Enhanced visual exploration for real objects compared to pictures during free viewing in the macaque monkey

Faiz Mustafar [a,b,1], Paolo De Luna [a,b], Gregor Rainer [a,b,*]

[a] *Visual Cognition Laboratory, Department of Medicine, University of Fribourg, Chemin du Musée 5, Fribourg CH-1700, Switzerland*
[b] *Swiss Non-Human Primate Competence Center for Research, Switzerland*

The question of whether animals perceive pictures as representation of real objects remains still unsolved. Object-picture perception is generally studied requiring animals to learn some information about real objects and transfer that knowledge to the pictorial domain, or vice versa. Here, we tackle the issue of object-picture perception from a different perspective, examining visual exploration behavior of two naïve macaque monkeys during free-viewing of objects and pictures of these objects on a computer monitor. Our main finding is that monkeys looked spontaneously longer at object rather than picture stimuli. However, we find striking similarities in temporal dynamics of gaze allocation within the time course of a single stimulus presentation, as well as in habituation rates within and across behavioral sessions. We also highlight differences between stimulus types in terms of spatial gaze patterns and looking strategies. Stimulus features that attract overt attention during spontaneous visual exploration are thus better predicted for object stimuli by a visual saliency model. Moreover, we provide evidence for a consistency in stimulus preference for objects and pictures, suggesting a correspondence of in how macaques perceive objects and their pictorial stimuli. Taken together, our data suggest that macaque monkeys exhibit evidence for correspondence between objects and pictures. This validates spontaneous visual exploration as a method for studying object-picture correspondence without a need for extensive behavioral training. We discuss the potential advantages of using object over picture stimuli in the context of studies on visual cognition.

1. Introduction

The degree to which animals can recognize pictorial representations of objects is a subject of current interest (Bovet and Vauclair, 2000; Fagot et al., 2010). A widely used approach to study this question relies on examining transfer effects in learning paradigms, where animals are trained on a task involving objects and then tested with the corresponding pictures or vice versa. A considerable literature has examined object-picture perception in birds (Fagot and Parron, 2010) including chicken (Railton et al., 2014) and pigeon (Cabe, 1976; Spetch and Friedman, 2006; Watanabe, 1997) as well as other species such as tortoise (Wilkinson et al., 2013). These studies have revealed different degrees of transfer, usually partial transfer, from objects to pictures, in dependence on task and stimulus variables.

Object-picture perception has also been studied in different primate species. For example, capuchin monkeys trained in an object matching task are able to transfer this knowledge to correctly match objects to pictures (Truppa et al., 2009). By showing that capuchin monkeys are able to match objects to objects when a matching picture is used as a distractor, these authors also provide evidence that transfer is not simply due to confusion between objects and pictures. The results by Truppa and colleagues reveal a large degree of transfer between objects and pictures, and thus suggest a close correspondence between objects and pictorial representations. Other studies have shown that transfer can often be more limited. Thus, macaque monkeys require considerable training to learn that selecting the picture of a preferred food item results in the delivery of that food as a reward (Judge et al., 2012). This study has highlighted large inter-individual differences between participant macaques, with the fastest learner requiring only about 15% of the training needed by the slowest learner to achieve criterion behavioral performance. Similarly, another study demonstrated

* Corresponding author at: Visual Cognition Laboratory, Department of Medicine, University of Fribourg, Chemin du Musée 5, Fribourg CH-1700, Switzerland. Tel.: +41 26 300 8689; fax: +41 26 300 9734.

E-mail address: gregor.rainer@unifr.ch (G. Rainer).

[1] These authors contributed equally to this work.

partial transfer between objects and pictures in marmoset monkeys, in that animals pre-exposed to a real snake model showed enhanced fearful responses to a picture of the same snake (Emile and Barros, 2009). In related work, it was shown that baboons and gorillas, but not chimpanzees that have been trained to select real bananas over stone pebbles tend to choose bananas over pebbles when confronted with their pictures (Parron et al., 2008), thus, demonstrating transfer from the object to the picture domain. In fact, the animals frequently ate the banana pictures, which the authors interpreted as evidence for confusion between objects and pictures. The observation that animals reliably chose the real banana over the banana picture however suggests that the transfer of task-related behavior (that is, eating) from the object to the picture situation might be an alternative explanation. Curiously, object-picture perception has been only rarely studied in macaques, for example (Zimmerman and Hochberg, 1970), despite the fact that this species is used so extensively in visual neuroscience investigations.

The existing literature on object-picture perception can be broadly classified into two cases: some studies have first trained animals on tasks involving objects, such as the stimulus matching paradigm, and then tested their performance with pictures of the same objects to estimate transfer (Truppa et al., 2009; Wilkinson et al., 2013). Other studies have relied on spontaneous behaviors elicited by certain objects, such as approach behaviors for preferred fruits or fearful behavior for snakes, and tested how these behaviors transfer to picture stimuli (Emile and Barros, 2009; Judge et al., 2012; Parron et al., 2008). In both cases, animals are required to transfer acquired or innate knowledge about objects to pictures. Such transfer is likely to depend heavily on cognitive abilities and flexibility, as animals with extensive previous experience on task-related behaviors appear to perform better on transfer to pictures than naïve animals. An alternative to studying the transfer between objects and pictures would be to compare the spontaneous behaviors elicited by both stimulus types when not associated with emotional content or foraging behavior. Similarities in such spontaneous directed behaviors for objects and pictures provide evidence for correspondence in their mental representations, in analogy to high transfer performance from objects to pictures on the transfer tasks. Visual explorative behavior is an example of this kind of spontaneous behavior directed toward elements of the visual environment.

Visual exploration has been studied extensively, and can provide a wealth of information about which aspects of visual scenes attract overt attention, and what kinds of information subjects are extracting from the visual environment (Gottlieb et al., 2013; Gunhold et al., 2014; Schutz et al., 2011). In visual exploration studies, subjects are generally not rewarded for fixating on particular visual scene elements, to avoid biasing the gaze patterns toward these rewarded locations. A part of the visual exploration literature focuses on gaze patterns during free-viewing of ethologically or socially relevant stimuli, such as faces or body parts of conspecifics or other species (Dal Monte et al., 2014; Ghazanfar et al., 2006; Gothard et al., 2004; Kano and Tomonaga, 2009, 2010; Sigala et al., 2011). In this context, gaze patterns are thought to be driven by the need to make predictions about the intentions of other agents (Cannon et al., 2012; Kano and Call, 2014; Southgate et al., 2007), which might be relevant for guiding future behaviors. This can be considered as a form of top-down gaze control, because it is primarily the information requirements of the observer that determines which scene elements are fixated. By contrast, a largely separate body of work has examined bottom-up contributions to gaze control, by attempting to predict gaze patterns based on local statistical properties of image patches in natural scenes (Berg et al., 2009; Berger et al., 2012; Einhauser et al., 2006). In this context, the gaze patterns are determined exclusively on the properties of the visual

input, without considering the information requirements of the observer. An emergent concept in this research area is the saliency map, which is computed by combining various low-level statistical image properties and can be compared to the experimental eye fixation map to determine how well it can predict the allocation of overt attention for a particular visual stimulus (Hou et al., 2012; Itti and Koch, 2000; Koehler et al., 2014; Tatler et al., 2011).

A pertinent property of primate visual explorative behavior is that it tends to habituate with repeated exposure to stimuli, such that less time is spent on successive viewing periods (Butler and Alexander, 1955; Humphrey, 1972, 1974; Jutras and Buffalo, 2010). Habituation has been observed to occur at different timescales of separation between observation periods, ranging from minutes to 24-h intervals (Gammon et al., 1972; Rabedeau and Miles, 1959; Wilson and Goldman-Rakic, 1994). Habituation is generally stimulus specific, and thus not a consequence of general fatigue resulting from repeated exposure. It reflects reduced allocation of overt attention, because there is a reduced need to gather information about stimuli that have been previously explored. Habituation is thought to depend on implicit memory formation, and can thus be used to study recognition memory, as is typically done in novelty preference paradigms (Antunes and Biala, 2012; Hannula et al., 2010; Snyder et al., 2008).

Thus, previous studies on object-picture perception in non-human primates have generally employed manual interactions with the stimuli as well as substantial behavioral training. Here, we address the issue of object-picture perception using gaze fixation instead of manual interaction as a behavioral measurement and rely on the inherent motivation of the animals for visual exploration rather than a trained behavior that results from operant conditioning. This approach can complement existing literature and provide important evidence for perceptual correspondence between objects and pictures in non-human primates. Specifically, we have compared the spontaneous visual exploratory behavior of task-naïve monkeys for sets of arbitrary three dimensional shapes and the corresponding pictures presented on a computer monitor. The use of arbitrary shapes ensured that monkeys had no relevant semantic knowledge that might bias their behavior during free-viewing, thus, emphasizing bottom-up saliency as the main factor determining gaze control.

