lateral occipital complex (LOC), an area in the occipitotemporal cortex known to be involved in the analysis of visual shapes (Kanwisher et al. 1996; Malach et al. 1995) and object recognition (Bar et al. 2001; Grill-Spector et al. 2000), and to establish a theoretical framework for interpreting these measurements. To this end, we manipulated the SNR of visual stimuli by adding spatially correlated noise to noise-free images of natural scenes while keeping constant the root-mean-square (RMS) contrast[1] of the images (Fig. 1; see also METHODS). Subjects' task was to decide if the two pictures presented on either side of a fixation point were taken from the same scene. We measured the effect of stimulus SNR on BOLD responses across different cortical visual areas. We present empirical evidence and theoretical arguments in the context of the signal-detection theory (Green and Swets 1966) to show that the relationship between stimulus SNR and BOLD amplitude is correlated with the sequential ordering of cortical areas in a processing pathway. Our analysis and theoretical discussion mainly focus on the ventral pathway, as the stimuli (static images) and task (object/scene discrimination) used in our experiment are thought to engage primarily processing in the ventral occipitotemporal pathway (Ungerleider and Mishskin 1982). Our empirical findings suggest that the analysis of visual scenes in the occipitotemporal cortex is mediated by a cascade of decision stages with increasing degree of invariance and signal uncertainty. Our analytical method establishes a new approach for determining the sequential ordering of the processing stages along a decision cascade based on the relation between fMRI response and stimulus SNR.

METHODS

Subjects

Eight students from the University of Tuebingen participated in this study. All subjects had normal or corrected-to-normal vision. Six of the eight subjects had reliable retinotopy. Behavioral and fMRI data from these six subjects (3 males, 3 females, age: 22–35) were analyzed and reported here.

Stimuli

In each trial, two image panels of 152 × 216 pixels each were presented side by side (7.5° center to center) with the fixation point presented midway between the panels (Fig. 1, *A* and *B*). These panels were two large cutouts from the left and right halves of an indoor or outdoor photograph selected from the Corel Stock Photo Library. A subject's task was to judge if the two panels were taken from the same scene. The original photographs (364 × 236 pixels) were in color. The selected set was converted to black and white, and the gray levels of each picture were linearly adjusted such that all pictures had the same mean luminance and RMS contrast. Throughout the experiment (4 scans), the subjects were presented with each picture twice. The

[1]Root-mean-square (RMS) contrast of an image is the standard derivation of the pixel values expressed in units of Weber contrast. The Weber contrast of a pixel is defined as (pixel luminance – mean luminance)/(mean luminance).

pictures used in the experiment were randomly assigned into four groups, which were rendered at one of four noise levels. The pairing of the four noise levels to the four picture groups was partially counterbalanced within subjects and fully counterbalanced across subjects using a Latin Square.[2] Trials from each noise level were evenly divided across scans.

For any image noise to be effective in perturbing the later processing stages, the noise must survive the noise-reduction/signal-enhancement stages in early vision. Noise lacking spatial correlation can be reduced by averaging, and noise with different spatial scale than the stimuli can be excluded by filtering. To minimize the removal of the external noise by the visual system, we used a spatially correlated (or "pink") noise, the power spectrum of which matched exactly that of the stimulus. We expected (and verified) that such pink noise survived the generic noise-reduction steps in the visual system and could effectively affect higher-level processing. A fresh sample of pink noise was generated for each trial. To prevent the noise from conveying any task-relevant information, we generated the noise by replacing the amplitude spectrum of a white noise with the average amplitude spectra of the two pictures that would be used if the trial were a "different" trial. The resulting pink noise had the same mean luminance and RMS contrast as the pictures.

Signal and the corresponding pink noise were added in different proportions to form four SNR levels from 1/9 to 9 in equal log steps. To keep RMS contrast constant for all SNR levels, the signal and noise were added "on the circle" about the mean luminance. That is, if $K/(1-K)$ was the intended SNR and L_0 the mean luminance of the signal (and noise), the image used in the trial would be

$$\text{image}=L_0+\sqrt{K}(\text{signal}-L_0)+\sqrt{1-K}(\text{noise}-L_0) \quad (1)$$

For the LOC localizer scans, we used grayscale noise-free images (250 × 250 pixels or 10.9° × 10.9°) of novel and familiar objects as well as scrambled versions of each set as described previously (Kourtzi and Kanwisher 2000). Grid lines were present in both the intact and the scrambled images. For the localizer scans of the early retinotopic regions, we used rotating triangular wedge stimuli for the mapping of the borders between visual areas and concentric rings for eccentricity mapping. All the stimuli compensated for the expanded foveal representation and consisted of black-and-white checkerboards counter-flicking at a temporal frequency of 2 Hz as described in previous studies (DeYoe et al. 1996; Engel et al. 1994, 1997; Sereno et al. 1995).

All stimuli were displayed using a gamma-corrected LCD projector (NECGT950). Estimation of display gamma was obtained using a visual calibration procedure provided in PsychToolbox (Brainard 1997; Pelli 1997).

Design

Each subject was run in one session consisting of eight scans: two LOC localizer scans, two localizer scans for the early retinotopic areas, and four event-related scans for the main experiment. The order of the scans was counterbalanced across subjects.

The event-related scans for the main experiment consisted of one epoch of experimental trials and two 8-s fixation epochs, one at the beginning and one at the end of the scan. Each scan consisted of 25 experimental trials for each of the four SNR conditions and 25 fixation trials

[2]A Latin square is a square matrix consisting of *n* sets of the numbers 1 to *n* arranged in such a way that no row or column contains the same number twice. Using the rows (or columns) of a Latin square to order the *n* conditions in an experiment across blocks of trials ensures that each condition appears in each sequential position exactly once (or for equal number of times if repeated).

that were nonperiodically interleaved with the experimental trials. Each trial lasted for 3 s, and each stimulus was presented for 300 ms at the beginning of a trial, followed by a blank screen at mean luminance. As in previous studies (Kourtzi and Kanwisher 2000, 2001), the order of presentation was counterbalanced so that trials of each condition, including the fixation condition, were preceded (2 trials back) equally often by trials of each of the other conditions. Different stimuli were presented across conditions, but all the stimuli were presented in all conditions across subjects. Subjects were instructed to judge whether the two image panels presented in a trial were from the same or different scenes. In a "same" trial, both image panels were taken from the same picture (1 from each side and with their left-right order reversed so that they were never identical nor spatially aligned—Fig. 1*A*). In a "different" trial, panels were taken from different pictures. Behavioral data were recorded during the scans using a response button box.

For the LOC localizer scans, each stimulus condition (intact and scrambled images of objects) was presented for four 16-s stimulus epochs with interleaved fixation periods in a blocked design that balanced for the order of the conditions (Kourtzi and Kanwisher 2000, 2001). Twenty different images of the same type were presented for each condition. Each image was presented for 250 ms with a blank interval of 550 ms between images. The subjects were instructed to perform a 1-back matching task that engaged their attention in all the stimulus types, i.e., the intact and the scrambled images of objects. For the retinotopic mapping, we used eight positions for the wedges and eight eccentricity levels for the rings that were presented each for 8 s and repeated eight times. Subjects were instructed to fixate at the center of the display and to perform a dimming task on the fixation dot (detect change in fixation color).

Imaging

For all the experiments, data were collected on the 1.5 T Siemens scanner at the University Clinic in Tuebingen, Germany. A Gradient Echo pulse sequence (TR = 2 s, TE = 40 ms for the localizer scans; TR = 1 s, TE = 40 ms for the event-related scans) was used. Eleven axial slices (5 mm thick with 3 × 3 mm in-plane resolution) were collected with a head coil.

Data analysis

fMRI data were processed using BrainVoyager 4.6. Preprocessing of all the functional data included head movement correction, highpass temporal filtering, and removal of linear trends. The two-dimensional (2D) functional images were aligned to three-dimensional (3D) anatomical data, and the complete data set was transformed to Tailarach coordinates, inflated, unfolded, and flattened. For each individual subject, the LOC and the ventral and dorsal retinotopic areas (V1, V2, V3v/VP, V4, V3d, V3a) were defined as regions of interest (ROIs). We restricted these ROIs to voxels significantly activated by our stimuli, based on a 3D statistical map computed by correlating, for each voxel, the signal time course with a reference function derived from the hemodynamic response properties (Boynton et al. 1996; Cohen 1997; Dale and Buckner 1997). For each individual subject, the LOC was defined as the set of continuous voxels in the ventral occipitotemporal cortex that showed significantly stronger activation ($P < 10^{-4}$, corrected) to intact than scrambled images based on the average data from the two localizer scans. Two subregions of the LOC were identified: the posterior lateral occipital region (LO) and the anterior regions in the posterior fusiform sulcus (pFs) as described in previous studies (Grill-Spector et al. 2000). The early visual areas (V1, V2, V3v/VP, V4, V3d, V3a) were identified based on standard retinotopic mapping procedures (DeYoe et al. 1996; Engel et al. 1994, 1997; Sereno et al. 1995).

