Report

Discrimination Strategies of Humans and Rhesus Monkeys for Complex Visual Displays

Kristina J. Nielsen,[1] Nikos K. Logothetis,[1] and Gregor Rainer[1,*]
[1] Max-Planck Institute for Biological Cybernetics
D-72076 Tübingen
Germany

Summary

By learning to discriminate among visual stimuli, human observers can become experts at specific visual tasks. The same is true for Rhesus monkeys, the major animal model of human visual perception. Here, we systematically compare how humans and monkeys solve a simple visual task. We trained humans and monkeys to discriminate between the members of small natural-image sets. We employed the "Bubbles" procedure [1] to determine the stimulus features used by the observers. On average, monkeys used image features drawn from a diagnostic region covering about 7% ± 2% of the images. Humans were able to use image features drawn from a much larger diagnostic region covering on average 51% ± 4% of the images. Similarly for the two species, however, about 2% of the image needed to be visible within the diagnostic region on any individual trial for correct performance. We characterize the low-level image properties of the diagnostic regions and discuss individual differences among the monkeys. Our results reveal that monkeys base their behavior on confined image patches and essentially ignore a large fraction of the visual input, whereas humans are able to gather visual information with greater flexibility from large image regions.

Results and Discussion

We investigated the performance of monkey observers trained to discriminate among natural images. Natural images contain structure at many spatial scales distributed nonhomogenously across the image and are thus good examples of complex, redundant visual forms. The monkeys were trained to discriminate between three natural images by performing a saccade task with these stimuli (see Figure 1). Every stimulus presentation was followed by the presentation of three response targets, each of which was associated with one of the stimuli. Upon presentation of a particular stimulus, a saccade to the associated target was rewarded with a drop of juice.

After the monkeys reached a performance level of at least 80% correct for a particular stimulus set, "Bubbles" was used to identify the diagnostic regions for each stimulus in the set. In Bubbles, stimuli are sampled from a parametric search space. Here, we search the image space by presenting the stimuli behind occluders, which consist of a mid-gray mask punctured by a number of randomly located windows ("bubbles") through which the occluded image was visible (see Experimental Procedures for details). Unique occluders were generated on every trial by randomly placing the bubbles. The monkeys continued to perform their discrimination task on the partially visible images. Whether they could identify the partially visible stimuli depended on whether the occluder uncovered image parts critical for task performance. For quantitative analysis, we grouped the occluders from trials in which a stimulus was correctly identified. We similarly grouped occluders from incorrect trials, and we determined diagnostic regions by comparing these two groups. At each pixel, the distributions of occluder values for correct and incorrect trials were compared with the Kolmogorov-Smirnov test. Image pixels at which occlusion systematically influenced performance should show a different distribution of occluder values in correct and incorrect trials, whereas similar distributions should arise for pixels with no influence of occlusion. The p values of the Kolmogorov-Smirnov test were Bonferoni corrected for the number of image pixels, and diagnostic regions were taken to be image regions where the corrected p values were below 0.01.

Because monkeys had not been tested with Bubbles before, we first established that the technique is suitable for the study of visual recognition in monkeys. For this purpose, we used a custom-designed set of geometrical shapes for which we a priori determined the diagnostic regions. The results of this experiment are reported in the Supplemental Data available online. We then proceeded to use Bubbles to study visual-information use in a task that required the discrimination among the members of natural-image sets. The diagnostic regions for two image sets and both monkeys are shown on the left side of Figure 2. Diagnostic regions covered on average 7% ± 2% of each image and were similar in size for the two monkeys [G00: 4.5%; B98: 9.3%; paired t test: t(5) = −1.34, p = 0.24]. Note that no diagnostic region could be determined for one of the images in monkey B98, suggesting that the monkey used no region consistently to identify the image.