2. Material and methods

2.1. Subjects

Experimental subjects were two male macaque monkeys (*Macaca fascicularis*), aged of 12 years and referred to as monkey C (7.5 kg) and monkey D (8.4 kg). The monkeys lived within the same social group, in an enriched environment with indoor residence and outdoor compound. They were maintained on a diet of fresh fruit, vegetables, and monkey chow with water available ad libitum. The monkeys were familiar with the behavioral setup but they had not been trained on any behavioral task, except brief fixation on flashed light spots during eye tracker calibration (De Luna et al., 2014). They had never encountered the stimuli in the laboratory or in their home-cage. To perform good quality gaze tracking, monkey's head movements were non-invasively restrained using a helmet made of thermoplastic material (Uni-frame Thermoplastics MTAPUI2232, Civco Medical Solutions), that was manually shaped to achieve a comfortable fit for each animal (De Luna et al., 2014; Machado and Nelson, 2011). All experimental procedures and protocols were in fully compliance with Swiss and European Union animal experimental regulations and approved by the Fribourg cantonal veterinary authorities.
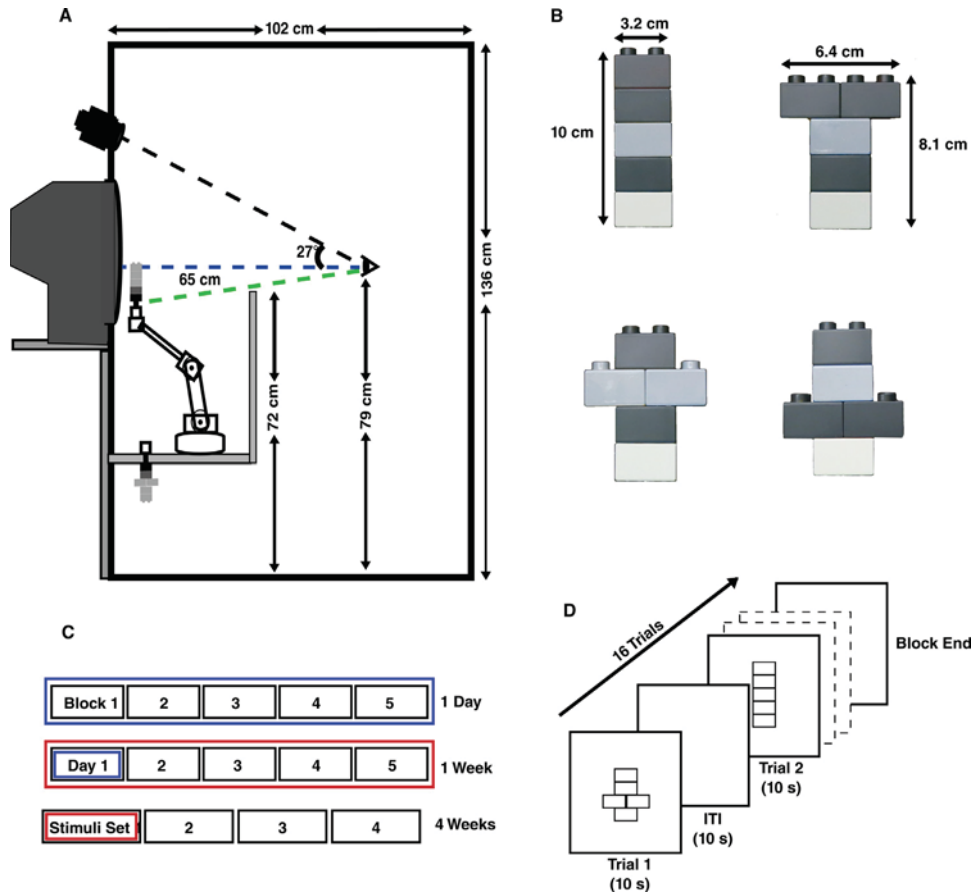
Fig. 1. Behavioral experimental setup. (a) Side view of the behavioral setup. The monkey's central visual field is shown by the blue dash line and the dashed green line indicates the limit of the monkey's lower visual field. (b) Example of four shapes employed as visual stimuli. (c) The overall schedule of the behavioral recording sessions during the entire four weeks period. (d) Schematic illustration of the experimental procedure for a block of trials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. Apparatus

2.2.1. Setup

The monkey was seated in a custom-made primate chair, whose position could be fixed inside a wooden recording box, as shown in Fig. 1a. The inside of the box was covered with black matte plastic panels. Illumination was provided by nine white LED lights, placed on the sidewalls, and a cathode-ray tube monitor (Compaq P900 19-inch CRT, at resolution of 1024×768 pixels, with the refresh rate of 75 Hz), placed at 65 cm from the subject. Since this monitor was used to present visual stimuli, the gamma color was calibrated via sampling luminance levels with a Minolta TVCA-II color analyzer and passing these values to the software used to show the stimuli. During the whole procedure, the screen showed a uniform background of intermediate intensity (50% grayscale), corresponding to a luminance of 13 cd/m^2. Auditory white noise was delivered via a speaker to mask external distractions.

Inside the recording box, a robotic arm (model AL5D-KT with the medium duty wrist upgrade, RobotShop Inc., acquired by Lynx-motion Inc., Swanton VT, US) was placed on a shelf below the monitor (see Fig. 1a). Robot movements were controlled by five independent servo-motors that moved three main joints. The robot input was sent through a serial connection and movements were programmed to grasp an object that was hanging from the shelf and present it at very close distance from the monitor (about 3 cm, which is about 62 cm from the animal). A black plastic panel occluded the lower subject's view up to the bottom edge of the screen, such that the monkeys never saw the robot moving or the object stimuli before they were presented (Fig. 1a, green dashed line). From the monkey's point of view, the stimulus was situated approximately at the center of the screen, in the center of his visual field (Fig. 1a, blue dashed line).

2.2.2. Visual stimuli: objects and pictures

Experimental stimuli were LEGO^{TM} objects and photographs of these objects. To avoid any bias due to object's semantic content and monkeys' color preference, we employed simple, abstract, and symmetric shapes and three neutral colors such as silver, light, and dark gray.

Objects were made of five LEGO^{TM} blocks each. Four basic shapes were chosen: one where blocks were arranged in a vertical straight line (sized 3.2×10.0 cm, width $\times$ height); the remaining three shapes had three blocks arranged vertically and interleaved by two blocks of the same color attached side-by-side (6.4×8.1 cm) (Fig. 1b). The bottom block of each object was attached to a plastic holder (0.8×7.5 cm) (not shown in Fig. 1b) that allowed the robotic arm to grasp and present each object without interfering with the monkey's view of the stimulus. Colors' relative position was kept fixed across the four shapes. Thus, for instance, one possible arrangement was from bottom to top, dark gray–light gray–silver–light gray for one shape, thus, the same order was also used for the remaining three shapes (see Fig. 1b). A particular arrangement of colors defined a set of visual stimuli that was presented repeatedly during one week of experimental recordings (see Section 2.4 and Fig. 1c). A different color pattern was introduced on the first day of the week.

3

Picture stimuli were photograph of the objects described above, shot with a 12 mega-pixel digital camera at a distance of about 60 cm. These pictures were prepared using free photo-editing software (Paint.NET, v4.0.3) to cut unnecessary parts such as the background as well as the plastic holder. To avoid any possible bias due to stimulus position and size, picture stimuli were adjusted to match the corresponding objects. Color saturation and luminance of picture stimuli were adjusted by eye by two researchers using the photo-editing software, comparing side-by-side an object and the corresponding picture under the same illumination that would have been provided to the monkeys.

Object and picture stimuli presentation was automated in a custom-written MATLAB program (version R2010a, The Mathworks Inc.). Low-level functions for stimuli presentation and exact time-stamping of stimuli animation were managed via functions of the Psychophysics Toolbox for MATLAB (v 3.0.11; (Brainard, 1997; Kleiner et al., 2007)).

2.3. Data collection

2.3.1. Eye tracking

Monkey's gaze was tracked employing a fully non-invasive setup, consisting of a helmet-based head restraint and an infra-red sensitive camera. The eye tracking acquisition hardware and software consisted of an infra-red sensitive camera (ISCAN ETL-200 system, ×2 lens attached, 120 Hz frame-rate) and a desktop computer running the acquisition software (DQW, v. 1.20N). The eye-tracking camera was placed in the recording box in the center of the wall on the opposite side of the monkey, at an elevation angle of +27° at a distance of 65 cm from the animal. Gaze position was measured tracking only the right eye for both monkeys.

Collected eye data were horizontal and vertical gaze position estimated using the relative positions of pupil center and corneal reflection. Raw eye data samples and their timestamp were collected at an average sample-rate of about 330 Hz. All the hardware components, except the robotic arm, and our MAT-LAB script have been described in an earlier publication from our group (De Luna et al., 2014). The software is publicly available on our website, at http://www.unifr.ch/inph/vclab/home/internal/eye-tracking-software

2.4. Procedure

Presentations of visual stimuli were organized in daily sessions of five blocks of sixteen trials each. Five daily sessions were administered every week according to a typical Monday-to-Friday schedule, at the same time of the day for both monkeys. Experiments were performed for a total of four weeks (Fig. 1c). Each recording session began with the eye tracking calibration performed on nine positions, eight of which arranged along a circle of 6° radius, regularly spaced 45° around the ninth in the center. We employed black squares of 0.5° diameter as fixation targets which were randomly presented. After the monkey had fixated the target for 200 ms within a 2° fixation window, the point disappeared from the screen and reward was delivered together with auditory feedback. About 20–25 trials were required to adjust gains and offsets of the signals for horizontal and vertical dimension.