For each event-related scan, the fMRI response (percent BOLD signal change relative to fixation baseline) was extracted by averaging the data from all the voxels within each of the independently defined ROIs. In each scan, we averaged the signal intensity across all the trials

in each condition at each of the 15 corresponding time points (seconds) and converted these time courses to percent signal change relative to the fixation trials, as described previously (Kourtzi and Kanwisher 2000, 2001). We then averaged the time courses for each condition across scans for each subject. Because of the hemodynamic lag in the fMRI response, the peak in overall response and therefore the differences across conditions are expected to occur at a lag of several seconds after stimulus onset (Boynton et al. 1996; Cohen 1997; Dale and Buckner 1997). We quantified the peak response in each ROI by fitting a parametric model of the hemodynamic response function defined by the difference of Gamma functions (Boynton and Finney 2003; Boynton et al. 1996) to the average fMRI responses across observers. This analysis showed that the peak BOLD response across conditions and ROIs was within 4–6 s after stimulus onset. Thus the average response between 4 and 6 s was defined as the *peak response* of the fMRI time courses. The *BOLD response function* was defined as the peak response versus SNR. The quantity of interest is the log-log slope of the BOLD response function. The difference in the log-log slope of the BOLD response function across the ventral visual pathways was tested using ANOVA with ROI (V1, V2, V3v/VP, V4, LO, pFs) as the within-subject factor and log-log slope as the dependent variable. (Separately, we also tested along the dorsal pathways: V1, V2, V3d, V3a.) To statistically evaluate if the log-log slope increased monotonically along a visual stream, we adopted an ordering of the ROIs that corresponded to the previously proposed processing hierarchy (e.g., Lerner et al. 2001): ventral: V1→V2→V3v/VP→V4→LO→pFs, and dorsal: V1→V2→V3d→V3a. With respect to a putative ordering, polynomial contrasts (i.e., polynomial regressions between the log-log slope and the putative ordering of an ROI along a processing stream) were computed for significance as part of the ANOVA. A significant linear term and the absence of any significant higher-order terms are indicative of a monotonic trend (Keppel and Wickens 2004). Between-subjects concordance regarding the rank order of the log-log slopes across ROIs was assessed using Kendall's *W* (Kendall coefficient of concordance). To assess the effect of behavioral accuracy on the log-log slopes, we also extracted the time courses, the BOLD response functions, and their log-log slopes separately from correct and incorrect trials. A two-way ANOVA with correctness (correct vs. incorrect trials) and ROI as within-subject factors and the log-log slope as dependent variable was performed separately for the ventral and dorsal ROIs.

RESULTS

Over the range of tested SNRs, subjects' behavioral accuracy increased continuously from 52% (near chance) to 79% correct as stimulus SNR increased (Fig. 2), indicating that the noise manipulation was effective in affecting performance [$F(3,15) = 26.3$, $P < 0.00001$]. In contrast, subjects' reaction times (for correct trials) were unaffected by SNR [mean RT = 0.9*s*, $F(3,15) = 1.25$, $P = 0.33$].

For all cortical areas where the peak BOLD response varied with SNR, the variation was monotonic (Fig. 3B). The BOLD response functions of each cortical area, based on either the time courses from individual subjects or the averaged time course across subjects, showed a progressive increase in their log-log slopes along the ventral processing stream (Figs. 4, *A* and *B*). A repeated-measures ANOVA applied to the ROIs along the ventral pathway (V1, V2, V3v/VP, V4, LO, pFs) revealed a highly significant effect of ROI on the log-log slope of their respective BOLD response functions [$F(5,25) = 8.65$, $P < 0.001$].

More importantly, a planned polynomial contrast revealed a monotonic relationship between the log-log slope of a cortical area and the putative sequential ordering of the area along the ventral pathway. Specifically, we found a significantly positive linear term [$F(1,5) = 15.3$, $P < 0.05$] and a significantly positive quadratic term [$F(1,5) = 13.0$, $P < 0.05$]. No other higher-order terms were significant. As suggested by Fig. 4*B*, the quadratic term was due to the noisy and near-zero log-log slope of V1 and V2, rather than indicating any strong nonmonotonicity.

Indeed, when either V1 or V2 was excluded from the analysis, the quadratic term vanished, whereas the linear term remained. Between-subject agreements on the rank order of the log-log slopes across visual areas were highly significant [Kendall's $W = 0.819$, $\chi^2(5) = 24.6$, $P < 0.001$] (Fig. 5). In sum, our data showed that the log-log slope of the BOLD response functions increased progressively from −0.015 (not significantly different from zero) in V1 to 0.22 in pFs in the ventral stream.

To determine if behavioral accuracy played a role in the observed progression of the log-log slope of the BOLD response functions along a visual pathway, we extracted the BOLD response separately for the correct and incorrect trials. Figure 4, *C* and *D*, shows the BOLD response functions obtained and their log-log slopes, respectively. A two-way (ROI × correctness) repeated-measures ANOVA applied to the ventral ROIs found no significant effect of Correctness (correct vs. incorrect trials) on the log-log slope of the BOLD response functions [$F(1,5) = 1.08$, $P = 0.31$] even though the log-log slope values from incorrect trials appeared lower on average (Fig. 4*D*). There was also no significant interaction between correctness and ROI [$F(5,25) = 1.26$, $P = 0.31$]. The effect of ROI remained highly significant [$F(5,25) = 7.42$, $P < 0.001$], and so was the monotonicity progression of the log-log slopes for both correct and incorrect trials, in spite of a reduction in reliability for estimating the log-log slopes using the smaller correctness-contingent data set. When the analysis was limited to the two subregions in LOC (LO and pFs), where the BOLD response has previously been shown to correlate with performance (Grill-Spector et al. 2000), the effect of correctness on the log-log slope of the BOLD response functions exhibited only a trend toward significance [$F(1,5) = 5.1$, $P = 0.07$], and there was no interaction between correctness and ROI [$F(1,5) = 0.03$, $P = 0.86$]. In short, the monotonic increase of the log-log slope of the BOLD response functions along the ventral pathway remained robust regardless of accuracy in performance.

Interesting, we also observed the same pattern of results along the dorsal pathway (V1, V2, V3d, V3a) even though our static stimuli and scene-discrimination task were not expected to optimally engage the dorsal pathway. Briefly, we found a highly significant effect of ROI on the log-log slope of the BOLD response function [$F(5,25) = 7.41$, $P < 0.01$], a monotonic progression of the log-log slope along the dorsal pathway [significant linear $F(1,5) = 10.0$, $P < 0.05$ and quadratic terms $F(1,5) = 8.67$, $P < 0.05$, and the quadratic term vanished when V1 was excluded from the analysis], no effect of correctness on the log-log slopes [$F(1,5) = 0.032$, $P = 0.87$] and no interaction between ROI and correctness [$F(3,15) = 0.70$, $P = 0.57$].

At the lowest stimulus SNR, objects and scenes were almost invisible, and the images were devoid of any sharp edges; however, the mean luminance, RMS contrast, and the distribution of contrast energy across spatial frequencies remained the same as the noiseless originals. It appears that the BOLD responses in V1/V2 were not contingent on the presence of any form or edges in the image because the BOLD responses in V1/V2 did not appear to depend on SNR (V2 showed a nonsignificant negative slope). BOLD responses remained independent of SNR in V1 and V2 when the analysis was limited to voxels that are closer to the foveal confluence (average log-log slope was 0.01 for V1 and −0.01 for V2, not significantly different from 0).

Might general attention/arousal differences across conditions modulate the fMRI signals and result in the differences observed across visual areas? We think that these confounds could not account for our findings for the following reasons. First, it is highly unlikely that observers could selectively choose to attend to particular conditions as trials were presented in quick succession and were randomly interleaved. Second, if the results in the higher visual areas were simply due to task difficulty, we would expect stronger BOLD responses in the hardest condition. In contrast, we observed the lowest BOLD responses for the hardest condition. Third, the reaction times in the hardest condition were statistically indistinguishable from those in the easiest condition; they would have been faster if the observers had given up and were

simply guessing. This suggests that observers were engaged in the task across all conditions and not responding randomly. Fourth, there was no effect of response accuracy (correct vs. incorrect trials) on the log-log slopes, suggesting that our findings could not be accounted for by differential attention across conditions depending on the behavioral performance of the observers.

To control for differences across conditions in eye movements, we measured eye movements in three subjects in the scanner using an infrared eye tracker with remote optics (ASL 504 LRO). The eye-movement recordings showed no differences across conditions in the mean eye position [x position: $F(3,6) = 0.53$, $P = 0.68$; y position: $F(3,6) = 0.94$, $P = 0.48$], the number of saccades [$F(3,6) = 0.70$, $P = 0.58$], the mean saccade amplitude [$F(3,6) = 1.80$, $P = 0.91$], the fixation duration [$F(3,6) = 0.71$, $P = 0.58$], and the number of blinks [$F(3,6) = 0.40$, $P = 0.76$]. These controls suggest that it is not likely that our results could be significantly confounded by differences in eye movements across conditions.

DISCUSSION

Our study reports three main findings. First, the observers' accuracy in the scene discrimination task increased with increasing stimulus SNR. Second, when stimulus contrast was held constant, the effect of stimulus SNR on fMRI responses varied across cortical areas. Third, and most importantly, the dependency of BOLD signal on stimulus SNR varied systematically along the ventral and dorsal visual streams and appeared to be contingent on the cortical area's "distance" from the retinal inputs. Specifically, the further away an area is from retinal inputs, the more dependent the BOLD signal was on input SNR. The log-log slope of BOLD signal-change versus image SNR increased progressively: V1/V2 < V3v/VP < V4 < LO < pFs along the ventral stream, and V1/V2 < V3d < V3a along the dorsal stream.[3] How could such regularity be explained?