Whereas the size of the diagnostic regions was similar for the two monkeys, their diagnostic regions tended to contain different amounts of spatial structure and cover different image locations. For monkey B98, diagnostic regions were located closer to the image border and contained image regions that largely lacked spatial structure. In contrast, G00 used image regions with more spatial structure, located near the center of the images. Accordingly, diagnostic regions for the two monkeys differed in terms of their distance from the image center [average distance for B98: 2.93° of visual angle; G00: 1.99°; paired t test: t(4) = −5.67, p < 0.01]. Image structure was characterized by luminance and amount of edges. Both parameters were computed at four decreasing levels of spatial scale (see Experimental Procedures). The mean

*Correspondence: gregor.rainer@tuebingen.mpg.de
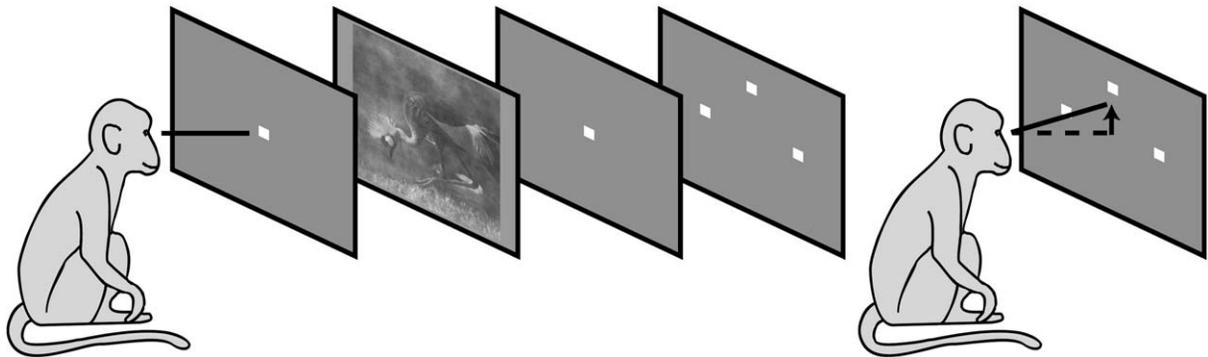
Figure 1. Task Design for the Monkey Observers

Each task began with the presentation of a central fixation spot, which the monkeys had to fixate. While the monkey continued to fixate, the fixation spot was replaced by the stimulus for 300 ms, after which time the fixation spot reappeared. Finally, three targets appeared in the periphery, each of which was associated with one of the stimuli. The monkey had to make a saccade to the correct target to receive a reward.

luminance of the diagnostic regions was similar for both monkeys, independent of spatial scale [paired t test: t(4) ≤ 0.75, p ≥ 0.5 for the four scales]. However, the diagnostic regions of B98 contained significantly fewer edges at the finest resolution [paired t test: scale 1: t(4) = 3.33, p = 0.03; scale 2: t(4) = 2.63, p = 0.06; scale 3: t(4) = 2.05, p = 0.11; scale 4: t(4) = 2.63, p = 0.06], confirming that the diagnostic regions of monkey B98 contained less spatial structure. Thus, diagnostic-region size, but not its location or spatial structure, was consistent across both monkeys. During task performance, we introduced catch trials on which the unoccluded images were shown to ensure that monkeys were maintaining high performance discriminating the unoccluded images. Both monkeys performed equally well on these catch trials [G00: 95% correct, B98: 98% correct, paired t test: t(5) = −1.24, p = 0.27].