During each behavioral session, four objects and their pictures were used for a total of eight visual stimuli. Visual stimuli were randomly presented twice per block (for a total of 16 trials per block), one at a time for about 10 s followed by 10 s inter-trial interval (Fig. 1d). Each trial started with the stimulus outside the subject's view and it required about 1.5 s to reach the center of the screen. Once there, it remained in a stationary position for 7 s and then it moved away along the same path of when it entered, which required approximately 1.5 s to totally disappear from

subject's view. No food reward was provided during the behavioral recordings. However, during the 2 min inter-block interval, fifteen 45 mg pellets were given to the monkeys regardless of their gaze behavior.

2.5. Data analysis

2.5.1. Fixations identification

In our experiments, eye position tracking was conducted with a fully non-invasive setup which avoids unnecessary psychophysical stress on monkeys. However, collected eye data are generally noisier compared to those collected with invasive procedures, such as head-post and scleral search coil implants. In order to identify periods of gaze fixations in the presence of noise, we have designed a dispersion-threshold algorithm that relies on changes in the signal variance. This algorithm is similar to a previous implementation (Salvucci and Goldberg, 2000). Our algorithm involves the following elements:

1) Eye blinks removal. Blinking periods were removed from the signal by discarding samples that could not be associated to physiological eye positions (velocity higher than 1000°/s, or absolute eye position more eccentric than 100°).
2) Signal variance estimation. Eye position variance was computed independently for the horizontal and vertical eye signals in a 61-sample sliding temporal window (corresponding to about 180 ms). Next, for each trial, we set a threshold at 1% of the maximum variance recorded during that trial for both horizontal and vertical dimensions. This allowed identifying the saccadic eye movements that corresponded to peaks in signal variance. For subsequent analysis, the two dimensions of variance, x and y, were combined in a unique signal z, where each sample $z(t)$ was the maximum between the values $x(t)$ and $y(t)$ at each time point t.
3) Saccades identification. Saccades were identified via template matching with the signal z. The saccade template was a parabola described by the equation $y = -x^2$, evaluated in the range −30 to +30 ms such that the center of the curve was at 0 ms, corresponding to the center of the template. The magnitude of template matching was defined as the inverse of the squared mean difference between the signal variance and the template moving in a sliding window. After this operation, peaks corresponded to the periods of optimal matching with the saccade template, and troughs to fixations. Peaks were identified using the MAT-LAB's function findpeaks, part of the Signal Processing Toolbox (v 6.18), that controlled for the minimum peak height, that is the threshold, as well as the minimum distance between peaks. This inter-saccade delay was also the minimum fixation duration and was set to 120 ms.
4) Fixations identification. Finally, fixations were defined according to the following criteria: they were never interrupted by eye blinking, they had to follow and precede a saccade, and their minimum duration was 120 ms. If all these criteria were met within any interval, mean x and y eye positions within that interval were stored as a correct fixation. Each two dimensional gaze position was then compared with the region-of-interest of the stimulus, and only if a fixation was inside such window it was used for the subsequent analyses.

For the two main parameters of the algorithm, namely variance threshold and minimum fixation duration, several values were tested. The choice of 1% of maximum within-trial variance and 120 ms minimum fixation duration were estimated as optimal for our data after visual inspection by two researchers on a random subset of 200 trials. The MATLAB code for this algorithm, which worked exceedingly well for our data, can be downloaded from
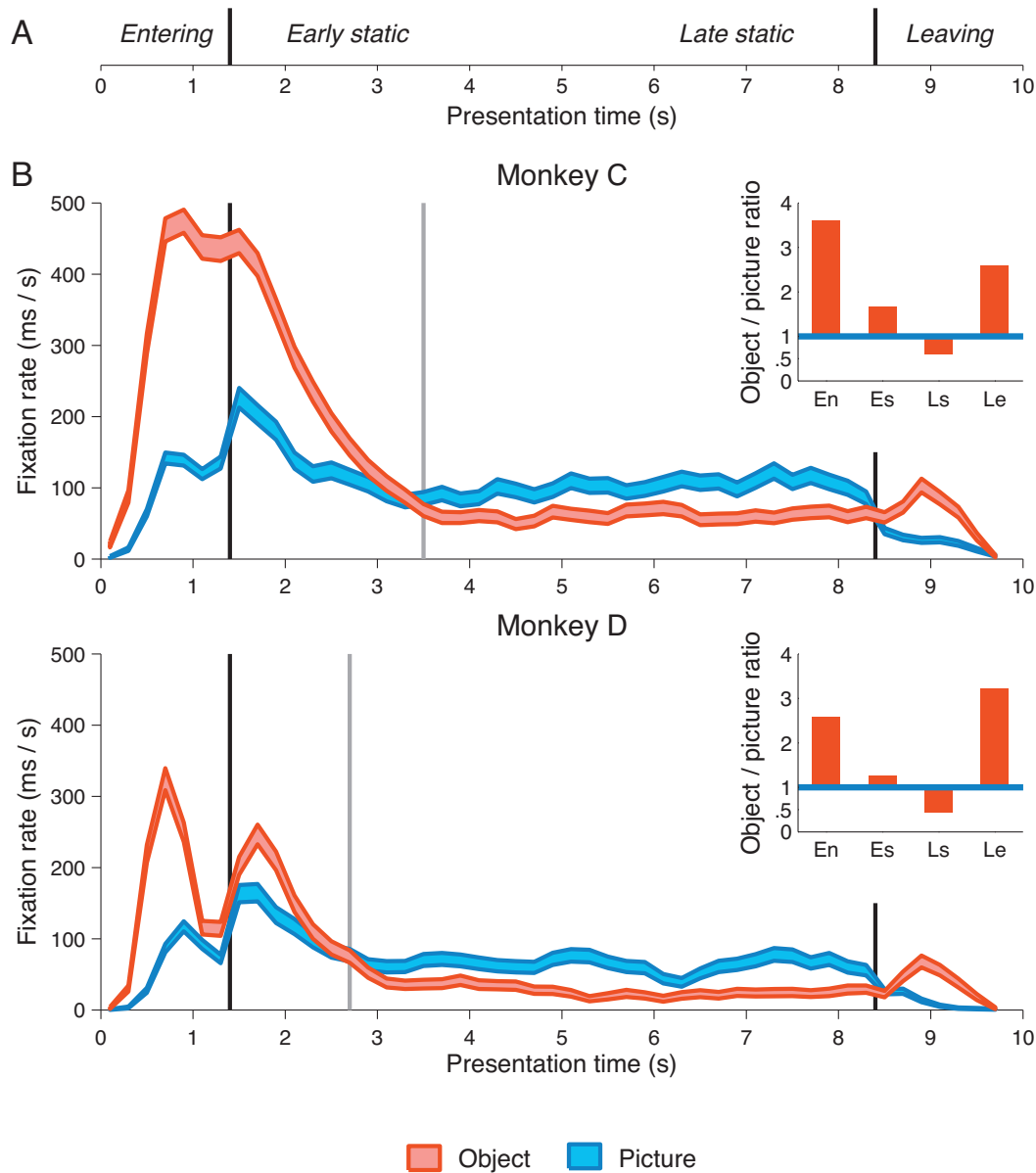
Fig. 2. Temporal dynamics of fixation rate within a trial. (a) Timeline of a single trial with three main epochs which we refer to as entering, static, and leaving epochs. The static epoch was further divided in two periods, early and late static epochs, with the values of 3.5 and 2.7 s respectively for monkey C and D. (b) Fixation rates (measured in ms/s) for objects and pictures averaged across all trials during the 4 trial epochs. A vertical gray line in each graph marks the separation between early and late static epochs. Ribbon graph is mean fixation rate ± standard error of mean, binned every 200 ms. For each graph, an inset shows the total fixation rate exhibited for object (red bars) compared to picture stimuli (blue horizontal line) during the four epochs. Abbreviations are: En, entering; Es, early static; Ls, late static; Le, leaving. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

our website (http://www.unifr.ch/inph/vclab/home/internal/eye-tracking-software).

2.5.2. Regions of interest

Regions of interest (ROI) were computed using the pictorial version of each stimulus. Each picture was processed in MATLAB in order to extract the outline of the stimulus shape. Then, each ROI was defined as the portion of visual scene that included the stimulus shape as well as the space up to 2° around it. The position of the ROI during entering and leaving epochs was estimated on a sample-by-sample basis considering the beginning and end of each stimulus animation and that stimuli moved at constant speed.

