Color Sensitivity of Cells Responsive to Complex Stimuli in the Temporal Cortex

Robin Edwards,1 Dengke Xiao,1 Christian Keysers,2 Peter Földiákov1, and David Perrett1
1School of Psychology, University of St. Andrews, St. Andrews, Fife KY16 9JU, United Kingdom; and 2Istituto di Fisiologia Umana, Università di Parma, 43100 Parma, Italy

Submitted 8 July 2002; accepted in final form 9 February 2003

INTRODUCTION

The visual system is often described as a hierarchy of processing modules, each specialized to a particular attribute of the visual scene. However, more recent work (Sugase et al. 1999) has suggested that there may also be a temporal hierarchy of information processing in the visual system with coarse information about a stimulus becoming available before the fine detail. Recording from neurons in inferior temporal cortex (IT), they found that the early part of a cell’s response related only to the global stimulus category (e.g., face or nonface) with the more detailed information about the stimulus (e.g., facial expression) available only in the later (>50 ms) component of the response. The idea that processing in the visual system proceeds in a coarse to fine manner has been proposed by a number of authors (Carpenter and Grossberg 1987; Nowak and Bullier 1997; Schyns and Oliva 1994; Ullman 1995). Coarse to fine processing in terms of spatial frequency is proposed by Parker et al. (1992) to explain differential reaction times to sinusoidal gratings. Delorme et al. (2000) found that, in a rapid categorization task, color makes little difference to speed or accuracy, leading to their suggestion that the first wave of visual information is essentially low spatial frequency and achromatic, with color and the high spatial frequencies following later.

A physiological basis for these theories may lie in the differential processing speeds of the magnocellular (M) and parvocellular (P) pathways from the retina to area V1 of the cortex. The distinction between M and P pathways originates in the primate retina (for reviews, see Merigan and Maunsell 1993; Milner and Goodale 1985) with two distinct classes of retinal ganglion cells. One class (the P β cells) has large cell bodies and dendritic radiations, a transient response to visual stimulation, and because all three cone types make excitatory synapses on these cells, they are spectrally broad-band. In contrast, the second class of cells (P β) has medium- to small-sized cell bodies, small dendritic radiations, produce sustained responses, and generally receive excitatory synapses from only one or two types of cones, thus conferring spectral sensitivity. The outputs from these two classes of retinal cells form two anatomically distinct pathways to visual cortex. The M pathway begins with P α cells, which project to the magnocellular layers of the lateral geniculate nucleus (LGN) and thence to layer 4Cα of cortical area V1. In contrast, P β cells project to the parvocellular layers of the LGN and thence to layer 4Cβ of V1.

The two systems can be said to transmit different regions of the “window of visibility” (Watson and Ahumada 1985) in that the P system appears to provide greater spatial resolution and color selectivity and responds to slowly changing stimuli. In contrast, the M system is effectively color-blind but is much more sensitive to rapidly changing stimuli (Merigan and Maunsell 1993). It was once thought (Livingstone and Hubel 1988) that the ventral visual stream was dominated by input from the P system, but more recent work (Ferrera et al. 1994) has shown that, in V4, neurons may be driven by either system, and there

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
is an almost equal contribution by the two inputs across the whole population of V4 cells.

Consistent differences have been observed between the latency of the magnocellular and parvocellular pathways. In the LGN, the earliest magnocellular responses precede parvocellular responses by 10 ms (Marrocco 1976; Maunsell et al. 1999) and, in V1, the latencies in the parvo-recipient layer 4C are ~20 ms longer compared with 4Ca (Nowak et al. 1995).

If an achromatic low-frequency signal about a stimulus reaches V1 prior to the color high-frequency information, does the visual system compensate for this latency difference in the later stages of processing or might this be a plausible explanation for the latency difference between global and fine information found by Sugase et al. (1999)?

Evidence from physiology

The extent to which visual neurons in IT cortex are sensitive to color is not clear from the literature. In the macaque, Tanaka et al. (1991) found that color was relevant for only ~10% of their sample and that proportion fell to ~7% for those cells whose receptive fields were considered to be elaborate (responding to a face or particular combination of shape and texture). An earlier study (Gross et al. 1972) had detected color sensitivity in some cells in IT, but the extent of its importance was not accurately quantified.