We propose that the progressive increase in the log-log steepness of the BOLD response function is due to a progressive increase in invariance along the forward direction of a visual pathway and that increase in invariance is a general property of information processing. Our theory leads to the intriguing possibility that an external-noise method can be used with fMRI to index the sequential ordering of cortical regions within an information-processing pathway. We present this theory next and discuss additional empirical support for it as well as potential challenges. For clarity, we focus this theoretical discussion on the ventral pathway, as the stimuli (static images) and task (object/scene discrimination) used in our experiment are thought to engage primarily processing in the ventral occipitotemporal pathway (Ungerleider and Mihskin 1995). Given the stimulus complexity (natural images), the fact that we found similar results in retinotopic areas of the dorsal visual pathway is confirmatory and of interest for future study.

Sequential ordering of visual processing and the log-log steepness of BOLD versus SNR

Intuitively speaking, the log-log steepness of the BOLD response function is related to the accumulation of nonlinear decision stages, ordered sequentially, in visual processing. We formalize this intuition in the context of signal detection theory in four steps. First, we will present the view that visual processing in the cortex can be seen as a cascade of classifiers. Second, we will argue that an increase in invariance is a necessary consequence of visual processing. Third, we will link invariance with the concept of *intrinsic uncertainty* from the signal detection theory and describe a relationship between intrinsic uncertainty and the slope

[3]The < signs represent the monotonic ordering as revealed by the polynomial contrast analysis. They are not meant to suggest significant pairwise difference, which is a more stringent test of monotonicity.

of a psychometric function. Fourth, we will relate the "psychometric" function for a population of neurons to the BOLD response function.

VISUAL PROCESSING AS A CASCADE OF CLASSIFIERS—Visual processing entails the extraction of "features" from retinal inputs that are relevant to behavior. This feature extraction process is often thought to be hierarchical, resulting in features that are progressively more complex, invariant, and task-relevant. The neural correlates for this stage-wise escalation of feature complexity and invariance seem apparent. In the human ventral cortex, lower (V1, V2) and higher (V4, LOC) visual areas have been implicated in the analysis of simple and complex features, respectively. Despite its complex dependence on feedback and recurrent connections, information processing along the ventral visual stream appears to follow a discernable order: V1, V2, V3v/VP, V4, and LOC (e.g., Lennie 1998) as suggested by the anatomical connectivity (Felleman and Van Essen 1991; Van Essen et al. 1992), response latency (Raiguel et al. 1989; Schmolesky et al. 1998), neural selectivity for image features (Hubel and Wiesel 1977; Livingstone and Hubel 1988), and response invariance (Fujita et al. 1992; Grill-Spector et al. 1999; Ito et al. 1995) across visual areas. This succession of neural stages is consistent with the classical theories of object perception (e.g., Marr 1982) that propose a cascade of computational stages. Each stage can be described as a classifier that maps configurations of simpler input features into output categories that correspond to more complex features.

INVARIANCE AND VISUAL PROCESSING—Along the visual processing pathway, different inputs to a classifier are either mapped to different outputs if the differences are important for the subsequent stages of analysis or to the same output if the differences are irrelevant. An edge detector, for example, may distinguish between edges of different orientations but remain indifferent to the cause of an edge (e.g., whether the edge is an abrupt change in luminance or in texture). The operations in each classifier can be functionally divided into two kinds: AND operations, which integrate simpler input features to form a more complex target feature, and OR operations, which allow for different versions of a target feature. The AND operator increases feature specificity and complexity, whereas the OR operator increases feature invariance. This AND/OR structure is evident in the classic models for the analysis of early visual features (Hubel and Wiesel 1977) and complex objects (Riesenhuber and Poggio 1999).

If we assume that a later stage in visual processing does not undo what has been done by the earlier stages, then the accumulation of AND and OR operations along a visual processing pathway means that both specificity and invariance must increase along the pathway. Increase in invariance, regardless of the specific type of invariance, has a distinct psychophysical signature, on which we shall elaborate next.

UNCERTAINTY, INVARIANCE AND THE SLOPE OF A PSYCHOMETRIC FUNCTION—A psychometric function relates performance (e.g., percent correct, d', hit rate, etc.) to task-relevant signal intensity (e.g., signal-to-noise ratio, contrast, etc). Here, we will derive a mathematical link between invariance and the slope of a psychometric function for a feature detector. Early studies in audition (cf. Green 1964) and vision (e.g., Foley and Legge 1981; Nachmias and Sansbury 1974; Stromeyer and Klein 1974; Tanner and Swets 1954) found that when a subject was asked to detect a faint but precisely defined signal, the resulting psychometric function had a slope much steeper than what would be expected of an ideal observer who knows the signal exactly and who is only limited by internal and external noise. Tanner (1961) pointed out that if an observer did not know the signal exactly and, therefore had to consider a number of possibilities, the observer, which was otherwise optimal, would have a steeper psychometric function similar to those obtained with human subjects. This led

to the important notion of "uncertainty"—the number (M) of orthogonal signals that an observer is effectively considering.

Uncertainty relates closely to invariance. Suppose the target in an experiment is one specific stimulus, but the detector, for whatever reasons, uses a mechanism that responds equally to any of the M possible stimuli. Such a detector is said to have *intrinsic* uncertainty (uncertainty that is not in the stimulus but in the observer) (cf. Graham 1989). Feature invariance is a form of intrinsic uncertainty. Consider a detector for an object that is invariant to orientation. Such a detector will respond equally to a target regardless of which way it is orientated. It is as if the detector is obligatorily considering all admissible versions of the target. Such a detector will exhibit an amount of intrinsic uncertainty equal to the effective number of orthogonal input patterns that activate the detector. The notations of "invariance" and "uncertainty," albeit different in their historical and theoretical origins, are essentially interchangeable.

Consider a psychometric function that relates hit rate to stimulus SNR for a "yes-no" feature-detection task. Hit rate is the probability of responding yes when the target feature was present. There has been at least a 30-yr history of describing a psychometric function with the best-fitting Weibull function

$$p(c)=1-(1-\gamma)\exp(-(c/\alpha)^{\beta}) \tag{2}$$

where $p(c)$ is the hit rate, c is the input SNR, γ the false-alarm rate, α is the threshold (the SNR required for the hit-rate to reach 63%, or $1-e^{-1}$, of the way between the false-alarm rate and the asymptotic hit rate), and β is proportional to the log-log slope of the Weibull function at $c=\alpha$.

Based on the findings from a seminal paper by Pelli (1985) on uncertainty, we can show that β, and hence the log-log slope of the psychometric function of hit rate, is monotonically related to uncertainty M if we assume that the detector retains a constant level of false alarm rate, γ_0, independent of uncertainty. Specifically, we show (APPENDIX A) that

$$\beta=k\Phi^{-1}((1-\gamma_0)^{1/M}) \tag{3}$$

where k is a proportional constant, and Φ^{-1} is the inverse of the cumulative normal probability distribution.

If we assume that the false alarm rate (γ_0) of a feature detector (e.g., a neuron) is low, then the false alarm will effectively be a constant because false alarm is lower-bounded by zero. Thus for a detector with a low false alarm rate, its intrinsic uncertainty M (or equivalently, invariance) and the log-log steepness β of its psychometric function is monotonically related according to Eq. 3, with higher intrinsic uncertainty correlates with higher log-log steepness. Measuring the log-log steepness of the detector's psychometric function will therefore provide us with a way to ascertain the detector's intrinsic uncertainty.

To link the log-log steepness β to a detector's intrinsic uncertainty, we required the false alarm rate γ_0 to be relatively constant across detectors. However, we do not need to assume the same for the threshold α of the detectors, which depends in part on the specificity and complexity of the features it tries to detect. α specifies the horizontal position of a psychometric function along the abscissa of log SNR and is separable from the log-log steepness β.

FROM PSYCHOMETRIC FUNCTION TO BOLD ACTIVATION—We adapt the view that neural activity (e.g., average firing rate) is positively correlated with the presence of a

target in a neuron's receptive field; in other words, firing rate in a neurometric function is a surrogate for hit rate in the corresponding psychometric function.