2.5.3. Statistics

Temporal dynamics of gaze patterns were systematically analyzed during the different epochs of the trial time course (Fig. 2a).

The entering and leaving epochs were defined as the interval during which the stimulus moved, respectively, toward or away from the subject's central visual field. The period during which the object remained stationary, was divided in early and late static for the two monkeys independently, according to the time when the grand-average fixation rate (shown in Fig. 2b) for picture became greater than for object. Epochs nomenclature used in this paper is shown in Fig. 2a. Statistical hypothesis testing was performed by analysis-of-variance (ANOVA) and t-tests, for which we reported F and t values, respectively, along with their p-value and degrees of freedom (df). Statistical significance threshold criterion was set at $p < 0.05$. For significant ANOVA effects and t-tests, we also reported, respectively, η^2 (eta-squared) and Cohen's d values as estimates of effect size. ANOVA analyses were followed by post-hoc Tukey's honest significant difference (HSD) test, which adjusted the p-values accounting for multiple comparisons.
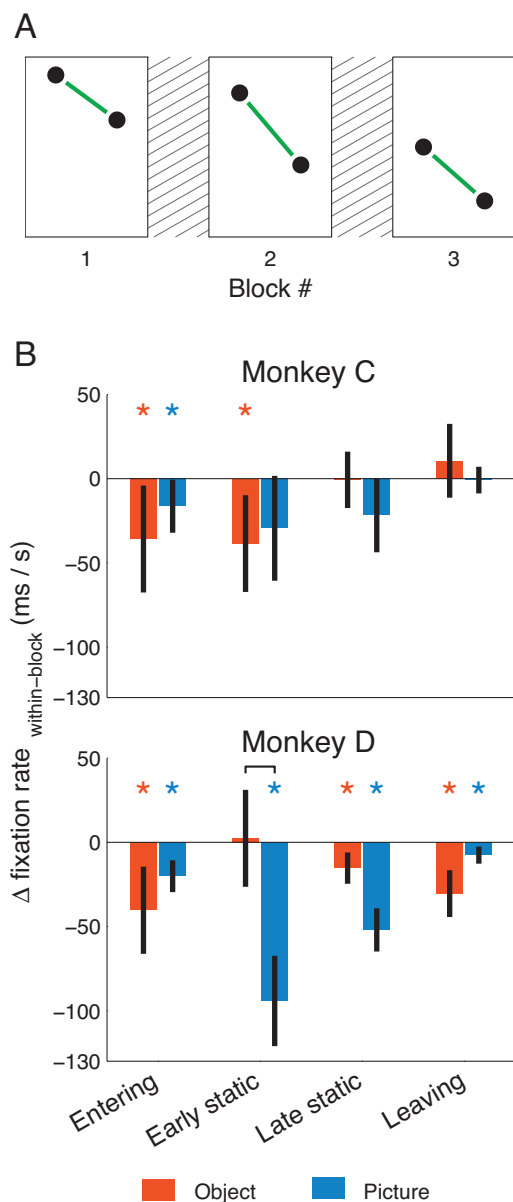
Fig. 3. Habituation within a block. (a) Schematic representation of data used for within-block habituation calculation. Black circles within a rectangle are fixation rates during first and second presentations for a stimulus during a block. Difference between values is indicated by a green line. (b) The two graphs show within-block habituation for monkey C and D, respectively. Bars are mean and 95% confidence intervals of the mean as estimated from a normal distribution. Stars denote p-values that are lower than 0.05. Horizontal square bracket in the bottom graph indicates a significant difference in the post-hoc test between object and picture conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Across-block habituation. (a) Diagram is similar to Fig. 3A. It shows that only the first (black circles), but not the second (gray circles), presentations of a stimulus during a day were used to compute across-block habituation. The green line indicated the slope of the regression line. (b) The two graphs represent the across-block habituation for object (red bars) and picture stimuli (blue bars) during the four epochs for the two monkeys separately. Bars are mean and 95% confidence intervals of the mean estimate. Stars denote rejection of the null-hypothesis of no habituation, which is slope equal to 0 (one-sample t-test, $p < 0.05$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.5.4. Within-trial dynamics of fixation rate

Temporal dynamics of gaze behavior occurred during the course of a trial were estimated using on-stimulus fixation samples. Mean fixation time was estimated in 200 ms wide bins, from time $t = 0$ to $t = 10$ s, corresponding to start and end of each trial. Fixation duration in each bin was then normalized to 1 s time unit, thus, obtaining the fixation rate, in ms/s. A two-way ANOVA with object/picture and epoch as main factors was performed for each monkey independently, on the within-epoch mean fixation rates. ANOVA was followed by post-hoc HSD test.
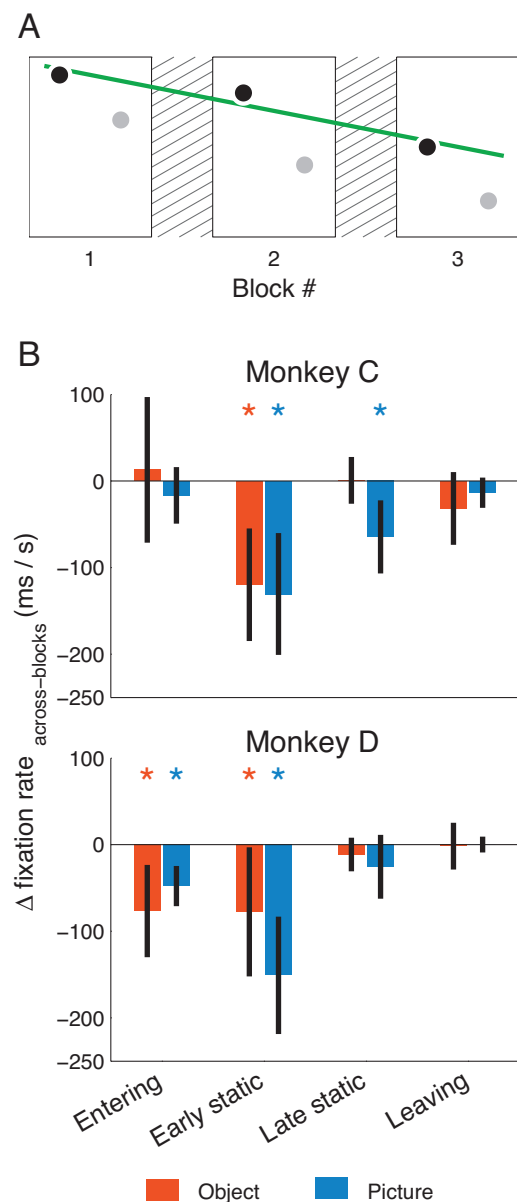
2.5.5. Habituation of fixation rates

Since habituation occurs at different rates on different timescales, variations of fixation rate were estimated within a block, across blocks and across days. For these analyses, on-stimulus fixation times exhibited during a trial were estimated within each of the four epochs. It is noteworthy that if a fixation spanned across two epochs, it was assigned to the first one. Since the trial time course was split in four epochs of different length, each fixation time was then divided by the duration of the epoch to which it was assigned. The resulting rates, measured in ms/s, represented the fixation time exhibited during a stimulus presentation of 1 s.
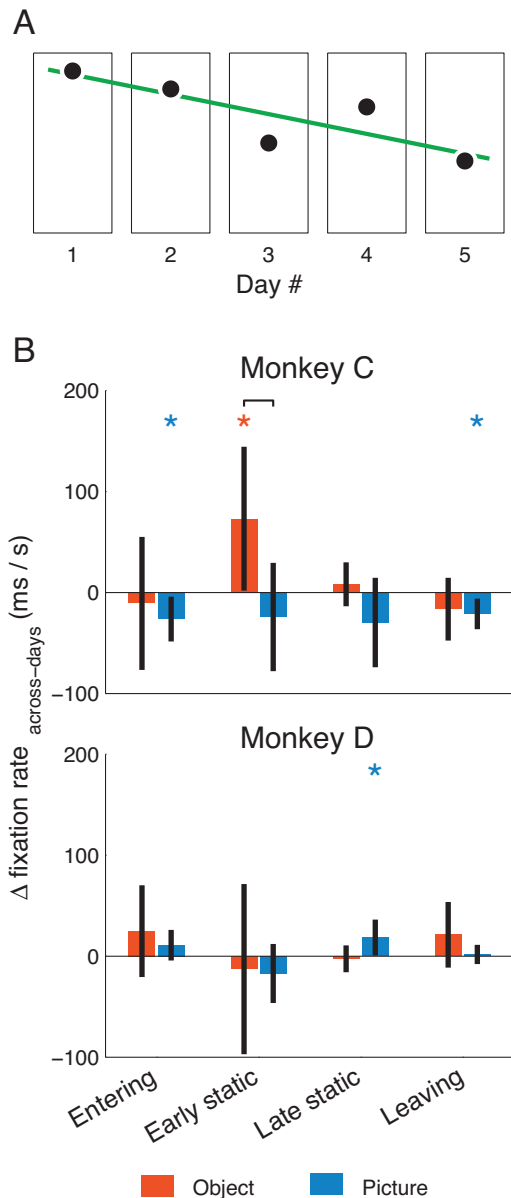
Fig. 5. Across-day habituation. (a) Schematic representation of mean fixation rate (black circles) for a stimulus during five consecutive daily sessions. The slope of the linear trend is indicated by the green line. (b) The two graphs show across-day habituation for objects (red bars) and pictures (blue bars) for the two monkeys. Bars are mean and 95% confidence intervals of the mean estimate. Stars are *p*-values that reached at the least the 0.05 significance threshold in the one-sample *t*-test against the hypothesis that habituation did not occur across days. Horizontal square bracket in top graph marks the only comparison that reached the 0.05 significance level in the post-hoc test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Within-block and across-blocks habituation indicated the extent of adaptation that occurred within a behavioral daily session. Thus, they were computed using the fixation rates of the ten daily presentations of a single stimulus. Within-block habituation was estimated as the mean difference between the second and the first presentations occurred during each block of a behavioral session (Fig. 3a), estimated independently for each monkey during each trial epoch and object/picture condition. Results are shown in Fig. 3b. The across-block habituation, instead, was defined as the slope of the regression line of the first presentations only. To obtain the total extent of habituation, slope values were multiplied by the number of blocks within a day minus one, that is 4 (Fig. 4a). We