Do the diagnostic regions identified by Bubbles bear any relation to performing the task outside the Bubbles paradigm? To address this question, we investigated whether monkeys could correctly identify images when presented with their diagnostic regions alone. For this purpose we constructed "diagnostic" stimuli that consisted of image regions with high diagnosticity by revealing the 10%, 30%, and 50% most diagnostic pixels (see Figure 3). Similarly, we constructed "nondiagnostic" stimuli consisting of the 10%, 30%, and 50% least diagnostic pixels. A unique stimulus set was generated for each monkey on the basis of that monkey's Bubbles results. All six stimuli thus constructed were matched to the original image in terms of mean luminance and contrast. Monkeys performed the discrimination task with the modified stimuli with no additional behavioral training.
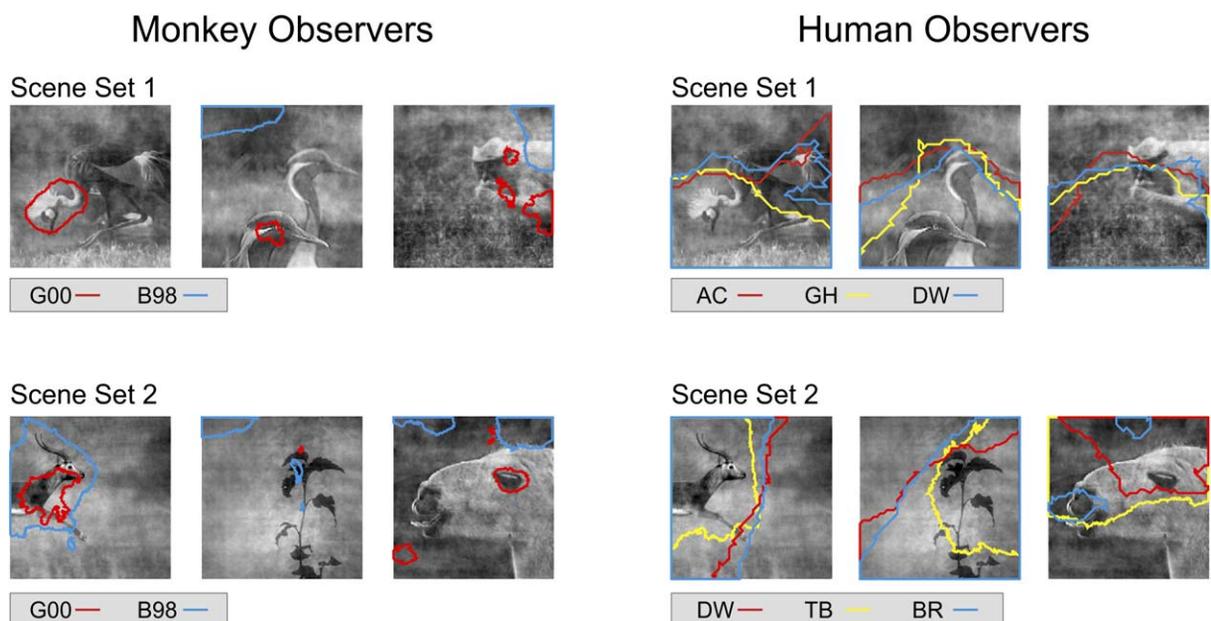
Figure 2. Diagnostic Regions in Natural Scenes for Monkeys and Humans

The left side shows the results for the two monkey observers, the right side for the human observers. Lines encircle the diagnostic regions, with each color corresponding to one observer (observer identity is given in the legend below each plot).
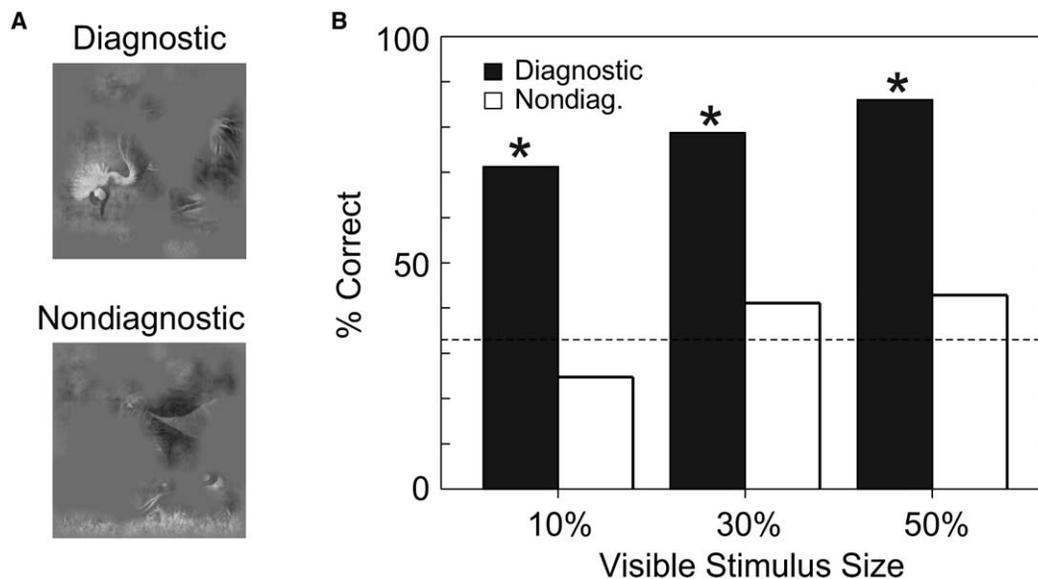
Figure 3. Verification of the "Bubbles" Results

(A) Exemplar stimuli showing a "diagnostic" and a "nondiagnostic" stimulus. These stimuli were generated based on the Bubbles results for the first scene in the first image set for monkey G00. In both stimuli, 50% of the original stimulus are exposed.