BOLD signal correlates with neural activities (cf. Heeger and Ress 2002; Heeger et al. 2000; Logothetis 2002; Logothetis et al. 2001). Whether BOLD signal corresponds to the input or the output of a cortical region (Logothetis et al. 2001) will not fundamentally affect our analysis because the inputs to an area are the outputs from the preceding areas, thus the ordinal relationship remains unchanged. We assume that within a region of interest (ROI), BOLD signal is related to average "neural activity" (e.g., local field potential, single- and multiunit activities, etc.) by a power function. This assumption is consistent with data from Logothetis et al. (2001). We will further assume that, within the ROI, the average "neural activity" is related to the *average hit rate*, $p_A(c)$, by another power function. These two power functions can be combined into a single power function, and we have

$$\%\,\Delta\,\mathrm{BOLD}(c)=g\cdot(p_{\mathrm{A}}(c))^{h} \tag{4}$$

where $\%\Delta BOLD(c)$, the BOLD response function, is the percent BOLD signal change relative to fixation baseline as a function of stimulus SNR c, and g, h are parameters of the combined power function. Combining the analytical relationship between uncertainty and the log-log slope of a single detector's response function (Eq. 3) with assumed power-function relationship between BOLD signal and the average response function of detectors in an ROI (Eq. 4), we derived (APPENDIX B) a specific relationship between the log-log slope of the BOLD response function and uncertainty

$$\left.\frac{\mathrm{d}(\log(\%\Delta\mathrm{BOLD}(c)))}{\mathrm{d}(\log(c))}\right|_{c=\alpha}=kh\Phi^{-1}((1-\gamma)^{1/M})(1-\gamma)\exp(-1)/(1-(1-\gamma)\exp(-1)+\gamma(1-\xi)/\xi) \tag{5}$$

where M is the amount of intrinsic uncertainty (or feature invariance), k is the proportional constant from Pelli (1985) relating β to the decision criteria λ, h is the exponent relating % ΔBOLD to hit rate, γ is the false alarm of a single neuron, and ξ is the fraction of neurons in the ROI that are selective (and responding) to the presented target. k is an analytical constant, unrelated to biology. We assume that h is roughly constant over the cortex (or at least within the visual areas). We also assume that the false alarm rate (γ) of all neurons is small and similar (i.e., it does not make sense for a neuron to signal the presence of a target if such signaling was often erroneous). With these two assumptions, the log-log steepness depends only on the intrinsic uncertainty (M) and the fraction (ξ) of neurons responding to the noise-free versions of the stimuli. The effect of ξ on the log-log steepness is reduced if the false alarm rate is sufficiently small or when ξ is sufficiently large. The later can be assumed for the ROIs where activity is highly modulated by the stimuli compared with fixation baseline as was the case in our experiment. Under these considerations, we can state that the log-log steepness of BOLD response function reflects mainly the intrinsic uncertainty (M), or in other words, the feature invariance, of a cortical area.

SUMMARY—The analysis can be briefly summarized as follows. Along the forward direction of a visual processing pathway, feature invariance increases. In the context of signal detection theory, feature invariance corresponds to the system's intrinsic uncertainty. Our derivation shows that across visual areas, an increase in uncertainty leads to an increase in the log-log steepness (β) of the "psychometric" function of a neural mechanism, which relates the probability of target detection to the stimulus SNR. With a small set of assumptions, we showed that an increase in β correlates with an increase in the log-log steepness of the BOLD response function (%ΔBOLD vs. SNR). We would thus predict that along the forward direction of a

visual processing pathway, the log-log slope of the BOLD response function must increase, which is consistent with our empirical findings.

Hierarchical decision cascade with feedback

We linked the log-log steepness of the BOLD response function to the degree of invariance (intrinsic uncertainty) of the features extracted by a cortical region, and the degree of invariance to the sequential ordering of cortical region along the "forward" direction of visual processing. However, reciprocal connections between cortical areas are prevalent (cf. Felleman and Van Essen 1991; Van Essen et al. 1992) and functionally relevant for visual processing. Because the information flow is clearly bidirectional, how can one define what is the forward direction?

A forward direction of information processing is defined in terms of how information is being processed and does not preclude the use of feedback to implement a decision hierarchy. For example, a forward direction is well defined if the dependency between feature types is directional—the detection of a contour depends on the detection of some local edge fragments consistent with the contour, but the detection of an edge fragment need not be consistent with the presence of a contour. Feedback from a contour-detection stage may modulate the sensitivity of an edge detector but does not alter its basic function of being an edge detector. A corollary of this type of dependency is that more than one set of low-level features can signify the presence of a higher-level feature (e.g., the same contour can be detected based on different sets of edge fragments). In other words, there is a progressive increase in feature invariance, in addition to feature complexity, along the forward direction. Our analysis showed that the log-log steepness of the BOLD response function is monotonically related to feature invariance and can thus be used to indicate the forward direction of information processing. The advantage of using the log-log steepness is that we do not need to define any ad hoc metrics for feature complexity or invariance.

Effects of noise, feature complexity, and contrast

Our main empirical finding is that image-level noise is more effective in disrupting BOLD activation in higher than lower visual areas. We attributed this to a progressive increase in invariance (i.e., intrinsic uncertainty) along the ventral pathway. Empirical evidence for an increase in invariance across image changes along the ventral pathway has already been established (Fujita et al. 1992; Grill-Spector et al. 1999; Ito et al. 1995). Our contribution is the theory that links the degree of invariance to the sensitivity of BOLD activation to image-level noise. This link is general and does not depend on the specific type of invariance, such as invariance in size or position.

One could argue that the observed effect between image-level noise and BOLD activation is due to the specific noise we used or related to the specific features the noise happened to disrupt. It is unlikely that our empirical findings are noise specific. A number of other manipulations that affect the front-end SNR, such as image scrambling, showed similar results—i.e., a strong modulation effect on BOLD at the later stages along the ventral pathway (Grill-Spector et al. 1998; Lerner et al. 2001; Rainer et al. 2002). When visual information processing is conceptualized as a sequence of AND and OR operations as here, both feature specificity (or complexity) and invariance are expected to increase along a processing pathway. In a fundamental sense, the alterative explanation, that a highly specialized feature formed at a higher-level stage is easier to be disrupted by random noise—cannot be completely ruled out. However, the effect of feature complexity per se on a psychometric or neurometric function is less specific, making it more difficult to predict a priori its effects on BOLD without additional assumptions on the type of feature complexity.

Avidan et al. (2002) showed that the sensitivity to image contrast decreased along the ventral pathway. This finding appears to contradict our theory. There were two critical differences between Avidan et al. and the present study. First, Avidan et al. manipulated contrast in the absence of noise, essentially relying on an observer's internal noise to limit performance. As a result, performance could remain high at very low contrast (lowest nonzero contrast tested in Avidan et al. was 4%). Second, and more importantly, the observers' accuracy in Avidan et al. remained relatively constant in all the tested contrast levels except the zero-contrast condition. For line-drawings, performance often remains >80% correct for a contrast as low as 2% if no external noise is present (e.g., Tjan et al. 1995). A sharp transition of the BOLD response in the LOC would therefore be expected if one titrates a narrow range of contrasts between 0.5–3%. The result would be a steeper log-log slope of BOLD response versus contrast for LOC and a comparatively much shallower slope for V1—a pattern consistent with our current results. This abrupt change in the contrast response function followed by saturation could also be inferred from data in Olman et al. (2004) for extrastriate areas but not for V1. In general, the log-log slope of interest is the one obtained at threshold SNR (Eq. 5), where the slope of the BOLD response function is the steepest.

We note that the log-log slopes of the BOLD response function measured in the same cortical region (e.g., V1) can be different across different studies. Equation 5 shows that the log-log slope is affected by the fraction (ξ) of neurons responding to the noise-free versions of the stimuli and the false-alarm rate (γ). ξ depends on the stimuli, whereas γ depends on a combination of the stimulus-level noise and noise internal to the observer. To use the log-log slope as an index for feature invariance or intrinsic uncertainty (M), we have to assume that within an experiment, γ is relatively constant across cortical areas, and *if* γ is not adequately small, ξ must be sufficiently close to unity. Because the values of ξ and γ vary across experiments, we do not expect the log-log slopes to stay constant across studies. However, with respect to our theory, we do expect the ordinal relationship of the log-log slopes between different cortical areas to be consistent across studies.

Finally, Rainer et al. (2001) used a noise manipulation very similar to ours and showed that for anesthetized monkeys, BOLD signal varied nonmonotonically with SNR in V1 and extrastriate areas. This nonmonotonicity was not observed in our data. In Rainer et al., SNR was manipulated using a simple weighed sum

$$image = K\ signal + (1 - K)\ noise \tag{6}$$

This manipulation caused RMS contrast to co-vary with SNR, such that at 0 and 100% noise, the RMS contrast is at maximum, whereas at 50% mixture of signal and noise, the RMS contrast is at its minimum (Dakin et al. 2002). Since BOLD signal from V1 is known to be sensitive to contrast (Boynton et al. 1999; Olman et al. 2004; Ress and Heeger 2003; Zenger-Landolt and Heeger 2003), it is most likely that the nonmonotonicity observed in Rainer et al. is due to a variation in contrast. In the current study, we manipulated SNR in a manner that ensured constant RMS contrast across noise levels.

Conclusions

When random visual noise was used to disrupt the retinal input from visual images, its effect on the BOLD activation of different cortical areas varied systematically along the ventral and dorsal visual pathways. In particular, the log-log slope of the BOLD response function (% BOLD signal change relative to fixation baseline versus stimulus SNR) increased monotonically from early to late visual areas along a visual pathway. An increase in SNR from 1/9 to 9 resulted in little change to the BOLD response in V1/V2 but caused progressively larger changes in cortical areas that are further away from the retinal inputs. We adapted the

analysis of uncertainty from signal detection theory to interpret this systematic change in the BOLD response function along a visual pathway. We showed that the observed increase in the log-log slope of the BOLD response function areas along a visual pathway is a general property of visual information processing—a progressive increase of invariance (or equivalently, intrinsic uncertainty) along a cascade of classifiers.

External noise methods are powerful tools in vision psychophysics. They provide a quantitative interpretation of behavioral data at a level of detail that permits computational modeling of the underlying mechanisms. Here we provide novel evidence that external noise methods can be used to study the representations of visual scenes along the ventral visual pathway. Investigating these representations in the light of findings from signal detection theory, such as the relationship between uncertainty and the log-log slope of a psychometric function, provides novel insights in understanding the degree of invariance in the features extracted by different cortical areas without making specific assumptions about what these features are. Contingent on the validity of a few general assumptions about signal detection theory and the relationship between neural activity and BOLD signal, this method can also "image" the ordinal flow of information processing between different cortical areas based on the effects of stimulus noise on their processing.