used only the fixation rates for the first presentations to avoid any bias due to within-block habituation. Slope values were estimated during each epoch separately for the two stimulus conditions and animal subjects (Fig. 4b).

Besides, we have estimated the across-day habituation for a stimulus computing the slope of the regression line of the within-day mean fixation rates. Each slope was then multiplied by the number of daily sessions within a week minus one (namely 4) to obtain the total amount of habituation due to repeated daily exposure to the same stimuli set (Fig. 5a). Slopes were computed during all epochs for each monkey and epoch separately (Fig. 5b).

Habituation rates at each timescale were analyzed separately for each monkey. The values of habituation during an epoch were tested against the hypothesis of mean equal to 0 with the one-sample Student's *t*-test. Subsequently, habituation rates were compared within and between epochs with a two-way ANOVA with object/picture and epoch as main effects, followed by post-hoc Tukey's HSD test. Furthermore, we used the χ^2-test to examine the hypothesis that on any given timescale, a monkey could exhibit a larger degree of habituation than the other one. We compared the number of times that the post-hoc test reached significance out of the total eight possible times (object/picture condition for four epochs) in the two monkeys. We used the results of the ANOVA and the χ^2-test to give a more general interpretation about habituation rates.

2.5.6. Fixation strategies

A more systematic description of gaze strategies was performed on fixation durations, after discarding the entering and leaving epochs because the smooth-pursuit eye movements, which normally occur when stimuli move, can bias such analyses. Distributions of fixation duration were compared independently for early and late static epochs using a two-way ANOVA with monkey and object/picture as main effects, followed by the Tukey's HSD post-hoc test. Fixation duration histograms, as well as fixation frequency, are shown in Fig. 6 for the two epochs of interest (Fig. 6a and b).

2.5.7. Shape rank

The ability to create a hierarchical preference among very similar stimuli is a sign that subjects can actively distinguish among them, and look longer at some of the stimuli while being less interested in the remaining ones. One would also expect that if subjects are able to spontaneously create such preference in the real-world domain, they can also retain it when the stimuli are pictures of those objects presented on a computer monitor. We tested this hypothesis estimating a ranked-order for the four shapes during the early and late static epochs according to the mean fixation rate exhibited during each week of recording. The rank was first determined for the object stimuli and then it was applied to the picture stimuli. Data were analyzed with a two-way ANOVA for each epoch and monkey independently with object/picture and rank as main effects. Results are shown in Fig. 7 .

2.5.8. Saliency

The so-called saliency map is a spatial representation of visual relevance within an image, where pixel intensities attempt to quantify the probability that an observer will direct his gaze toward a particular region of the picture (Koch and Ullman, 1985). Thus, saliency maps were computed for each picture stimuli to assess if monkeys were selectively exploring regions predicted by a theoretical model and whether this behavior changed between early and late static epochs. In order to quantify the accuracy of the saliency map in predicting gaze behaviors, we have first computed fixation maps, also known as heatmaps, binning fixation positions in a 0.1° regularly spaced grid (Fig. 8a). Heatmaps were smoothed with a 25-pixel radius Gaussian kernel and divided by the total duration
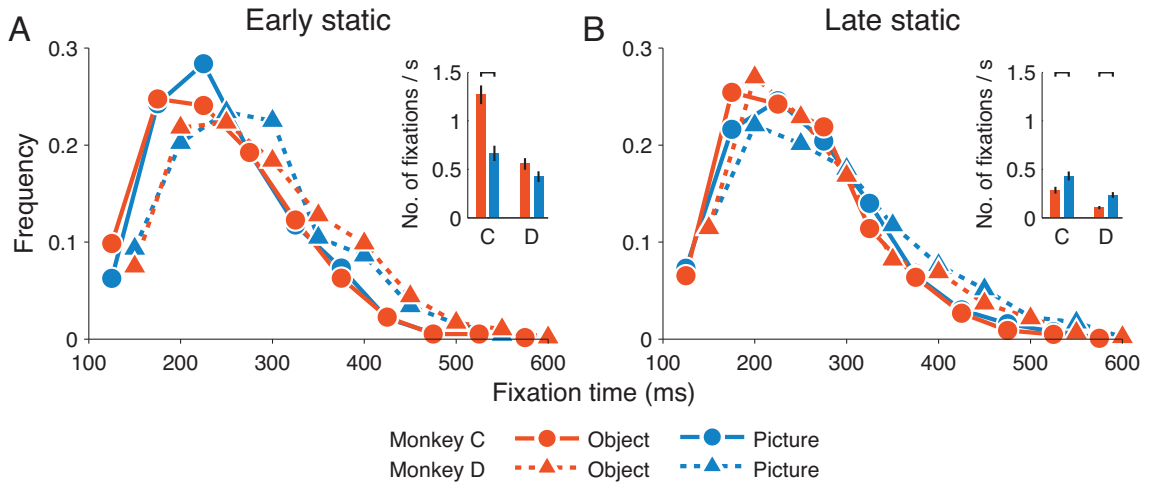
Fig. 6. Fixation statistics. Curves are frequency histograms of the fixation duration distributions. Values are grouped in 50 ms wide bins. An offset of half bin between data for monkey C (circles) and monkey D (triangles) has been introduced for clarity. For each epoch, an inset shows on-stimulus fixation frequency for the two monkeys (C and D) looking at object (red bars) or picture (blue bars) stimuli. Bars are mean and 95% confidence intervals of the mean estimate. Horizontal square brackets at the top of each graph mark comparisons that have reached the 0.05 significance level in the post-hoc test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
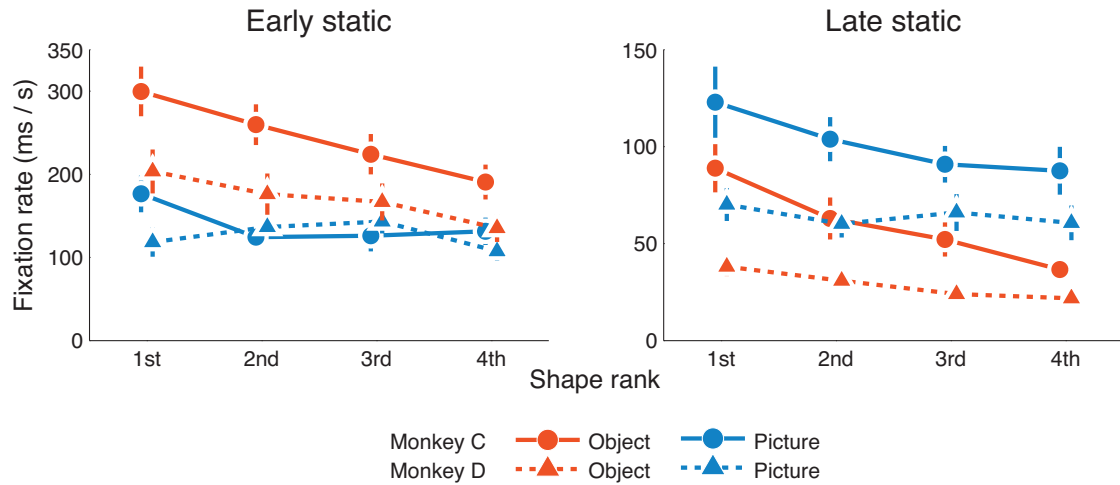

Fig. 7. Shape relative preference. Shape preference was estimated from the within-week mean fixation rate for each stimulus independently. Mean fixation rates were rank-ordered among objects and then the same order was applied to the corresponding picture stimuli. Solid and dashed lines are values for monkey C and D, respectively. Values are mean fixation rate ± standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the fixations that contributed to that map (Fig. 8b). Therefore, each point in the grid represented the frequency of fixation in that region. The empirical fixation map was next compared to a theoretical saliency map, obtained by employing an algorithm based on spectral analysis, which aims to separate foreground elements from their background (Hou et al., 2012); task that the visual system of an observer performs naturally in order to detect objects in a scene. Saliency maps were estimated in the $L^*a^*b^*$ color space, also known as CIELAB, and were smoothed with a 25-pixel radius Gaussian kernel and normalized in the range from 0 to 1 (Fig. 8c).