(B) Performance of the monkeys with the control stimuli. Data are averaged across both monkeys and stimulus sets. Black bars correspond to the performance for diagnostic stimuli, open bars to the performance for nondiagnostic stimuli. The dashed line shows the chance level (33.3%). Stars indicate deviations in the performance from chance level significant at $p < 0.05$, as assessed by a χ^2 test.

Performance levels were averaged across all six images in the two sets and both monkeys (see Figure 3). All performance levels were compared against the chance level of 33% correct responses with a χ^2 test. Monkeys performed significantly better than chance for all diagnostic stimuli (χ^2 tests: $\chi^2 \geq 37.6$, $p < 10^{-8}$ for the three stimulus sizes). However, monkeys performed at chance level for the three nondiagnostic stimuli ($\chi^2 \leq 2.48$, $p > 0.16$ for the three tests). This indicates that when monkeys were confronted with image regions of high diagnosticity, they treated these as the unoccluded images and were able to perform the task. In the absence of high-diagnosticity regions, monkeys were unable to perform above chance.

To compare the visual information use of monkeys with that of humans, we tested human observers with Bubbles on the identical image sets. The diagnostic regions for human observers are shown on the right side of Figure 2. With an average size of 51% ± 4% of the full image, diagnostic regions for human observers were an order of magnitude larger than the diagnostic regions determined for the monkeys. A t test showed this difference to be significant [$t(28) = -9.44$, $p < 0.001$]. These results are summarized in Figure 4A, which contrasts the diagnostic-region size for the two species.

On most of the trials, the diagnostic regions were not completely exposed, but only a small portion of them was visible. To analyze how much of the diagnostic region was visible on an average correct trial, we focused on the last trials of each testing session. Performance levels were similar for monkeys and humans for this data set [t test on the performance levels, $t(25) = 0.05$, $p = 0.96$]. For each trial, we then computed what fraction of the diagnostic region was visible through the occluder. These data, averaged across correct and incorrect trials separately, are plotted in Figure 4B, showing

that monkeys needed to see more of their diagnostic regions to identify an image than human observers. On average, 42.9% of a diagnostic region needed to be visible for the monkeys to correctly identify a scene, whereas on incorrect trials only 32.9% of the diagnostic region was visible. This difference was statistically significant [$t(8) = 2.67$, $p = 0.03$]. For human observers, only an average of 4.2% of the diagnostic region was visible on correct trials, compared to 2.1% on incorrect trials. This difference was also statistically significant [$t(17) = 8.60$, $p < 0.001$], as were the differences between monkeys and human observers [correct trials: $t(25) = 12.65$, $p < 0.001$; incorrect trials: $t(25) = 11.98$, $p < 0.001$].

The results imply that monkeys needed to see more of the diagnostic image regions for a correct identification. However, the diagnostic regions of monkeys cover a smaller extent of the full image. Considering of each occluder only the bubbles that fall into the diagnostic regions, we found that for the monkeys, on average 2.0% of an image was visible on correct trials. This value was similar to the result for humans, for which 2.2% of an image was visible on an average correct trial [$t(25) = -0.41$, $p = 0.69$]. On the incorrect trials performed by monkeys, only 1.4% of the images was visible, whereas 1.1% of the images was visible on incorrect trials for human observers. Again, the two values were not significantly different [$t(25) = 0.91$, $p = 0.37$]. In conclusion, when only the diagnostic image regions are considered, monkeys and humans required the same amount of the full stimulus to be exposed for a similar performance.

Finally, we examined individual differences among human and monkey observers. We estimated the degree to which the diagnostic regions of different observers overlapped. Averaged across the two image sets, the diagnostic regions of the two monkeys overlapped in 1.2% of the full image, or 17.3% of an average
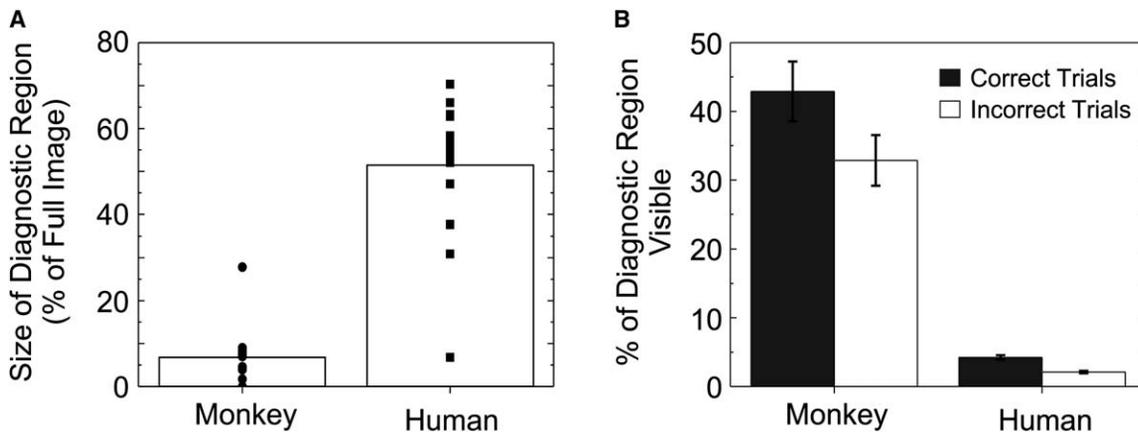
Figure 4. Comparison of the "Bubbles" Results between Monkeys and Humans