A study specifically looking at color sensitivity in anterior IT was carried out by Komatsu et al. (1992). The color of simple geometric shapes varied systematically, and color was found to influence the response of the vast majority of cells tested (~90%). Unfortunately, color was only studied with those cells that could be driven by simple shapes, and cells responding to more complex stimuli (such as faces and natural scenes) were not tested.

A second study (Komatsu and Ideura 1993) varied shape (simple geometric figures) and texture in addition to color. Again, a high degree of color sensitivity was found, with 69% of neurons selective for color (similar to the proportion found to be shape selective). There was no evidence of any interaction between shape and color selectivity (i.e., the color preference of a cell did not depend on the shape preference).

Perrett et al. (1982) recorded from face selective cells in rhesus superior temporal sulcus (STS) and noted that color did not seem to play an important role, with only 1 of 18 cells showing a reduction in response when the faces were viewed through a color filter. However, there has been little systematic study of the effect of color in neurons with complex stimulus selectivity (or indeed the latency of color information), and we must look beyond physiology for further evidence of its role.

Lesion and imaging studies

In addition to the role that IT cortex plays in object recognition, there have been a number of studies examining the effect of lesions on color processing. Heywood et al. (1995) found almost complete impairment of hue discrimination after ablation of the inferior temporal lobe in macaque monkeys yet luminance discrimination was relatively spared. Similarly, Dean (1979) found no retention of a color discrimination task after IT ablation, although after retraining, hue discrimination thresholds were found to be unaltered with respect to their preoperative levels. It is possible, however, that slight departures from isoluminance in this study may have allowed the monkeys to learn to discriminate color based on spared luminance discrimination. Horel (1994) used cold-suppression of the dorsal aspect of IT to examine color and form discrimination in trained macaques, finding disruption of color discrimination even though form discrimination was spared.

Imaging studies also indicate that IT has a role in the processing of color. Takechi et al. (1997) used PET to look at the cortical areas involved in color, luminance and positional discrimination (again in the macaque). Using simple square stimuli, they found significant activation in the posterior part of IT cortex in the (color-brightness) and (color-position) subtraction pairs.

fMRI imaging has been used in the human (Zeki and Marini 1998) to look at the areas involved in color processing with more natural stimuli. Their stimuli included common objects and landscapes that were presented in full color, achromatic, or false color conditions. The (color-achromatic) subtraction resulted in an area of activation extending anteriorly beyond V4 in the temporal lobe into areas the authors suggest may be analogous to monkey IT. Interestingly, this activation is not found in the (false-achromatic) subtraction, suggesting that the role of this inferior temporal area anterior to V4 is specific to the role color plays in object recognition rather than simply processing color in a more abstract capacity.

Together, the preceding studies suggest that color processing is an important property of inferior temporal cortex in both the monkey and human, though the extent to which this is specific to object processing is unclear (particularly in human).

Psychophysics

Psychophysical studies provide the majority of evidence that color information is delayed with respect to form in the visual system, although there has been a great amount of debate in the literature about the precise role it plays in object recognition.

Several studies have indicated that color is unimportant for object recognition with subjects responding just as quickly and accurately to black and white line drawings or photographs when compared with color photographs (Biederman and Ju 1988; Davidoff and Ostergaard 1988; Ostergaard and Davidoff 1985). However, there has been a greater number of studies that do not support this position, finding that color and detail consistently enhance the speed of reaction or accuracy (Humphrey et al. 1994; Lee and Perrett 1997; Price and Humphreys 1989; Wurm et al. 1993).

There are major methodological differences between all the studies and a potential explanation for the conflicting data are given by Price and Humphreys (1989), who suggest that no advantage for color is found when the subjects are required to make very fast decisions or the stimuli are masked. This contrasts with the situation when subjects can react in their own time where a consistent advantage for color is found. Perhaps subjects reacting quickly have access only to coarse representation, which does not include color.

Delorme et al. (2000) required both monkey and human subjects to make a rapid categorization (food or nonfood/animal or nonanimal) of briefly presented (32 ms) stimuli, which were either color or achromatic photographs. Color was found to make little difference in terms of either accuracy or
reaction time in the majority of subjects. Interestingly, they noted that there was a strong correlation between the accuracy impairment for achromatic images and reaction time with the slowest reacting subjects showing the highest impairment with achromatic images and the faster subjects categorizing equally well in both conditions. Like Price and Humphreys (1989), they suggest that color is not an important cue when reactions must be made quickly, but that it can be used as a relevant feature when a subject takes longer to respond.