Then, we computed the overlap between fixation map and saliency map of each stimulus by point-wise multiplication of the two maps (Fig. 8d). In this way, frequency values contained in the fixation maps were weighted according to the predicted saliency of each pixel. The sum of all the values contained in this new map defined the degree of similarity between experimental data and the theoretical model. Values of overlap were in the range 0–1, where 0 indicated that no fixation occurred at the estimated salient pixels and 1 when there was a complete accordance between eye

fixation map and saliency model. These values have been reported in Fig. 8e as percent overlap for clarity. Values of overlap were used to perform a two-way ANOVA with object/picture and epoch as main effects, followed by the post-hoc Tukey's HSD test.

3. Results

We measured gaze fixations of two monkeys during the interleaved presentation of real objects and pictures of these objects presented on a computer monitor (see Fig. 1a and b). Monkeys were free-viewing individual stimuli as they entered, remained static and left the visual field in front of them (see Fig. 1d), and they were not rewarded for fixating the stimuli. Monkeys completed five sessions consisting of five blocks of sixteen trials on consecutive days with the same stimuli, and the experiment was repeated four times with objects with different color arrangement and their pictures being introduced at the beginning of each week (Fig. 1c).
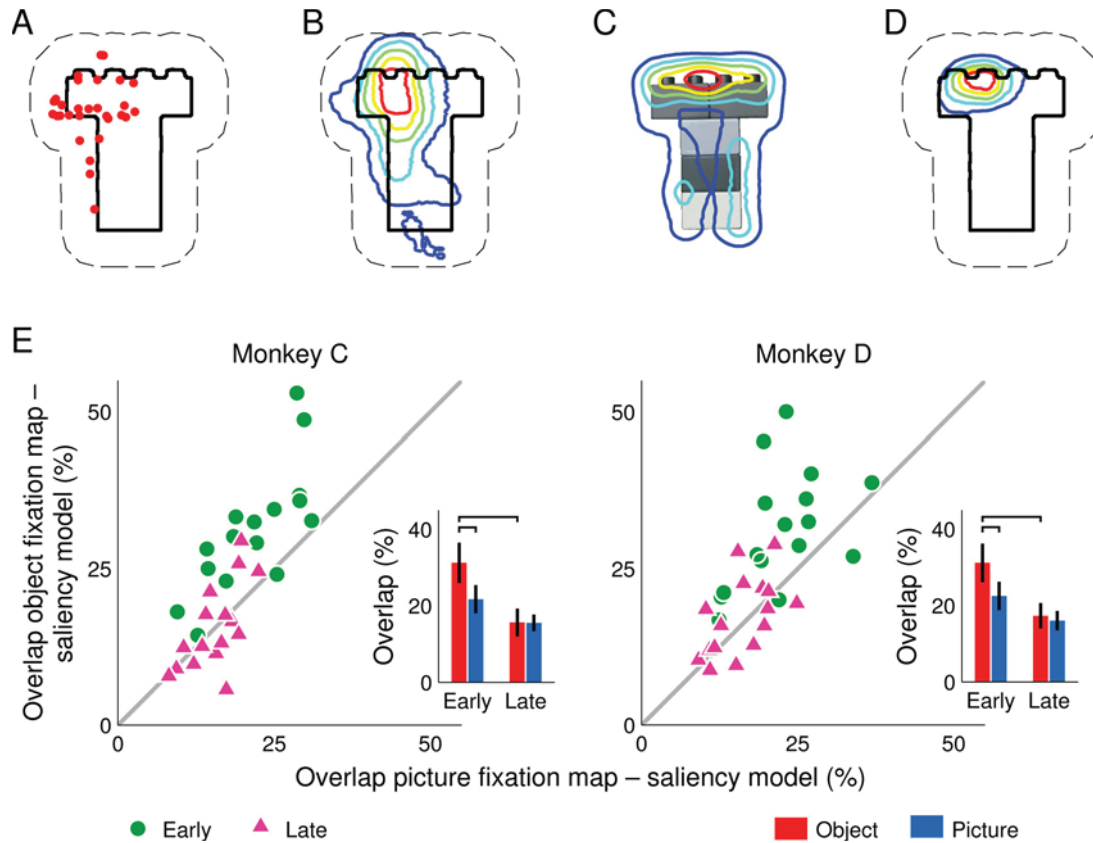
8

Fig. 8. Gaze patterns and saliency estimation. (a) For each trial, mean horizontal and vertical position (red circles) of fixations occurred during early and late static epochs were estimated. Shape 3 (solid line) and its region-of-interest (ROI; dashed line) are represented (compare with Fig. 1A). (b) Fixation map is represented as concentric lines where colors from blue to red indicate higher frequencies of fixation for particular regions of a stimulus. (c) Saliency map was normalized in the range from 0 (dark blue line) to 1 (red line). (d) Example of weighted-frequency map obtained multiplying pixel-by-pixel the fixation and the saliency maps represented in panels (b) and (c), respectively. The extent of overlap is color-coded, with dark blue lines indicating poorer overlap than red contours. (e) The two graphs show the amount of overlap for the two monkeys separately. Green circles and pink triangles indicate overlap during early and late static epochs, respectively. Gray line is the identity line. For each graph individually, object/picture differences are highlighted in an inset as bar plot. Bars for object (red bars) and picture stimuli (blue bars) are mean and 95% confidence intervals of the mean estimates of overlap between fixation map and saliency model. Horizontal square brackets indicate comparisons in which the *null*-hypothesis of equal mean was rejected with 95% confidence, according to post-hoc Tukey's HSD test. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1. Fixation of objects and pictures: within-trial dynamics

We first examined how much time monkeys spent fixating the stimuli during the different phases of the each trial, averaged across all trials (Fig. 2b). For both objects and pictures, we observed two peaks in fixation rate as the stimuli entered the visual field, and during the beginning of the static epoch. During the remaining static epoch, fixation rate was stationary and at a relatively low level. These findings suggest that stimuli captured the monkeys' overt attention when they were moving into view, as well as when they had just stopped moving. By contrast, overt attention was directed only rarely at the stimuli as they remained stationary for several seconds, with 100 ms/s fixation rate corresponding to a mere 10% of the gaze position directed to the stimuli during this epoch. Despite the overall similarities in gaze allocation, monkeys visually explored the objects longer than their pictures during stimulus movements as well as at the beginning of the static epoch. Unexpectedly, fixation rate for pictures was higher than for objects toward the middle and end of the static epoch. Based on this reversal of preference for objects versus pictures, we defined early and late static epochs. A two-way ANOVA with object/picture and epoch as factors revealed main effects of object/picture and epoch, as well as a significant interaction for both monkeys (object/picture × epoch, $F(3, 6392) = 165.99$, $\eta^2 = 0.06$ for monkey C, and $F(3, 6328) = 57.08$, $\eta^2 = 0.02$ for monkey

D, $p < 0.001$ for both monkeys), with post-hoc tests confirming significant differences in fixation rate between objects and pictures in all epochs (monkey C: $p < 0.001$ for all epochs; monkey D: $p < 0.01$ for all epochs). Monkeys fixated objects up to three times longer than pictures, particularly during movement epochs when stimuli were entering or leaving the visual field, while pictures were preferentially viewed up to two times longer during the late static epoch (see insets Fig. 2b). The two monkeys showed overall differences in fixation rate during the entering and early static epochs, with monkey C being more motivated for visual exploration with peak fixation rate of 500 ms/s, compared to about 300 ms/s for monkey D. Despite of this, the patterns described above were highly consistent across animals, suggesting that they do not depend on overall motivation but reflect stimulus-dependent gaze dynamics.

3.2. Fixation of objects and pictures: dynamics across trials

To examine how gaze allocation varied with stimulus repetition, we compared fixation rates on repeated trials where monkeys were viewing identical stimuli, at three different timescales: within blocks, across blocks and across days. Since each stimulus was shown twice during each block, the within-block variation was estimated as the difference in fixation rate between the two presentations (Fig. 3a). We observed generally reduced fixation rates on the second compared to the first presentation for both objects

and pictures (Fig. 3b). This habituation was significant during the entering epoch for objects and pictures in both monkeys (one-sample t-test; monkey C: $t(70) = -2.26$, $d = -0.3$, and $t(79) = -2.07$, $d = -0.2$, for objects and pictures, respectively, $p < 0.05$; monkey D: $t(68) = -3.11$, $d = -0.4$, and $t(79) = -4.22$, $d = -0.5$, for objects and pictures, respectively, $p < 0.05$), whereas in the remaining epochs monkey D showed more extensive evidence for habituation than monkey C (7/8 vs. 3/8 cases, χ^2-test: $p < 0.05$). Thus, it was monkey D, which exhibited already relatively lower overall fixation of visual stimuli, whose values habituated strongly on repeated presentation, whereas monkey C showed high overall fixation, and these high values habituated relatively little. It is noteworthy that fixation rate habituation of monkey D was significantly larger for pictures than objects during the early static epoch (post-hoc test, $p < 0.001$), suggesting reduced interest for visual exploration of pictures but not objects during this epoch.