(A) Size of the diagnostic regions in percentage of original image size. The bars indicate the mean across all observers and images. Symbols indicate the values for individual diagnostic regions.

(B) Percentage of the diagnostic region visible on an average correct (black bar) or incorrect (open bar) trial. Bars show the average across all observers and images. The error bars correspond to the standard error of the mean.

diagnostic region. In contrast, the diagnostic regions of human observers overlapped on average in 35.6% of the full image, i.e., in 69.2% of the diagnostic regions. Thus, diagnostic regions of individual human observers tended to be more similar to each other than the diagnostic regions of the two monkeys. This raised the question of to what extent the behavior of human observers can be used to predict the monkey observers' behavior. We computed the overlap between the human observers' consensus diagnostic region and a monkey observer's diagnostic region. Across the two scene sets, the common diagnostic region for the human observers overlapped with 77.2% of a diagnostic region of monkey G00. For monkey B98, the data from human observers could be used to predict about 19.3% of the monkey's diagnostic region. This indicates that diagnostic regions estimated in human observers are not in general a good predictor for diagnostic regions in monkeys.

We have trained human and monkey observers to discriminate between natural images. Such images contain a wealth of features that observers might use to discriminate among them. Are observers using all available features in the images equally, or are they preferentially relying on certain features to guide their behavior? To answer this question, we used the Bubbles technique to determine diagnostic regions for each image. These diagnostic regions delineate the spatial location of the features that significantly contributed to observer performance in the discrimination task. Generally, diagnostic regions covered only a fraction of the entire visual stimulus, suggesting that observers were not drawing information equally from the entire image, but sampling preferentially from restricted image portions. Unavailability of diagnostic regions due to occlusion was associated with observers' inability to perform the discrimination task.

We observed robust differences in diagnostic-region size between monkey and human observers. Whereas monkey diagnostic regions covered only a small fraction of the images, they were approximately an order of magnitude larger for human observers, where they covered around half of the images. Diagnostic regions represent the image parts from which observers draw information, but how much visual information do observers need to see on an individual trial to enable them to perform correctly? Intriguingly, taking into account the diagnostic-region size, we found that around 2% of the entire image was visible on the average correct trial in both monkeys and humans. Thus, although the actual amount of visual information required for correct performance was similar for monkeys and humans, humans were able to gather this information from a much larger region. This suggests that human observers could extract task-relevant information from the visual environment with greater flexibility. A recent study applied the Bubbles technique to a face-classification task in human observers and pigeons and found general agreement between diagnostic regions in terms of size and location for those two species [2]. To what degree this is a result of the different visual stimuli used in that and our study remains to be determined.

Although diagnostic-region size was similar for both monkeys, they showed significant individual differences in location and image statistics of the diagnostic regions. Diagnostic regions in one monkey were located close to the image center and contained lots of spatial structure, whereas in the other monkey they were located close to the image border and contained little spatial structure. These differences in the diagnostic regions cannot be explained by different training histories, because the two monkeys received the same training. Our results therefore imply that the monkeys' individual biases led them to choose different strategies. Note that both monkeys performed the discrimination task with unoccluded images at similarly high levels. As our findings suggest, they achieved this performance by using very different strategies and focusing on different image regions. There is no way to infer this rather striking difference in visual information use on the basis of performance data on the discrimination task, and it can be detected only with a method, such as Bubbles, that directly visualizes information use.