The work summarized in the preceding text suggests that the color information is used by the visual neurons in temporal cortex, but the extent of its importance in the processing of complex visual stimuli is unclear. This study examines the importance of color to cell responses to complex images in the temporal cortex and evaluates the hypothesis that the color signal may be delayed with respect to the signal about form defined from luminance cues. We tested cells at different stimulus presentation speeds and predicted that as presentation rate increased, sensitivity to color information in the stimulus would diminish because the color signal would not arrive in time to contribute to the response.

**METHODS**

**Recording techniques**

The subject (male *Macaca mulatta*, age 6 yr) was seated in a primate chair and head restrained. Neural signals were recorded using standard methods (Oram and Perrett 1992). Neurons were localized to the upper and lower banks of the superior temporal sulcus and inferotemporal cortex (see Fig. 1). The subject’s eye position was monitored (accuracy ±1°; IView, SMI). A 486 PC and Cambridge Electronics CED 1401 interface recorded eye position and spike arrival times and measured stimulus onset times.

**Stimulus presentation**

Stimuli (256 x 256 pixels) were presented centrally on a Sony GDM-20D11 monitor (72-Hz refresh rate, image size: 10 x 10°) that was attached to an Indigo2 Silicon Graphics workstation. Stimuli were presented against a black background. Onset and duration of the stimuli were measured using light-sensitive diodes on the monitor screen. If the measured stimulus duration differed from the intended duration, the data for that stimulus sequence were discarded. Sequence presentation commenced when the subject’s gaze remained within a fixation window ±5° of the monitor center for >500 ms and terminated if the subject’s gaze moved outside the fixation window. Fixation was rewarded with fruit juice delivery. Activity relating to the first and last image of each sequence was discarded.

**Visual stimuli**

The stimulus set consisted of 38 color images (256 x 256 pixels) including photographs of human and monkey heads, animals, everyday objects, and abstract figures. The images included photographs of two monkeys and one human taken from all angles at 45° intervals, where 0° was the front (or facial) view of the head and 180° was the rear.

False-color and achromatic versions of each image were prepared as outlined below. First, images were transformed to YCbCr color space, which has separate luminance (corresponding to the CIE Y primary) and chromaticity components (Bhaskaran and Konstantinides 1997).

Achromatic images were generated simply by setting the chromatic components (Cb and Cr) to zero for each pixel, followed by a transform back to RGB. This process always produces a valid RGB triplet.

False-color images were prepared by reflecting each pixel in turn about the origin of the chromatic (CbCr) plane, keeping Y constant. The process often generates invalid RGB triplets that correspond to colors that cannot be produced on a standard monitor (e.g., very bright pure blue). When this occurred, points were moved back toward the origin of the chromatic plane until displayable colors were obtained (this had the effect of reducing the saturation of the color).

A digital photometer (Tektronix, Model J6523–2) was used to test the success of these transformations in maintaining both overall image luminance and contrast edges within images. Overall image luminance was measured by placing a perspex diffusion plate between the computer screen and photometer. Contrast borders were tested by
individually measuring a series of color patches (1° diam) before and after image transformation. The process was judged to be a satisfactory for the purposes of the experiment, with measured luminances falling within ±10% of the pretransformation levels.

Procedure

SCREENING PHASE. On detection of a neuron, we first conducted a screening test to establish whether any of the stimuli were sufficient to evoke a visual response.

Each cell was tested using either the 38 color images or 38 achromatic images in addition to the original 38 color images. Stimuli were presented in a random sequence with a stimulus presentation time of 111 ms (8 frames) and no gap between stimuli. In a previous study, we showed that stimulus selectivity measured at 111 ms/image is identical to that measured at slower presentation rates in IT and STSa (Keysers et al. 2001). Testing neurons at that presentation rate does therefore not select for a particularly fast-discriminating subset of neurons. Each stimulus was repeated several times during the screening phase (range 9–58). Preference was judged by eye from the set of peristimulus time histograms (PSTHs) computed on-line.