Acknowledgments

GRANTS

This study was supported by National Eye Institute Grant EY-016391 and University of Southern California Zumberge Fund to B. S. Tjan, funding from the Max-Planck Society and the Biotechnology and Biological Sciences Research Council (BB/D52199X/1) to Z. Kourtzi, and the Deutsche Forschungsgemeinschaft Grant TH812/1-1 to the University Clinic Tuebingen.

APPENDIX A

In this appendix, we derive the relationship between the log-log steepness (β) of the psychometric function of hit rate and intrinsic uncertainty (*M*), under the assumption of a constant force-alarm rate (γ_0). We express the psychometric function of hit rate using a Weibull function

$$p(c)=1-(1-\gamma)\exp(-(c/\alpha)^{\beta}) \tag{A1}$$

where $p(c)$ is a hit rate, c is the input SNR, γ the false-alarm rate, α is the threshold SNR, and β is proportional to the log-log slope of the Weibull function at $c = \alpha$

$$\frac{\mathrm{d}(\log(p))}{\mathrm{d}(\log(c))}\Big|_{c=\alpha}=\beta\cdot(1-\gamma)\exp(-1)/(1-(1-\gamma)\exp(-1)) \tag{A2}$$

Pelli (1985) considered a maximum-output model of uncertainty. The model consists of multiple input channels that are orthogonal, each responds by correlating the input with a template. The maximum response among all the channels is taken to be the *decision variable* of the model. The number (*M*) of the orthogonal channels is a measure of the intrinsic uncertainty.

Pelli (1985) showed that the log-log steepness (β) of this model's psychometric function for hit rate is proportional to the model's response criterion (λ), which is the threshold that the model's decision variable must exceed before the model responds yes. Thus

$$\beta = k\lambda \tag{A3}$$

where k is a proportional constant. A high criterion will reduce false alarms, but it also reduces hit rate. Pelli also showed that the false-alarm rate (γ) is related to the level of uncertainty (M) and the response criterion (λ)

$$\gamma = 1 - \Phi^{M}(\lambda) \tag{A4}$$

where Φ is the cumulative normal probability distribution. Although Pelli's derivation was based on a specific observer model (the maxoutput model), the qualitative relation between these parameters holds for almost all reasonable models of uncertainty proposed before and after Pelli (cf. Graham 1989; Green and Swets 1966; Swets 1964), including models that are optimal in a Bayesian sense (Tjan 1996).

If an observer (or a neuron) wants to maintain a fixed level of alarm false rate, say γ_0, independent of uncertainty or invariance, then according to Eq. A4, the response criterion (λ) must increase with uncertainty. Because β is proportional to the response criterion (Eq. A3), it also increases with uncertainty

$$\beta = k\lambda = k\Phi^{-1}((1 - \gamma_0)^{1/M}) \tag{A5}$$

where Φ^{-1} is the inverse of the cumulative normal probability distribution.

APPENDIX B

In this appendix, we derive the relationship between the log-log slope of the BOLD response function (%ΔBOLD vs. stimulus SNR) and intrinsic uncertainty (M), building on the results from APPENDIX A and the power-function assumption that links BOLD signal to the average psychometric function of hit-rate of the feature detectors in an ROI (Eq. 4 in main text)

$$\%\,\Delta\,\mathrm{BOLD}(c) = g \cdot (p_A(c))^h \tag{B1}$$

Because only a fraction of neurons will be activated by a stimulus, the average hit rate within the ROI is

$$p_A(c) = \xi p(c) + (1 - \xi) p(0) \tag{B2}$$

where $0 \leq \xi \leq 1$ is the relative density of the neurons in the ROI whose target features are in the (noiseless) stimulus, $p(c)$ is the hit rate for such an individual neuron in the ROI, and $p(0)$ corresponds to the false alarm rate. The average hit rate can be modeled with a Weibull function modified such that its asymptote is <1.0. From Eqs. A1 and B2, we can see that the average false-alarm rate is the same as that of a single neuron ($\gamma_A = \gamma$), and likewise for the average threshold SNR ($\alpha_A = \alpha$). If the false-alarm rate is sufficiently low or if ξ is sufficiently high, then the average log-log steepness (β_A) will be roughly the same as that for a single neuron ($\beta_A = \beta$); otherwise, β_A and β will be linearly related by a factor depending on ξ and γ

$$\frac{d(\log(p_A))}{d(\log(c))}\Big|_{c=\alpha}=\beta\cdot(1-\gamma)\exp(-1)/(1-(1-\gamma)\exp(-1)+\gamma(1-\xi)/\xi) \tag{B3}$$

By combining Eqs. A5, B1, and B3, we can relate the log-log steepness of BOLD versus SNR to the intrinsic uncertainty (M)

$$\frac{d(\log(\%\Delta\mathrm{BOLD}))}{d(\log(c))}\Big|_{c=\alpha}=kh\Phi^{-1}((1-\gamma)^{1/M})(1-\gamma)\exp(-1)/(1-(1-\gamma)\exp(-1)+\gamma(1-\xi)/\xi) \tag{B4}$$

where k is the proportional constant from Pelli (1985) in Eq. A5 relating β to the decision criteria λ, h is the exponent relating %ΔBOLD to every hit rate, γ is the single-neuron false alarm rate, and M is the amount of intrinsic uncertainty (or feature invariance).

REFERENCES

Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R. Contrast sensitivity in human visual areas and its relationship to object recognition. J Neurophysiol 2002;87:3102–3116. [PubMed: 12037211]

Bar M, Tootell RBH, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM. Cortical mechanisms specific to explicit visual object recognition. Neuron 2001;29:529–535. [PubMed: 11239441]

Boynton GM, Demb JB, Glover GH, Heeger DJ. Neuronal basis of contrast discrimination. Vision Res 1999;39:257–269. [PubMed: 10326134]

Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 1996;16:4207–4221. [PubMed: 8753882]

Boynton GM, Finney E. Orientation-specific adaptation in human visual cortex. J Neurosci 2003;23:8781–8787. [PubMed: 14507978]

Brainard DH. The psychophysics toolbox. Spatial Vision 1997;10:433–436. [PubMed: 9176952]

Burgess AE, Colborne B. Visual signal detection. IV. Observer inconsistency. J Opt Soc Am A 1988;5:617–627. [PubMed: 3404312]

Chung STL, Levi DM, Tjan BS. Learning letter identification in peripheral vision. Vision Res 2005;45:1399–1412. [PubMed: 15743610]

Cohen MS. Parametric analysis of fMRI data using linear systems methods. Neuroimage 1997;6:93–103. [PubMed: 9299383]

Dakin SC, Hess RF, Ledgeway T, Achtman RL. What causes nonmonotonic tuning of fMRI response to noisy images? Curr Biol 2002;12:R476–R477. [PubMed: 12176342]author reply R478

Dale AM, Buckner RL. Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Mapp 1997;5:329–340.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J. Mapping striate and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci USA 1996;93:2382–2386. [PubMed: 8637882]

Dosher B, Lu ZL. Mechanisms of perceptual learning. Vision Res 1999;39:3197–3221. [PubMed: 10615491]

Eckstein MP, Whiting JS. Visual signal detection in structured backgrounds. I. Effect of number of possible spatial locations and signal contrast. J Opt Soc Am A 1996;13:1777–1787.

Engel SA, Glover GH, Wandell BA. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex 1997;7:181–192. [PubMed: 9087826]

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN. fMRI of human visual cortex. Nature 1994;369:525. [PubMed: 8031403]

Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 2002;415:429–433. [PubMed: 11807554]

Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1991;1:1–47. [PubMed: 1822724]

Foley JM, Legge GE. Contrast detection and near-threshold discrimination in human vision. Vision Res 1981;21:1041–1053. [PubMed: 7314485]

Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. Nature 1992;360:343–346. [PubMed: 1448150]

Gold J, Bennett PS, Sekuler AB. Signal but not noise changes with perceptual learning. Nature 1999;402:176–178. [PubMed: 10647007]

Graham, NVS. Visual Pattern Analyzers. Oxford Psychology Series. Vol. 16. Oxford, UK: Oxford Univ Press; 1989.

Green, DM. Psychoacoustics and detection theory. In: Swets, JA., editor. Signal Detection and Recognition by Human Observers—Contemporary Readings. New York: Wiley; 1964. p. 58-91.

Green, DM.; Swets, JA. Signal Detection Theory and Psychophysics. New York: Wiley; 1966.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 1999;24:187–203. [PubMed: 10677037]

Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzchak Y, Malach R. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum Brain Mapp 1998;6:316–328. [PubMed: 9704268]

Grill-Spector K, Kushnir T, Hendler T, Malach R. The dynamics of object-selective activation correlate with recognition performance in humans. Nat Neurosci 2000;3:837–843. [PubMed: 10903579]

Heeger DJ, Huk AC, Geisler WS, Albrecht DG. Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? Nat Neurosci 2000;3:631–633. [PubMed: 10862687]

Heeger DJ, Ress D. What does fMRI tell us about neuronal activity? Nat Rev Neurosci 2002;3:142–151. [PubMed: 11836522]

Hubel DH, Wiesel TN. Functional architecture of macaque monkey visual cortex. Proc R Soc Lond B Biol Sci 1977;198:1–59. [PubMed: 20635]

Ito M, Tamura H, Fujita I, Tanaka K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol 1995;73:218–226. [PubMed: 7714567]

Kanwisher N, Chun MM, McDermott J, Ledden PJ. Functional imagining of human visual recognition. Brain Res Cognit Brain Res 1996;5:55–67. [PubMed: 9049071]

Keppel, G.; Wickens, TD. Design and Analysis: A Researcher's Handbook. Upper Saddle River, NJ: Pearson Prentice Hall; 2004.