We demonstrate that trained observers use particular spatial regions in complex scenes to perform

a discrimination task. Other forms of perceptual learning are based on enhancing sensitivity for orientation, spatial frequency, or other stimulus dimensions [3]. The spatial version of Bubbles employed here does not tell us whether observers are relying specifically on certain spatial frequency or orientation channels; however, in principle Bubbles can readily be adapted for the study of such effects. Indeed, it has already been used to identify how the performance of a task depends on different spatial-frequency channels [1], as well as on their phase [4]. The version of Bubbles we used relies on occlusion to study the contribution of image features to behavioral performance, raising the question how relevant our results are to real-world vision. Occlusion is common in everyday life, and we are generally able to recognize objects despite the fact that they are partially occluded. Thus, Bubbles can be thought of as a parametric unbiased method for simulating the occlusion that occurs in many real-world situations. In addition, several behavioral studies have provided evidence for the idea that chimpanzees [5], as well as macaque monkeys [6, 7], are able to recognize familiar stimuli despite partial occlusion.

Our findings reveal which features of a set of learned visual stimuli observers actually use during the performance of a task. In brain regions such as area TE of the inferior temporal cortex, learning has been associated with long-lasting modifications in neural activity to represent task-relevant attributes of visual stimuli [8]. After training, TE neurons become tuned to features diagnostic for a categorization or discrimination task [9, 10], or to the trained views of three dimensional objects [11]. Most previous studies on the effects of training on the perception of complex stimuli have used stimuli with predefined dimensions (for example, [12–14]). In these studies, stimuli are assigned to different classes according to experimenter-defined parameter ranges. For example, observers learned to sort Greeble stimuli into different classes based on the shapes of their components [15], or to sort face and fish stimuli into categories based on dimensions such as nose height or fin size [16]. Investigators have generally inferred that after learning, observer performance must be based on acquired expertise about which aspects of the stimuli are diagnostic. However, most complex visual stimuli are not parametrically defined according to simple generative models and contain many elements that are at different spatial scales and could be employed by observers. Direct methods such as Bubbles may thus prove particularly useful for understanding how such complex stimuli are encoded in the brain.

Experimental Procedures

Subjects
Two adult male Rhesus monkeys (*Macaca mulatta*) weighing 10 and 13 kg participated in the experiments. Before the experiments, a metal head post and a scleral search coil [17] were implanted under aseptic conditions [18]. Monkeys received their daily amount of liquid during the experimental sessions and were provided with dry food ad libitum. The monkeys were tested daily and performed between 500 and 1000 trials per day. About 20 sessions were collected per monkey for each stimulus set. All studies involving the monkeys were approved by the local authorities (Regierungspräsidium Tübingen) and were in full compliance with the guidelines of the European Community (EUVD, European Union directive 86/609/EEC) and the National Institutes of Health for the care and use of laboratory animals.

A total of eight human observers (3 males, 5 females) were tested. All subjects were naive as to the purpose of the experiments. Informed consent was obtained from all subjects. Subjects had normal or corrected-to-normal vision. Testing sessions usually lasted between 1 and 3 hr, with subjects completing between 1000 and 2000 trials in this time. Subjects returned to the lab for additional sessions, until a total of 3000 to 6000 trials had been collected.

Task and Stimuli
Two stimulus sets of three natural scenes each were used. All stimuli had a size of 256 × 256 pixels, corresponding to 6° × 6° of visual angle. The natural scenes were taken from Corel PhotoCDs and normalized to have equal Fourier amplitude spectra [19]. All stimuli were presented centrally. Both monkey and human observers worked with one stimulus set at a time. During each trial, one of the stimuli was randomly chosen and presented to the observer. Observers had to indicate which of the three stimuli they had just seen.

For monkeys, each trial began with the presentation of a yellow fixation spot in the center of the screen, combined with the sounding of a tone. After 100 ms fixation time, the spot was turned off and the stimulus was presented for 300 ms. During stimulus presentation, the monkeys had to maintain fixation at the center of the screen in a window with a radius of 3°. After another 100 ms of central fixation, three small white squares (the targets) were presented at 6° eccentricity. Each of the three members of a stimulus set was associated with one of the targets. A saccade to the correct target was rewarded by a drop of juice.