COLOR RESPONSE TEST. A total of 50 cells were found to have a preference for at least one of the stimuli during the screening phase. For each cell, we selected its “best” and “worst” images (those producing highest and lowest responses respectively) from the screening set, along with three other images that had produced intermediate responses. This was intended to reduce the contamination produced by adjacent stimuli because it has been shown that the neural response to a stimulus typically outlasts stimulus presentation time by ~60 ms (Keysers and Perrett 2002; Keysers et al. 2001).

A cell was then tested using a “test set” of stimuli for each of the images chosen above. Each set included the color image, achromatic, and false-color transformed versions as well as an additional set of four frequency-filtered color and achromatic versions of the image that will be reported elsewhere. In total, the complete test set was comprised of 35 stimuli/cell.

For the first 11 cells, no false-color images were present in the test set. These cells were tested using only six versions of each image, with a total test set of 30 stimuli/cell.

The stimuli were presented in four different stimulus duration conditions, as detailed in Table 1. Within each condition, stimuli were presented in a random order with the constraint that no two identical stimuli were ever presented in immediate succession. Presentations for each condition were also randomly interleaved. The total presentation time was identical for each of the four conditions with more stimulus repetitions of the shorter duration conditions. The median number of repeats of the slowest condition (S56G167) was 17 (range: 6–42) and the median number of repeats of the fastest condition (S14G0) was 209 (range: 119–333).

Condition S14G0 was only tested when a sufficiently strong response was obtained and was often carried out post hoc in isolation, after the data from the other conditions had been examined. Only seven cells were tested at this rate.

TABLE 1. Experimental conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Stimulus, ms</th>
<th>Gap, ms</th>
<th>Total, ms</th>
<th>Time Taken for 1 Complete Cycle, s</th>
</tr>
</thead>
<tbody>
<tr>
<td>S56G167</td>
<td>55.6</td>
<td>166.7</td>
<td>222.2</td>
<td>7.8</td>
</tr>
<tr>
<td>S56G0</td>
<td>55.6</td>
<td>—</td>
<td>55.6</td>
<td>1.9</td>
</tr>
<tr>
<td>S28G0</td>
<td>27.8</td>
<td>—</td>
<td>27.8</td>
<td>1.0</td>
</tr>
<tr>
<td>S14G0</td>
<td>13.9</td>
<td>—</td>
<td>13.9</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Response analysis

A cell’s response to a particular stimulus in the sequence was calculated by aligning segments, in the continuous recording, on each occurrence of that particular stimulus. Each segment lasted from 250 ms before stimulus onset to 550 ms after stimulus offset. The PSTH was generated by summing across all the aligned segments and represented the response triggered by that particular stimulus against a background of activity evoked by all the surrounding stimuli.

“Best” and “worst” stimuli for a cell at the end of the screening phase were simply judged by eye from the set of PSTHs calculated on-line and were selected (along with 3 intermediate images) for the color response test.

After the color-response test, the sets of stimuli were ranked from best to worst separately for each stimulus duration condition based on the cell’s response to the color. This ranking and the cell latency was calculated as follows: first, the responses were summed across trials (bin-size = 1 ms) and smoothed (Gaussian, σ = 20 ms). A control period was defined as the 200 ms preceding stimulus onset. The latency of response onset was measured as the first 1-ms time bin at which the firing rate exceeded the mean ±2.58σ (i.e., P < 0.005) of activity during the control period, for ≥15 consecutive bins (i.e., 15 ms). Where this criterion was not met, a fixed latency of 100 ms was assumed for that stimulus. If a latency could not be detected for any of the stimuli, the cell was excluded from the analysis of that condition.

Next, the response to each stimulus was measured in a time window starting with the latency measured in the preceding test and lasting for the length of the stimulus + half the Gaussian width of the smoothing filter (10 ms). Stimulus sets were ranked according to this windowed response, and cell latency was defined as being the onset time of the maximum response.

Population analysis

A spike density function (SDF) was calculated from the raw spike counts for every cell and stimulus by smoothing with a Gaussian (σ = 5 ms).

A single normalizing factor for each cell was calculated as the maximum value of the SDF for the color version of the best stimulus. Color, achromatic, and false-color responses to the best stimulus were weighted by this factor, so that every cell would have an equal contribution in the population response with the color response acting as baseline.