Knill DC. Discrimination of planar surface slant from texture: human and ideal observers compared. Vision Res 1998;38:1683–1697. [PubMed: 9747503]

Kourtzi Z, Kanwisher N. Cortical regions involved in perceiving object shape. J Neurosci 2000;20:3310–3318. [PubMed: 10777794]

Kourtzi Z, Kanwisher N. Representation of perceived object shape by the human lateral occipital complex. Science 2001;293:1506–1509. [PubMed: 11520991]

Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Res 1995;35:389–412. [PubMed: 7892735]

Legge GE, Kersten D, Burgess AE. Contrast discrimination in noise. J Opt Soc Am A 1987;4:391–404. [PubMed: 3559785]

Lennie P. Single units and visual cortical organization. Perception 1998;27:889–935. [PubMed: 10209632]

Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R. A hierarchical axis of object processing stages in the human visual cortex. Cereb Cortex 2001;11:287–297. [PubMed: 11278192]

Liu Z, Knill DC, Kersten D. Object classification for human and ideal observers. Vision Res 1995;35:549–568. [PubMed: 7900295]

Livingstone M, Hubel D. Segregation of form, color, movement, and depth: anatomy, physiology, and perception. Science 1988;6:740–749. [PubMed: 3283936]

Loftus GR, Masson MEJ. Using confidence intervals in within-subject designs. Psychonomic Bull Rev 1994;1:476–490.

Logothetis NK. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos Trans R Soc Lond B Biol Sci 2002;357:1003–1037. [PubMed: 12217171]

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. Nature 2001;412:150–157. [PubMed: 11449264]

Lu ZL, Dosher B. External noise distinguishes mechanisms of attention. Vision Res 1998;38:1183–1198. [PubMed: 9666987]

Lu ZL, Dosher BA. Characterizing human perceptual inefficiencies with equivalent internal noise. J Opt Soc Am A 1999;16:764–768.

Majaj N, Pelli DG, Kurshan P, Palomares M. The role of spatial frequency channels in letter identification. Vision Res 2002;42:1165–1184. [PubMed: 11997055]

Malach R, Reppas JB, Benson RB, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RBH. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA 1995;92:8135–8138. [PubMed: 7667258]

Marr, D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York: Freeman; 1982.

Nachmias J, Sansbury RV. Letter: grating contrast: discrimination may be better than detection. Vision Res 1974;14:1039–1042. [PubMed: 4432385]

Olman CA, Ugurbil K, Schrater P, Kersten D. BOLD fMRI and psychophysical measurements of contrast response to broadband images. Vision Res 2004;44:669–683. [PubMed: 14751552]

Pelli DG. Uncertainty explains many aspects of visual contrast detection and discrimination. J Opt Soc Am A 1985;2:1508–1532. [PubMed: 4045584]

Pelli, DG. The Quantum Efficiency of Vision. In: Blakemore, C., editor. Visual Coding and Efficiency. Cambridge, UK: Cambridge Univ. Press; 1990.

Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision 1997;10:437–442. [PubMed: 9176953]

Pelli DG, Farell B, Moore DC. The remarkable inefficiency of word recognition. Nature 2003;423:752–756. [PubMed: 12802334]

Raiguel SE, Lagae L, Gulyas B, Orban GA. Response latencies of visual cells in macaque areas V1, V2 and V5. Brain Res 1989;493:155–159. [PubMed: 2776003]

Rainer G, Augath M, Trinath T, Logothetis NK. Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. Curr Biol 2001;11:846–854. [PubMed: 11516645]

Rainer G, Augath M, Trinath T, Logothetis NK. The effect of image scrambling on visual cortical BOLD activity in the anesthetized monkey. Neuroimage 2002;16:607–616. [PubMed: 12169247]

Rees G, Friston K, Koch C. A direct quantitative relationship between the functional properties of human and macaque V5. Nat Neurosci 2000;3:716–723. [PubMed: 10862705]

Ress D, Heeger DJ. Neuronal correlates of perception in early visual cortex. Nat Neurosci 2003;6:414–420. [PubMed: 12627164]

Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nat Neurosci 1999;2:1019–1025. [PubMed: 10526343]

Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG. Signal timing across the macaque visual system. J Neurophysiol 1998;79:3272–3278. [PubMed: 9636126]

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 1995;268:889–893. [PubMed: 7754376]

Solomon JA, Pelli DG. The visual filter mediating letter identification. Nature 1994;369:395–397. [PubMed: 8196766]

Stromeyer CF 3rd, Klein S. Spatial frequency channels in human vision as asymmetric (edge) mechanisms. Vision Res 1974;14:1409–1420. [PubMed: 4446371]

Swets, JA. Signal Detection and Recognition by Human Observers—Contemporary Readings. New York: Wiley; 1964.

Tanner WP Jr. Physiological implications of psychophysical data. Ann NY Acad Sci 1961;89:752–765. [PubMed: 13775211]

Tanner WP Jr, Birdsall TG. Definitions of d' and n as psychophysical measures. J Acoust Soc Am 1958;30:922–928.

Tanner WP Jr, Swets JA. The human use of information. I. Signal detection for the case of the signal known exactly. IRE Trans Inf Theory PGIT 1954;4:213–221.

Tjan, BS. PhD thesis. Minneapolis, MN: University of Minnesota; 1996. Ideal Observer Analysis of Object Recognition.

Tjan BS, Braje WL, Legge GE, Kersten D. Human efficiency for recognizing 3-D objects in luminance noise. Vision Res 1995;35:3053–3069. [PubMed: 8533342]

Ungerleider, LG.; Mishkin, M. Two cortical visual systems. In: Ingle, DJ.; Goodale, MA.; Mansfield, RJW., editors. Analysis of Visual Behavior. Cambridge, MA: MIT Press; 1982. p. 549-586.

Van Essen DC, Anderson CH, Felleman DJ. Information processing in the primate visual system: an integrated systems perspective. Science 1992;255:419–423. [PubMed: 1734518]

Zenger-Landolt B, Heeger DJ. Response suppression in v1 agrees with psychophysics of surround masking. J Neurosci 2003;23:6884–6893. [PubMed: 12890783]

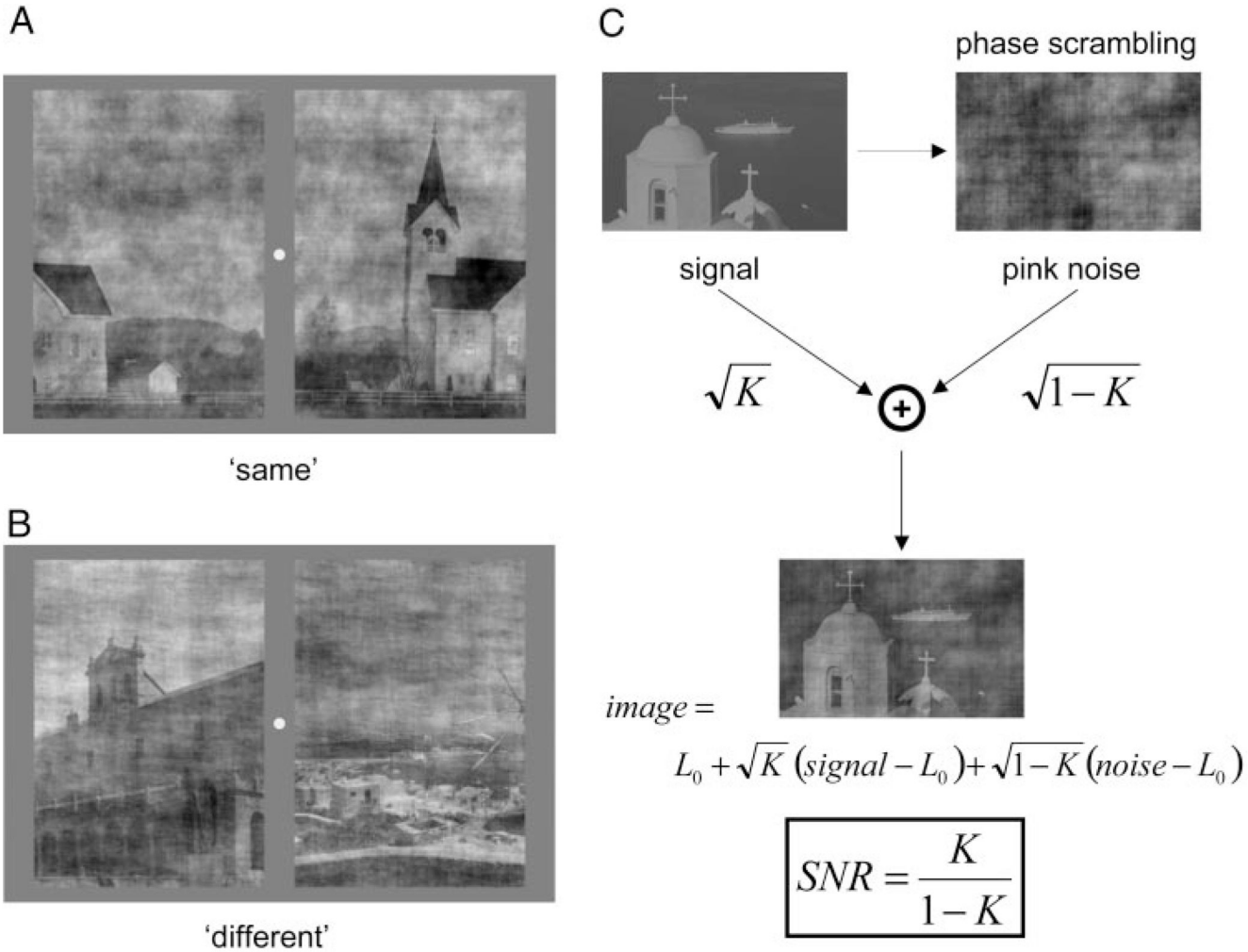

FIG. 1.