For human observers, trials began with the presentation of a yellow fixation spot for 500 ms, followed by one of the stimuli for 500 ms. Observers responded after the presentation of the stimulus by pressing designated keys on the numerical keypad of a standard computer keyboard. Each of the images in a stimulus set was associated with a specific response key. No constraints were imposed on reaction time. For the data described above, no fixation constraints were imposed because of the brief presentation time, and observers were not given feedback about the correctness of their answer to prevent learning and behavioral nonstationarity during "Bubbles." To ensure that these factors did not significantly contribute to the differences between humans and monkeys, we performed control experiments during which three human observers were required to maintain fixation within 3° of the center of the screen. The control experiments were performed with the second natural-scene set. The tested subjects did not participate in the previous experiments. During these control experiments, stimuli were presented for 300 ms, and we provided human observers with performance feedback such that after each trial a "+" or "−" sign on the screen indicated correctness of the response. These conditions thus recreated the exact parameters we used for the monkeys. We found that the characteristics of the diagnostic regions were not changed by these additional controls. In particular, the diagnostic-region size was 33.5% ± 5% and thus statistically indistinguishable from that obtained in the original experiments [46.1% ± 7%, t test, t(16) = −1.54, p = 0.14]. There was a 73% overlap between diagnostic regions determined in these control experiments and the regions obtained in the original experiment.

All observers were initially trained to associate each of the unoccluded stimuli with its assigned saccade target or button press. For monkeys, this was done by introducing a brightness cue in the saccade targets, with the correct target being brighter than the incorrect targets. This cue was gradually removed as the monkeys' performance improved. Monkeys were always trained with the entire stimulus set. Monkeys initially learned to associate visual stimuli with saccade directions prior to the stimulus sets reported here. This initial training lasted for a period of several months. Once the monkeys had learned the rules of the task, they quickly acquired new stimulus sets. Humans were provided with a printout that showed both the stimuli and their associated buttons in order to inform them about the mapping between stimuli and response buttons. They were then given a training period of 20 trials, in which they could use the printout to guide their responses. After these training trials, the printout was removed. All subjects performed

better than 90% correct on the original stimuli after these training trials.

After observers had acquired the task with unoccluded stimuli, we additionally introduced stimuli with occluders. The presentation of unoccluded images was maintained (10% of trials for human observers, 40% for monkeys) as a baseline control of the performance. The occluders were constructed as described in [1]. In brief, each occluded image appeared to be shown behind a surface punctured by round windows ("bubbles"), through which parts of the image were visible. Bubbles had the profile of a 2D Gaussian, so that they smoothly merged into the nontransparent background. Bubbles were randomly positioned, with the restriction that the center of each bubble fell onto an image pixel, and two bubbles could not have identical center coordinates. All bubbles had the same size, which was determined by setting the standard deviation of the 2D Gaussian profile to 14 pixels. For the human subjects, bubbles numbers were adapted to each subject's performance by a staircase protocol. Staircases were run independently for each image in a stimulus set and converged to a performance of 75% correct. After every fourth trial of an image, the bubbles number was updated. The number was decreased by three if the image had been identified correctly in the last four trials, and it was increased by two if fewer than three trials had been correct. For the monkeys, we employed the same staircase procedure in most sessions. As an additional control, we used a modified staircase procedure for one dataset (second set of natural images for monkey B98). During these sessions, the bubbles numbers were identical for all images, rather than adapted to each stimulus independently as in the original staircase procedure. The modified procedure thus showed all stimuli through the same bubbles number, preventing the number of bubbles itself from serving as a potential cue to the stimulus. They were initialized to a value at which the monkeys could perform the task at ceiling performance. After 15 trials, the bubbles numbers were successively decreased by a fixed amount until the monkey's performance dropped below 70% correct. At this point, the numbers of bubbles were reset to the initial value and the cycle was restarted.

Setup

Monkeys performed experiments in acoustically shielded chambers. Eye movements were monitored with the scleral-search-coil technique [20] and digitized at 200 Hz. Stimuli were presented on a 21" monitor (Intergraph 21sd115, Intergraph Systems, Huntsville) with a resolution of 1024 by 768 pixels and a refresh rate of 75 Hz. Background luminance of the monitor was set to 41 cd/m^2, and the monitor was γ corrected. The monitor was placed at a distance of 95 cm from the monkey. Stimuli were generated in an OpenGL-based stimulation program under Windows NT. Similar equipment was used for human observers, who were seated 85 cm from the monitor (background luminance of 27 cd/m^2). When eye movements of human observers were measured, the head position of the observers was restrained by using a chinrest. Eye movements were measured with iView 1.1 (SensoMotoric Instruments GmbH, Teltow, Germany).