Finally, population curves were calculated for each condition as the average SDF for color, achromatic, and false-color versions of the best stimulus.

For the latency-aligned population curves, an additional step took place prior to averaging, with each SDF shifted in time such that time 0 reflected the detected cell latency, as measured in the preceding text.

RESULTS

Stimulus preference

Stimuli were categorized as being abstract if no natural coloring existed for that image. Those stimuli that had natural coloring were broken down into faces (which included all head views) and non-faces. The category of the preferred (or “best”) stimulus for each cell is shown in Table 2. Where a cell’s preferred stimulus was not consistent across presentation rate conditions, or if no response was obtained for any condition (excluding S14G0 because only 7 cells were tested at this rate), the cell was categorized as having unknown preference. Stimulus preference is shown with cell location in Fig. 1.

We first consider the effect of color on the whole population...
of cells tested. Although it seems unlikely that screening with color stimuli alone introduced a sampling bias in favor of chromatically tuned cells (because luminance information is still present in these color images), this point is considered later in this section.

Whole population analysis

Neurons in STS/IT exhibit a high degree of chromatic tuning in addition to their shape tuning as illustrated by the response of a single cell (Fig. 2) and the population response to the best stimulus (Fig. 3). Achromatic versions of the best stimulus produce, at a population level, much weaker responses than the original color images. The reduction in response is even greater when stimuli are falsely colored, suggesting that these cells are not only have a preference for certain color profiles, but this tuning extends to inhibition of the shape response, when an incorrect color profile is present.

This initial qualitative description is backed by a formal statistical treatment of the color and shape responses in the following sections.

The population includes cells with a wide range of response latency (median: 91, range: 58–141 ms), and we can see the extent of chromatic tuning even more clearly if this variation in response latency is removed by aligning the data from each cell on the cell latency (Fig. 3B).

If we are correct in our hypothesis that the color signal will be delayed with respect to the luminance signal, then given the effects noted above, we should see two separate effects in the population response. 1) Color, achromatic, and false-color curves should be initially similar, then diverge, reflecting the delayed contribution of color-specific information to the response. 2) Any difference between the curves should be progressively reduced as presentation rate increases because the appropriate color information will not arrive in time to contribute to the response.

<table>
<thead>
<tr>
<th>Preferred Stimulus Category</th>
<th>Number of Cells</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Non-abstract</td>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>Face</td>
<td>(22)</td>
<td>(44)</td>
</tr>
<tr>
<td>Non-face</td>
<td>(4)</td>
<td>(8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>18</td>
<td>36</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

FIG. 2. Response of a single cell to color, achromatic, and false-color versions of its best stimulus in different presentation rate conditions [56 ms + 167 ms gap (top), 56 ms (middle), 28 ms (bottom)]. This particular cell was not tested at the fastest rate of 14 ms. Spike density functions (SDFs) were created by summing the data across trials and smoothing with a gaussian (\(\sigma = 10\) ms). \(\bullet\), stimulus presentation; \(\square\), previous or subsequent stimulus presentations. The latency of this cell (66 ms) was at the lower end of the cell population tested. Scale is indicated in the bottom right corner.
In fact, neither of these effects is seen. Preference for the color version of the stimulus appears almost immediately at the start of the population response (Fig. 3B) and this preference is consistent across the different presentation rate conditions, and is clearly apparent even at the fastest presentation rate of 14 ms.

**TIME COURSE OF COLOR AND SHAPE TUNING—POPULATION STATISTICAL ANALYSIS.** The time course of shape and color tuning was accurately established using a sliding window statistical test to measure the probability of discrimination between stimuli as a function of time.

For each cell, and each presentation rate, normalized and latency-aligned SDFs were calculated as previously described. Shape tuning was then determined by comparing the SDFs for the achromatic stimuli with related one-way ANOVA (5 levels, corresponding to the 5 different stimuli) and one entry per cell, performed separately for each 1-ms time bin. Color tuning was measured in a similar manner, with a one-way ANOVA (3 levels) comparing the responses to color, achromatic, and false-color versions of the best stimulus for each cell. The analysis was restricted to the subset of cells for which we had collected false-color responses. There was an insufficient number of cells tested at the 14-ms presentation rate to perform the analysis for this condition.