Sample stimuli for "same" (*A*) and "different" (*B*) trials. Even in the same trials, the left and right panels were never identical but were pictures taken from the same scene. *C*: pink noise was generated by replacing the phase spectrum of a scene by the phase spectrum of white noise. The signal-to-noise ratio (SNR) of a stimulus was manipulated by adding a pink noise pattern to a noise-free image "on the circle" about the mean luminance such that both the mean luminance and the RMS contrast of the resulting image was unaffected by SNR.

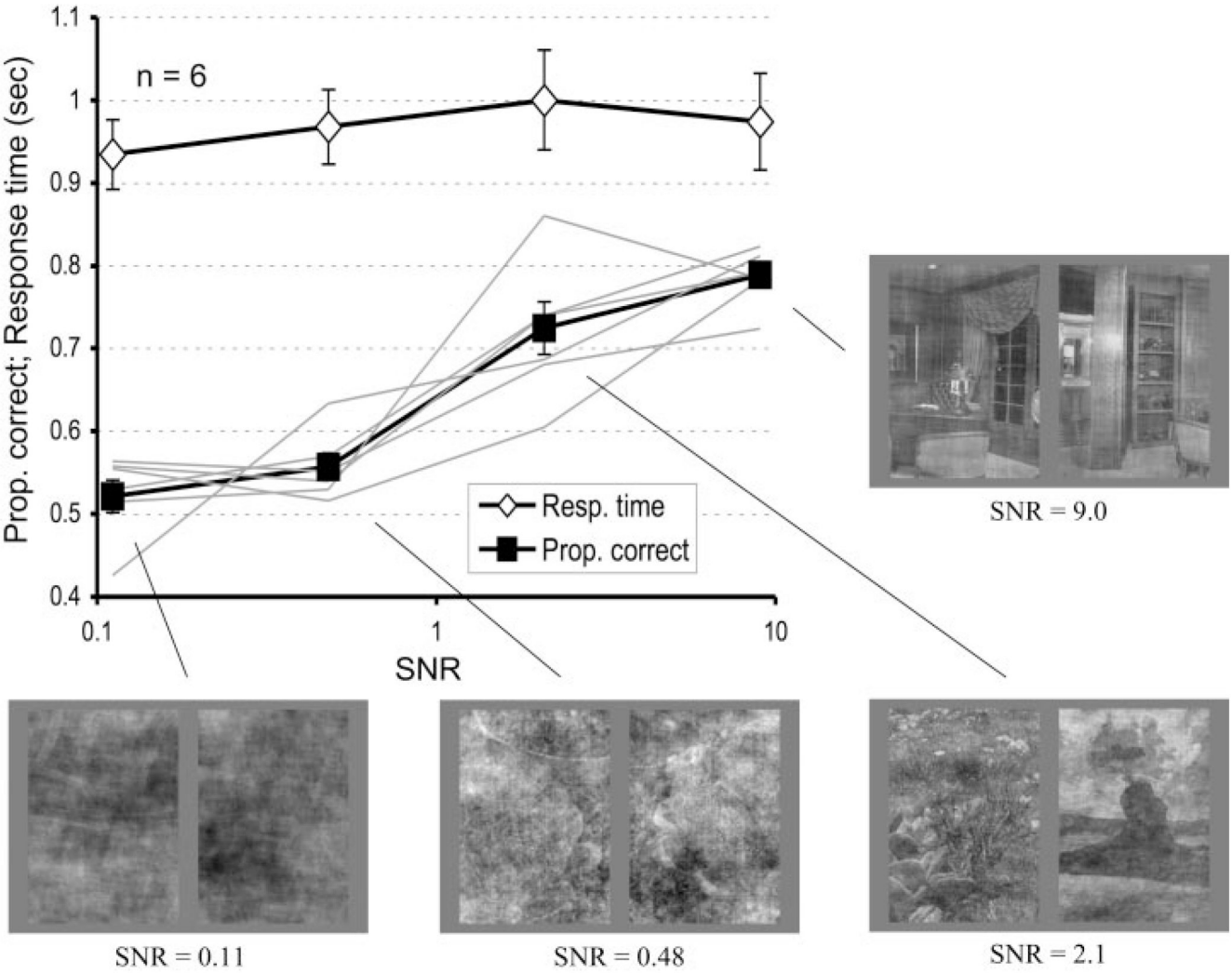

FIG. 2.

Accuracy and reaction times collected during scanning averaged across all subjects, weighed by the number of valid responses (the response box failed in one run for 1 subject). Average accuracy increased as a function of SNR, from an average of 52% at SNR = 0.11 to 79% at SNR = 9.0. Thin gray curves depict accuracy for individual subjects. Average reaction times for correct trials remained at a relatively constant level (0.9 s), unaffected by SNR. Error bars represent standard error of the mean.

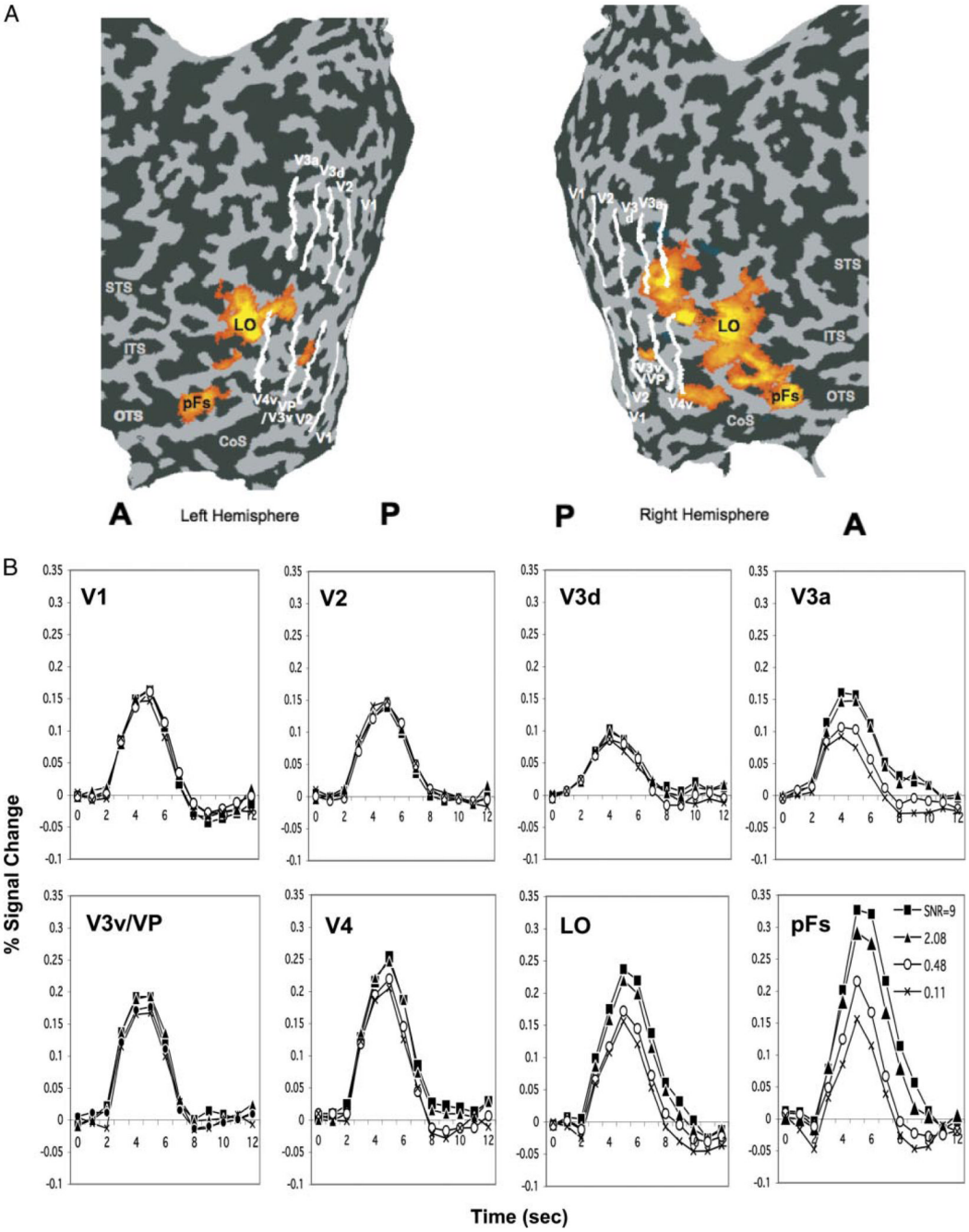

FIG. 3.