To estimate the across-block variation for fixation rate, we computed a linear trend for the fixation rates for the first stimulus presentations of each block (Fig. 4a). We observed a general decrease in fixation rate for objects and pictures in both monkeys (Fig. 4b). The habituation was significant in both monkeys for the early static epoch (one-sample t-test; monkey C: $t(79) = -3.67$, $d = -0.4$, and $t(79) = -3.69$, $d = -0.4$, for objects and pictures, respectively, $p < 0.001$; monkey D: $t(79) = -2.07$, $p < 0.001$, $d = -0.2$, and $t(79) = -4.43$, $p < 0.05$, $d = -0.5$, for objects and pictures, respectively), suggesting that visual exploration during this epoch was most sensitive to repetition. By comparison, habituation during the entering epoch was weaker in monkey D and absent in monkey C, suggesting that the moving stimulus captured the overt attention of the monkeys consistently during the course of daily sessions. A two-way ANOVA for each monkey revealed no significant effects of object/picture or object/picture × epoch interactions (object/picture main effect, monkey C: $F(1, 632) = 1.39$, $p = 0.24$; monkey D: $F(1, 632) = 0.80$, $p = 0.37$), suggesting that across-block habituation similarly affected objects and pictures.

Finally, we examined variations in fixation rate across days, on the basis of linear trends of the daily average fixation rates for each of the stimuli during each week (Fig. 5a). In contrast to the above findings, we did not see strong evidence for across-day habituation in either monkey (Fig. 5b). This suggests that monkeys were generally similarly motivated for visual exploration during each daily behavioral session. Indeed, monkey C even showed some evidence for sensitization to objects (effect size: 73 ms/s), but not pictures, during the early static epoch (post-hoc test, $p < 0.03$). This is consistent with a selective increase in visual exploratory behavior for objects over pictures, which is restricted to the early static epoch in monkey C. Monkey D also exhibited some evidence for sensitization for the picture stimuli in the late static epoch; however, this effect was small in magnitude (18 ms/s).

3.3. Visual exploration strategies for stationary stimuli

We next examined visual exploration strategies of the two monkeys, to reveal potential differences between object and picture gaze dynamics. We estimated the duration and number of fixation periods, that is periods of constant gaze position in each monkey during the early and late static epochs. The epochs with moving stimuli were not considered in this analysis, because pursuit eye movements often occur when observing moving stimuli and this complicates systematic comparison between epochs. In the early static epoch (see Fig. 6a), a two-way ANOVA revealed significant effects of monkey subject ($F(1, 3003) = 139.14$, $p < 0.001$, $\eta^2 = 0.04$), but not object/picture or factors interaction ($F(1, 3003) = 0.89$, $p = 0.3$ and $F(1, 3003) = 3.18$, $p = 0.07$, respectively). The mean fixation duration was greater in monkey D ($M = 282$ ms) than in monkey C ($M = 243$ ms), suggesting that monkey C appeared to be executing

the next saccadic eye movement more quickly than monkey D. The lower fixation duration in monkey C was offset by a greater number of fixations for both stimulus types during the early static epoch (see inset Fig. 6). A two-way ANOVA confirmed the significance of this effect (monkey × object/picture interaction, $F(1, 3083) = 39.16$, $p < 0.001$, $\eta^2 = 0.01$; post-hoc tests, object condition $p < 0.001$, picture condition $p < 0.001$). In addition, monkey C fixated more often on objects than pictures (post-hoc tests; $p < 0.001$); an effect that did not reach significance in monkey D ($p = 0.08$). Taken together, objects tended to attract gaze more often than pictures in the early static epoch and the duration of individual fixations was similar for objects and pictures.

In the late static epoch (see Fig. 6b), monkey D also exhibited longer mean fixation durations ($M = 278$ ms) than monkey C ($M = 251$ ms; two-way ANOVA, $F(1, 4585) = 74.38$, $p < 0.01$, $\eta^2 = 0.02$), in similarity to the early static epoch as discussed above. By contrast however, there was a significant effect of object/picture in absence of a significant interaction ($F(1, 4585) = 16.89$, $p < 0.001$, $\eta^2 = 0.004$, and $F(1, 4585) = 2.00$, $p > 0.1$, respectively), with fixations lasting slightly longer ($M = 12$ ms) for pictures than objects in both monkeys. In terms of the number of fixations (Fig. 6 inset), the late static epoch presented the opposite pattern to the early static period, in that pictures were fixated more frequently than objects by both animals (post-hoc tests, monkey C: $p < 0.001$; monkey D: $p < 0.001$). This suggests a shift in visual exploration strategy from the early to the late static period, with a minor prolongation of fixation duration preferentially for pictures accompanied by a pronounced shift in the number of fixations from favoring objects to favoring pictures. Note that the overall number of stimulus fixations decreased by about a factor of 3 from the early to the late static epoch, reflecting a reduction of overt attention to the stimuli.

3.4. Comparison of object and picture fixation preferences

Here, we were interested in examining whether there was any relationship between spontaneous free viewing of object and corresponding picture stimuli for each of the participant monkeys. We therefore rank ordered the four shapes according to the fixation rate during the object stimulus presentation, and applied this same rank to the picture stimuli to reveal fixation rate for 1st–4th shape (see Fig. 7). The red curves representing object fixations exhibit negative slopes, reflecting the rank ordering according to shape. Of interest are the blue curves corresponding to the picture fixations, where tendencies for negative slopes indicate consistent shape preference across object and picture conditions. A two-way ANOVA with shape rank and object/picture as factors revealed main effects of both factors (object/picture factor: $F(1, 152) = 45.95$, $p < 0.001$, $\eta^2 = 0.21$, and $F(1, 152) = 24.95$, $p < 0.001$, $\eta^2 = 0.13$, for early and late static epochs, respectively; shape rank factor: $F(3, 152) = 4.76$, $p = 0.003$, $\eta^2 = 0.07$, and $F(3, 152) = 5.26$, $p = 0.002$, $\eta^2 = 0.08$, for early and late static epochs, respectively) and no significant interaction ($p > 0.1$, for both epochs) for monkey C in both early and late static epochs. Monkey C thus provides evidence for systematic shape preference that was consistent between object and picture conditions. For monkey D, only the object/picture, but not the other factor or the interaction reached significance in both epochs (object/picture factor: $F(1, 152) = 10.84$, $p = 0.001$, $\eta^2 = 0.06$, and $F(1, 152) = 57.43$, $p < 0.001$, $\eta^2 = 0.27$, for early and late static epochs, respectively; remaining p-values > 0.1). Monkey D did not distinguish systematically between shapes, maybe owing to the overall similarity between the four shapes, and thus does not provide any evidence related to similar shape preference for the two stimulus types. Overall, when shape selectivity was present it did appear to be similar for objects and pictures, which suggests that monkeys do appear to treat objects and the corresponding pictures with a degree of equivalence.

10

Finally, we wanted to compare the monkeys' fixation patterns with saliency models that predict which spatial regions of visual stimuli attract gaze through bottom-up factors such as edges, color contrast or figure-ground segregation. Our procedure is illustrated in Fig. 8, showing fixation points on a single trial, the corresponding fixation map for a particular shape, the result of the saliency model as well as the overlap between the experimental fixation map and the saliency model output. The experiment-model overlap is illustrated in Fig. 8e for the two epochs and monkeys, where each data point in the scatter plot corresponds to the overlap estimated for a single shape based on behavioral data from one week. In the late static epoch, the saliency model predicted the fixation maps for objects and pictures equally well, but for both cases relatively poorly. On the other hand, during the early static epoch, the saliency model performed better at predicting the fixation map for objects than for pictures (post-hoc test, monkey C: $p < 0.05$; monkey D: $p < 0.05$). This suggests that even though monkeys spend more time visually exploring the pictures during the late static epoch (compare Fig. 2), this additional exploration is not directed particularly toward the salient features. Conversely, free viewing of objects in the early static period corresponds most closely to active visual exploration of salient features as predicted by the model.