The results of this analysis can been see in Fig. 4. Shape and color discrimination have an almost equal onset time, with shape leading color by not more than 5 ms at the uncorrected \( P = 0.01 \) level (lower dashed line) and no difference when considered the same criterion level with Bonferroni correction (upper dashed line). The Bonferroni level was calculated by dividing the criterion level \( (P = 0.01) \) by the number of time bins over which the analysis was performed (500) and is, in fact, over-corrected because there is a high degree of correlation between consecutive time bins due to the response properties of the cells themselves and second due to the smoothing procedure used to create the SDFs.

If we consider the overall pattern of color and shape discrimination, the data appear to suggest that in fact, the color signal dominates the earlier part of the response, with optimal color discrimination peaking before that of shape by 10–20 ms. The latter part of the response is dominated by shape discrimination, which outlasts color discrimination by ~20 ms in the zero gap conditions.

This pattern is particularly evident in slowest presentation condition, where a 168-ms gap is present between each stimulus. During the entire gap period, the response continues to carry information on the shape of the previous stimulus and...
this is extinguished only when the next stimulus is presented. However, this “visual memory” for stimulus shape appears to be color blind because color discrimination is as short lived as in the conditions where no gap exists.

COLOR TUNING OF INDIVIDUAL CELLS The color responses of individual cells were assessed by comparing windowed spike counts for each individual trial for the best color, achromatic and false-color (when presented) stimuli. The window was determined from the population analysis above, and was calculated separately for each condition as follows. The window began with the first 1 ms time bin when color discrimination exceeded the \( P = 0.01 \) (uncorrected) level. It ended when population color discrimination fell back below this criterion level with the constraint that the window was at least as long as the stimulus presentation time.

Color sensitivity was measured with a one-way ANOVA with two or three levels corresponding to color, achromatic, and false color responses (false color responses were only measured in a subset of the cells).

The variance of spike counts has previously been found to be approximately proportional to the mean response (Dean 1981; Tolhurst et al. 1981), therefore it was first necessary to perform a square-root transform on the raw spike counts prior to the statistical tests.

Table 3 shows the proportion of cells where a significant effect of color was found for each condition. Seventy percent of the cells in our population showed a significant effect of color in at least one of the conditions (\( P < 0.05 \), corrected). This breaks down into 72% of the cells screened using just color stimuli and 67% of the cells screened using color and achromatic stimuli. This difference was not significant (\( \chi^2 = 0.70, df = 2, \text{n.s.} \)), and it therefore seems unlikely that there
was any substantial sampling bias toward color sensitive cells when screening took place with color stimuli alone.

**EFFECT OF COLOR ON OBJECT DISCRIMINATION.** The results presented in the previous sections, showing a large decrease in response to a preferred pattern when color information is removed, suggest that neurons in STS/IT are less well able to discriminate between different objects when color cues are inappropriate or absent.

This hypothesis was explicitly tested by carrying out three one-way ANOVAs (each with 5 levels of object) separately over the populations’ responses to colored, achromatic, and falsely colored versions of the stimuli. This produces a measure of the populations’ ability to discriminate between different objects in the three color conditions, and the results of this analysis are shown in Fig. 5.

It can be seen that discrimination between different objects is greatest when the color versions of the stimuli are viewed. The probability of discrimination (based on the neural response) is reduced when achromatic versions of the stimuli are presented and reduced further still when the stimuli are inappropriately colored.

**COLOR SENSITIVITY INDEX.** To assess whether there was any relationship between color sensitivity and cell latency, a color sensitivity index was calculated for each cell

\[
\text{colour sensitivity index} = \frac{c - a}{c + a}
\]

where \(c\) was the mean spike count for the best color stimulus and \(a\) was the mean spike count for the achromatic version of the best stimulus. The window was calculated separately for each presentation rate as described in the previous section. The color sensitivity index ranges from \(+1\) (preference for color stimuli) to \(-1\) (preference for achromatic stimuli) with a value of 0 indicating that a cell responded identically to the color and achromatic stimuli. Extreme values were unlikely to be obtained, however, because spike counts were not corrected by subtraction of the background firing rate from stimulus response (this resulted in a measure that was too sensitive to noise).

Figure 6 shows color sensitivity plotted against latency for each cell. Latency and color sensitivity has been averaged across the three conditions (where possible) to produce a single figure for each cell. There is a negative correlation between latency and color sensitivity (\(r = -0.394, n = 50, P < 0.01, 2\)-tailed), with the most color sensitive cells tending to respond earliest at \(\sim 70–90\) ms.