Cortical map (*A*) from 1 subject showing the measured regions of interest (ROIs). The ROIs considered in the study were V1, V2, V3d, V3a, V3v/VP, V4, and 2 subregions of the LOC: lateral occipital cortex (LO) and the posterior fusiform sulcus (pFs). *B*: time courses of the blood-oxygen-level-dependent (BOLD) response (% BOLD signal change from fixation baseline) averaged across subjects illustrate the interaction between stimulus SNR and ROI. The modulation of the BOLD response by stimulus SNR increased from posterior (V3d, V3v/VP, V4) to anterior ROIs (V3a, LO, pFs). There was no noticeable effect of SNR on BOLD response in V1 and V2.

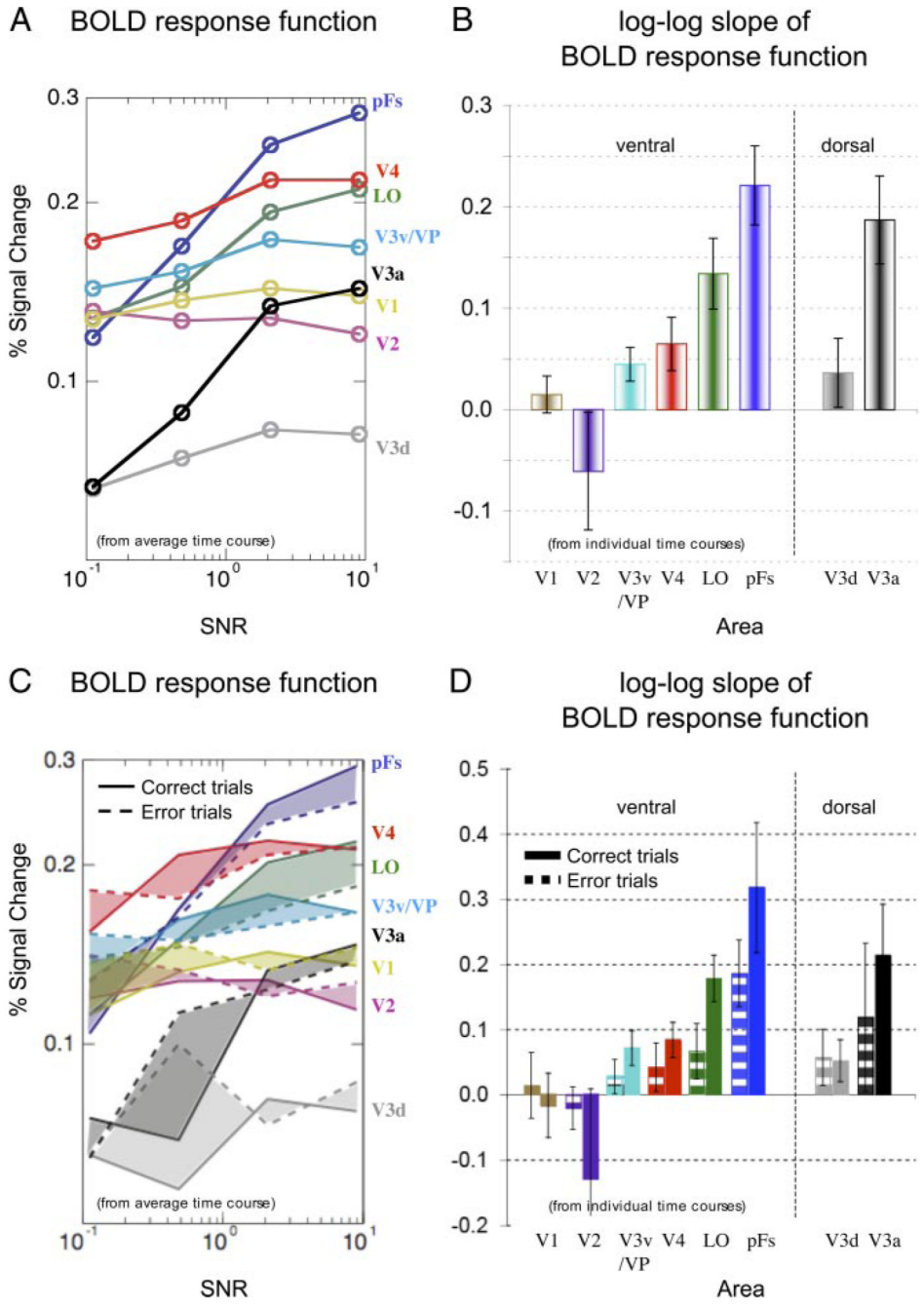

FIG. 4.

The BOLD response functions (average peak % BOLD signal change from fixation baseline vs. stimulus SNR) from the 8 ROIs derived from the average time courses of Fig. 3 are plotted in log-log axes (*A*). The progressive increase in the log-log slope of the BOLD response functions (*B*) along the ventral and dorsal visual pathways were confirmed by within-subject ANOVA and planned polynomial contrasts (see text). The BOLD response functions (*C*) and their log-log slopes (*D*) were also derived separately from the correct and incorrect trials. There was no statistically significant effect of response correctness on the log-log slopes, nor was there any interaction between behavioral correctness and ROI (see text). Error bars represent within-subject SE (Loftus and Masson 1994).

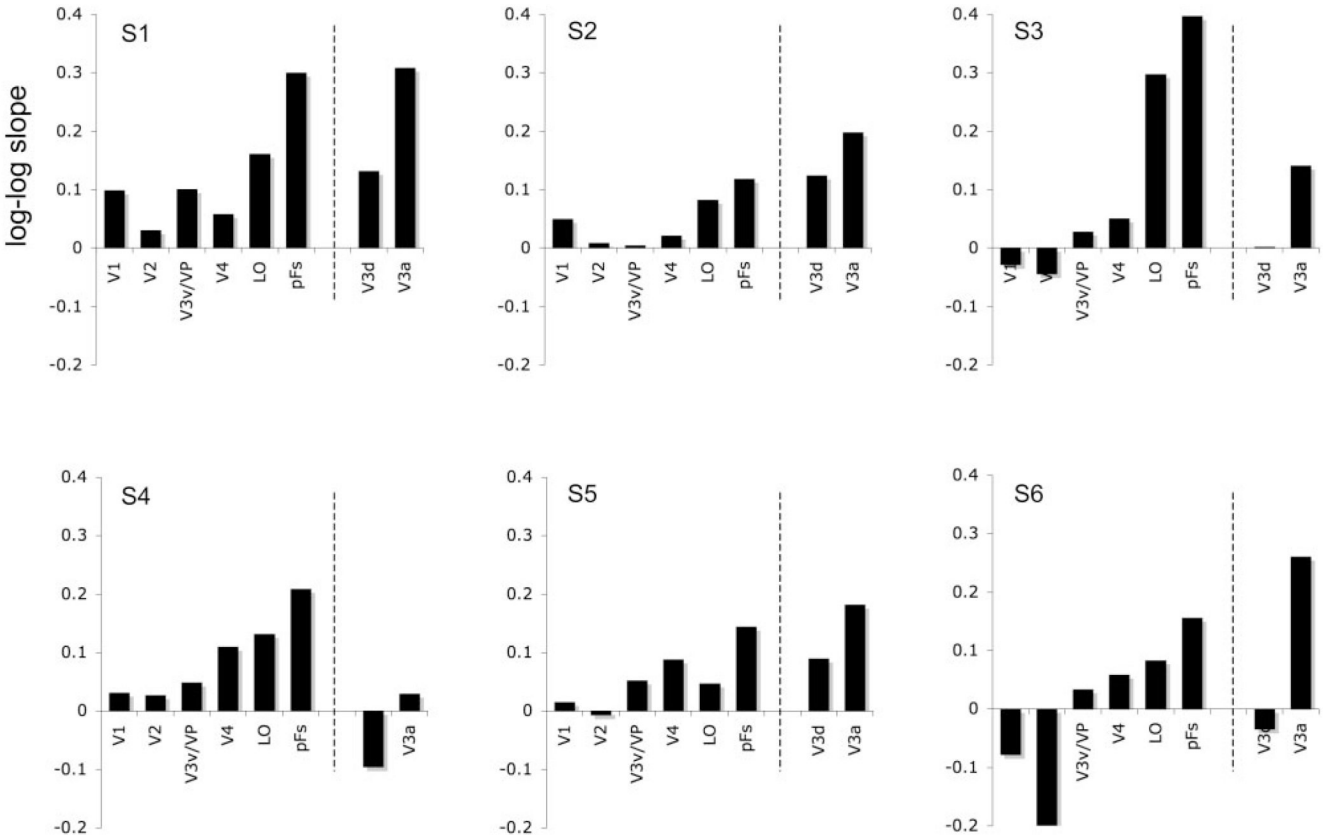

FIG. 5.

The log-log slope of the BOLD response functions from individual subjects. Between-subjects agreement regarding the rank order of the log-log slopes was confirmed with Kendall's *W* test (see text).