Data Analysis

Analyses were carried out in Matlab (The Mathworks, Natick). To determine how much of the diagnostic regions was visible through the occluder on any trial, we analyzed the last 40 trials of each staircase session for each stimulus. Bubbles numbers were stable throughout these trials. Monkey B98 was tested on the second set of natural scenes with a method of constant stimuli; therefore, this data set was excluded from the analysis. Because only four to six sessions were run for human subjects, we used only the last four testing sessions for the monkeys. An image pixel was considered to be visible when the occluder value for this pixel was equal to or larger than 0.5.

Physical properties of an image were characterized as the distribution of luminance, as well as edges across the image [21]. Both parameters were computed at four spatial resolutions, which were generated through progressively low-pass filtering and subsampling the image. The four resolutions corresponded to horizontal and vertical image-reduction factors of 1, 0.5, 0.25, and 0.125. Luminance information was computed at each resolution by convolution of the image with a 2D Gaussian with a kernel size of 20 by 20 pixels and a standard deviation of 4 pixels. Edges of four different

orientations (0°, 45°, 90°, 135°) were detected at each resolution. They were extracted by applying quadrature filter pairs to the images, i.e., pairs of similarly oriented sine and cosine Gabor filters. The standard deviation of the filters was set to 4 pixels, and the frequency to 1/10 pixels. Artifacts at the image borders were avoided by appending copies of an image to its borders. These copies were only present while convolutions were computed. All computed luminance and edge maps were rescaled to half the size of the original image, i.e., to 128 by 128 pixels.

Supplemental Data

Supplemental Data include Supplemental Results, one figure, and one table and are available with this article online at: http://www.current-biology.com/cgi/content/full/16/8/814/DC1/.

References

1. Gosselin, F., and Schyns, P.G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. Vision Res. *41*, 2261–2271.
2. Gibson, B.M., Wasserman, E.A., Gosselin, F., and Schyns, P.G. (2005). Applying bubbles to localize features that control pigeons' visual discrimination behavior. J. Exp. Psychol. Anim. B. *31*, 376–382.
3. Fahle, M. (2005). Perceptual learning: specificity versus generalization. Curr. Opin. Neurobiol. *15*, 154–160.
4. McCotter, M., Gosselin, F., Sowden, P., and Schyns, P.G. (2005). The use of visual information in natural scenes. Vis. Cogn. *12*, 938–953.
5. Fujita, K. (2001). What you see is different from what I see: Species differences in visual perception. In Primate Origins of Human Cognition and Behavior, T. Matsuzawa, ed. (Tokyo: Springer), pp. 29–54.
6. Kovács, G., Vogels, R., and Orban, G.A. (1995). Selectivity of macaque inferior temporal neurons for partially occluded shapes. J. Neurosci. *15*, 1984–1997.
7. Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. Nature *401*, 269–272.
8. Logothetis, N.K., and Sheinberg, D.L. (1996). Visual object recognition. Annu. Rev. Neurosci. *19*, 577–621.
9. Sigala, N., and Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. Nature *415*, 318–320.
10. Baker, C.I., Behrmann, M., and Olson, C.R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat. Neurosci. *5*, 1210–1216.
11. Logothetis, N.K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Curr. Biol. *5*, 552–563.
12. Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. J. Exp. Psychol. Gen. *115*, 39–57.
13. Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychol. Rev. *94*, 115–147.
14. Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science *291*, 312–316.
15. Gauthier, I., Williams, P., Tarr, M.J., and Tanaka, J. (1998). Training 'Greeble' experts: A framework for studying expert object recognition processes. Vision Res. *38*, 2401–2428.

16. Sigala, N., Gabbiani, F., and Logothetis, N.K. (2002). Visual categorization and object representation in monkeys and humans. J. Cogn. Neurosci. *14*, 187–198.

17. Judge, S.J., Richmond, B.J., and Chu, F.C. (1980). Implantation of magnetic search coils for measurement of eye position: An improved method. Vision Res. *20*, 535–538.

18. Lee, H., Simpson, G.V., Logothetis, N.K., and Rainer, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. Neuron *45*, 147–156.

19. Rainer, G., Augath, M., Trinath, T., and Logothetis, N.K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. Curr. Biol. *11*, 846–854.

20. Robinson, D.A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. IEEE Trans. Biomed. Eng. *BME-10*, 137–145.

21. Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. *20*, 1254–1259.