**Face-Selective subpopulation analysis**

Almost half our cells (\(n = 22\)) consistently responded best to an image of a face or head view. The effect of color on this subpopulation of face-selective cells (Fig. 7) is consistent with the effect of color on the population overall with the original (or naturally colored images) producing far greater responses that achromatic or falsely colored images.

![Figure 5](image_url) **FIG. 5.** The effect of color condition on shape discrimination. Shape tuning in each case was determined by a related 1-way ANOVA (5 levels) and compared responses to the color stimuli (top curve), achromatic stimuli (middle curve), or falsely colored stimuli (bottom curve). This analysis was performed separately for each 1-ms time bin. Probability is plotted on a negative log scale, and curtailed at \(-\log_{10}(P) = 17\). Bottom dotted line shows significance at the \(P = 0.01\) level (uncorrected). Top dotted line shows the same level Bonferroni corrected. (see Fig. 4 for additional details.)

![Figure 6](image_url) **FIG. 6.** Color sensitivity and cell latency. Color sensitivity was defined as the difference between color and achromatic responses over the sum of color and achromatic responses. Increasingly positive values represent increasing preference for color, increasingly negative values represent a preference for the achromatic stimuli. A value of \(-0\) indicates similar responses to color and achromatic stimuli. Color sensitivity and latency figures were averaged across all conditions. •, those cells that had a significant effect of color (\(P < 0.05\) corrected) in ≥1 condition. ••, cells where no significant effect was found. ••••, the least-squares regression for all cells.
COLOR TUNING OF INDIVIDUAL CELLS. The color tuning of individual face-selective cells was measured using the same technique as previously described in the overall population analysis (Table 4). Sixty-eight percent of these cells showed a significant effect of color in at least one of the conditions ($P < 0.05$, corrected).

COLOR SENSITIVITY AND LATENCY. While face-responsive cells are present across the full range of latencies, they tended to have shorter latencies than nonface cells (face cells, mean latency = 85 ms; non-face cells, mean latency = 102 ms). This difference was found to be significant ($t = 2.91$, df = 30, $P < 0.01$, 2-tailed).

The negative correlation between latency and color sensitivity found for the population as a whole is largely due to the contribution of a number of face-selective cells. If the analysis is restricted to the face-selective sub-population, a significant correlation is obtained ($r = -0.438$, $n = 22$, $P < 0.05$ 2-tailed), with those face-selective cells responding earliest tending to be the most strongly color tuned.

COLOR AIDS IN THE DISCRIMINATION OF FACE ORIENTATION. A particularly dramatic example of a color-sensitive face-tuned cell (color sensitivity index = 0.56) can be seen in Fig. 8 with data that were obtained during a screening phase. The cell responds best to the front view of a face with the response

![FIG. 7. Stimulus (A)- and latency (B)-aligned normalized average responses to the color, achromatic, and false-color versions of the best stimulus. This analysis is restricted to the subpopulation of cells whose preferred stimulus was a face. Top: slowest presentation rate; bottom: fastest presentation rate. □ stimulus presentation; □, earlier or subsequent stimulus presentations. Each set of responses for a cell was aligned to the detected latency of the color response as described in METHODS.](image)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Stimulus Presentation Time, ms</th>
<th>Gap Time, ms</th>
<th>n</th>
<th>Mean Response Onset, ms</th>
<th>Window Duration, ms</th>
<th>Percentage of Neurons Discriminating Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>S56G167</td>
<td>55.6</td>
<td>166.7</td>
<td>22</td>
<td>91</td>
<td>51</td>
<td>36</td>
</tr>
<tr>
<td>S56G0</td>
<td>55.6</td>
<td>—</td>
<td>22</td>
<td>87</td>
<td>59</td>
<td>72</td>
</tr>
<tr>
<td>S28</td>
<td>27.8</td>
<td>—</td>
<td>22</td>
<td>78</td>
<td>58</td>
<td>72</td>
</tr>
<tr>
<td>Any</td>
<td>—</td>
<td>—</td>
<td>22</td>
<td>—</td>
<td>—</td>
<td>68†</td>
</tr>
</tbody>
</table>

The last column is percent of neurons with a significant ANOVA testing the effect of stimulus color on neuron response ($P < 0.05$). * Percent of neurons with a significant effect of color in at least one condition ($P < 0.05$, Bonferroni corrected).
The proportion of color-tuned cells

The proportion of cells showing a significant level of color tuning is ~70%. This is remarkably similar to the figure obtained by Komatsu et al. (1992) using simple color patches and would suggest that there is a high degree of color tuning in IT regardless of whether the shape selectivity of a cell is simple (i.e., capable of being driven by simple geometric shapes), as...
The association between species-specific colors with particular objects. In their study, or complex (e.g., selective for face and head view), as in the present study.

The range of latencies of those cells categorized as face-selective was 66–123 ms, with the sample skewed toward the earlier part of this range (mean: 85 ms). The color insensitivity of face cells noted by Perrett et al. (1982) was from a population of cells with somewhat longer latencies than this (range: 80–180 ms, mean not provided but ~125 ms). This might suggest that the presence of separate color-sensitive and -insensitive populations of face cells in temporal cortex with those producing color invariant responses having longer latency. There is also evidence to support this view in our results (see also Fig. 6), with a significant negative correlation between latency and color sensitivity both for the face-selective subset and the population as a whole.

In a later study of face-selective cells in temporal cortex, Perrett et al. (1992) note that cells which are view-independent (i.e., respond equally to different views of the head) tend to have longer latencies (by ~10 ms) than those that are view-selective. They suggest this may be evidence of a hierarchical processing scheme, where the outputs from cells responsive to particular views of an object synapse on a single neuron upstream to produce the view-invariant response. We might speculate that the negative correlation we find between color sensitivity and latency is evidence of an analogous system of processing for object color, where several cells, each tuned for a different possible color of an object, make synapses on a single upstream neuron to produce a color-invariant response.

Effect of color on object discrimination

This study has shown that neurons in STS/IT are less well able to discriminate between different objects when color cues are inappropriate or absent (Fig. 5). This provides a neural basis for interpreting the results of the behavioral recognition studies discussed in the introduction.

It is also clear that the presence of color in itself is not sufficient to enhance the discrimination of objects, rather it is the association between specific colors with particular objects. Thus false color images, which contain both color contrast and luminance contrast are discriminated less well than achromatic images, which contain luminance contrast alone.

Nature of color tuning

Many of the stimuli used in this experiment had natural colors, i.e., colors that are generally associated with that particular pattern (e.g., a face), and the response to these stimuli was often greatly suppressed when the images were falsely colored. It seems possible that inferotemporal cortex contains cells that become tuned by visual experience to the specific conjunctions of color and shape that represent commonly occurring objects in the visual world. For instance, although many cells in temporal cortex respond to images of faces and bodies of different orientations and sizes, the majority are found to be tuned to real life sized and upright orientations (Ashbridge et al. 2000; Perrett et al. 1998). This might explain why naturally colored scenes activated a region of the human brain corresponding to inferior temporal cortex in the macaque (Zeki and Marini 1998) but falsely colored scenes did not. However, because the population of cells we tested was selected on the basis that they responded to at least one image from our screening set (and this set largely consisted of natural images) it is impossible to tell to what extent the cells in IT code for the shapes and colors of naturally occurring objects.

Concluding comments

It is clear from the present study that in temporal cortex, color information arrives simultaneously with the form information defined by luminance cues alone. Latency differences between color and achromatic processing may exist elsewhere in the visual system (perhaps between P and M pathways). Certainly, information about motion direction influences STS cells 30 ms before form information (Oram and Perrett 1996) as might be expected if direction is dependent on the M pathway and form on the P pathway. Timing differences in achromatic and color processing, however, do not appear to affect form processing within the STS and IT. This could be because processing prior to the IT/STS compensates for timing differences between M and P channels (perhaps through number of synapses) (Maunsell et al. 1999) or because temporal cortex form processing relies on exclusively on P channel input from V4 that already carries conjoint form and color sensitivity.

In conclusion, this study suggests a strong role for color in familiar object recognition and provides no evidence to support the idea of a first wave of form processing based on purely achromatic information.

DISCLOSURES

This work was funded by grants from the Biotechnology and Biological Sciences Research Council.

REFERENCES