4. Discussion

4.1. Dynamics of visual exploration

Our study presents comprehensive results on the temporal dynamics of visual exploration behavior in macaque monkeys, which showed an overall similar pattern for objects and pictures. Stimuli particularly captured the monkeys' overt attention when they were entering the visual field and when they stopped moving on this trajectory, as demonstrated by two peaks of fixation rate occurring in the first several seconds during the trial. The existence of a second peak may at first seem surprising, but it is consistent with previous findings showing enhanced visual exploration of objects just after they stopped their movement on a predictable trajectory (Berkson, 1965), and it is likely elicited by the violation of expectancy that the object will continue moving on the initial trajectory. There was a large preference for objects over pictures during this initial period of visual exploration, suggesting that real three-dimensional shapes were much more efficient in capturing monkeys' overt attention than pictorial representations on a computer monitor. This preference was more pronounced in one of the monkeys, probably because of greater curiosity and overall exploratory motivation in this animal. Later during the viewing period, this preference reversed such that pictures were actually explored more than objects. This finding represents within-trial habituation, which is more pronounced for objects that had just been in the attentional focus.

We quantify across-trial habituation at different timescales, and show that it occurs reliably within daily viewing sessions, consistent with previous reports (Jutras and Buffalo, 2010; Wilson and Goldman-Rakic, 1994). We found little evidence of across-day habituation, indicating that monkeys' overall exploratory motivation for the same stimulus set remains constant over five daily sessions. This is somewhat surprising, given that across-day habituation of exploration of similar object stimuli has been demonstrated on other species including rodents (Reger et al., 2009) and tree shrews (Khani and Rainer, 2012). We suggest that exploration in the form of locomotion may be more sensitive to repetition than gaze exploration, since it involves energy expenditure and exposure to environmental risks; factors that are not associated with eye movements.

4.2. Object-picture perception

Our findings suggest good correspondence between object and picture perception, due to the many similarities in spontaneous visual exploration of the two stimulus types. These similarities encompass on the one hand the general dynamics of visual exploration within and across trials, as discussed above. Another important aspect of similarity concerns the preference among the four shapes that are presented during each week, which provides evidence for consistent preference between the two stimulus types. Thus, the shape that was most explored as an object also tended to be heavily explored in the picture condition. Despite these similarities, a number of aspects of visual exploration differed substantially between objects and pictures. These differences include the greater overall exploration time for objects during and immediately following stimulus movements described above. In addition, monkeys' gaze patterns focused on the salient regions (Hou et al., 2012) of the objects to a greater extent than the corresponding pictures, suggesting that a greater degree of directed attentional effort is elicited by objects than pictures. Along related lines, we found partial evidence for visual exploration habituation to pictures compared to objects at two different timescales: with monkey D showing greater within-block habituation in the early static period and monkey C actually showing sensitization, which is increase, of visual exploration across-days. It is possible that the sensitization was a result of an initial fearful predisposition of this monkey to the real objects, which was associated with an avoidance of direct visual exploration. Taken together, the above evidence suggests that monkeys could perceive differences between objects and pictures, despite overall similar exploration behavior dynamics.

Independence, confusion, and equivalence have been advanced as three possible types of object-picture relationships (Fagot and Parron, 2010; Fagot et al., 2010). The consistent shape preference between stimuli observed in one monkey suggests that monkeys are able to make a connection between real objects and their pictorial representation regardless of the low-level features of the stimuli, which makes the notion of independence between objects and pictures unlikely. It seems unlikely that macaque monkeys perceive objects and pictures according to the confusion modality, which is the inability to distinguish between the two stimulus types, as we observed numerous differences in gaze-dependent parameters as mentioned above. We interpret these differences in eye movement parameters as evidence that the monkeys could distinguish between objects and pictures. It remains possible that picture-naïve macaque monkeys might nevertheless attempt to reach and manipulate pictures of objects, as previously described for baboons and gorillas (Parron et al., 2008). Whether this is in fact the case needs to be addressed in future experiments, and at present, we cannot completely rule out the confusion modality based on our experimental evidence. Our experimental design does not allow us to address the notion of equivalence, but taken together, we conclude that there certainly appears to be a correspondence in how crab-eating macaques perceive objects and their pictorial representations. This conclusion is in line with previous results in capuchin monkeys (Truppa et al., 2009) and lion-tailed macaques (Judge et al., 2012).

Our findings validate visual exploration behavior as a method for studying object-picture perception, with the potential advantage that it does not require behavioral training which may be useful for studying populations of primates also outside the laboratory.

4.3. Fixation strategies

An analysis of the fixation durations revealed substantial inter-individual differences in fixation duration between the two participant monkeys, a finding which has been also reported

recently for capuchin monkeys (Berger et al., 2012) and great apes (Kano et al., 2011). Surprisingly, monkey C exhibited shorter individual fixation durations for objects during the early static epoch than monkey D, despite the overall longer time spent in object exploration. This is explained by the greater number of on-stimulus fixations for this monkey. This suggests that during particularly active periods of visual exploration, a large number of stimulus fixations together with short fixation durations may optimize information acquisition by the visual system. While objects were more often targeted by fixations than pictures in the early static epoch, the pattern was reversed during the late static epoch where a greater number of fixations targeted pictures. This shift in visual exploration strategy may be related to findings in humans suggesting that gaze parameters such as fixation duration and saccade amplitude vary with viewing time (Pannasch et al., 2008).

We have applied a saliency model to compare on-stimulus gaze patterns to predicted maps based on bottom-up visual attributes. The performance of saliency models in predicting monkey visual exploration has been examined in a number of studies (Berg et al., 2009; Einhauser et al., 2006), which are however not directly comparable to our findings due to the stimulus material. Previous work has focused on predicting gaze patterns for natural scenes containing large amounts of distributed spatial structure, whereas our study has applied one of these models to isolated single shape stimuli on a homogenous background. Although saliency models were developed for the former situation, our findings suggest that they can also be applied usefully to more sparse visual scenes. Visual saliency related advantages for objects over pictures were largest in the early static epoch, the period during which the strongest habituation or sensitization occurred and which exhibited the most striking differences in fixation statistics. The early static epoch thus offers the most sensitive opportunity to measure cognitive modulations related to visual stimulus processing.

5. Conclusions

Since the advent of computer monitors, studies on the neural representation of objects is generally performed using pictures presented on a computer monitor (Logothetis, 2000; Tanaka, 1996), although some of the now classical early studies did in fact use real objects mostly on an informal basis (Desimone et al., 1984) without systematically comparing responses between the two stimulus types. However, conclusions based on pictures about the neural basis of object representation in relevant brain areas such as the inferior temporal cortex may only partially generalize to actual objects, given the substantial differences in spontaneous viewing behavior directed toward the two types of stimuli. This is particularly relevant for studies that have implicitly examined transfer of knowledge from the object to the picture domain (Booth and Rolls, 1998). Macaque monkeys often require many months of operant conditioning before they produce the required behavioral task at hand, and a substantial portion of this time is spent familiarizing animals with computer monitors. Our study suggests that because monkeys' overt interest in real objects is several times greater than for the corresponding pictures, this part of the training might be drastically shortened. Of course, the set of stimuli is limited to what can be physically arranged within the grasp of a robot arm, so certain studies involving large sets of stimuli or parametric manipulations may not translate easily to an object setting. Nevertheless, a large body of work in the areas of cognitive neuroscience, memory research or neuroeconomics could benefit from employing object stimuli.

Taken together, our findings are in line with previous studies on object-picture perception in non-human primates, suggesting that macaques are able to make a connection between objects and pictures while not treating them in an equivalent manner. We consider that the partial transfer of knowledge from the object to the picture domain seen in many previous studies is thus related to the greater overall looking time for objects compared to pictures in the present study. Previous work has focused on marmoset and capuchin monkeys, as well as baboons and great apes, with only limited work done on macaques, so our results complement the previous literature in this area. In addition, our findings may explain results in humans (Snow et al., 2014) and other animals (O'Hara et al., 2015) showing that objects tend to be remembered better than pictures. We suggest that the greater looking time spent on objects compared to pictures may be the reason for the superior memory performance, since the additional sensory stimulation may enhance memory formation for object stimuli.

References

Antunes, M., Biala, G., 2012. The novel object recognition memory: neurobiology, test procedure, and its modifications. Cogn. Process. 13, 93–110.

Berg, D.J., Boehnke, S.E., Marino, R.A., Munoz, D.P., Itti, L., 2009. Free viewing of dynamic stimuli by humans and monkeys. J. Vision 9, 1–15, 19.

Berger, D., Pazienti, A., Flores, F.J., Nawrot, M.P., Maldonado, P.E., Grun, S., 2012. Viewing strategy of Cebus monkeys during free exploration of natural images. Brain Res. 1434, 34–46.

Berkson, G., 1965. Eye fixation responses to kittens and infant chimpanzees to onset and offset of a stimulus. Percept. Motor Skills 21, 619–624.

De Luna, P., Mohamed Mustafar, M.F.B., Rainer, G., 2014. A MATLAB-based eye tracking control system using non-invasive helmet head restraint in the macaque. J. Neurosci. Methods 235, 41–50.

Booth, M.C.A., Rolls, E.T., 1998. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb. Cortex 8, 510–523.

Bovet, D., Vauclair, J., 2000. Picture recognition in animals and humans. Behav. Brain Res. 109, 143–165.

Brainard, D.H., 1997. The psychophysics toolbox. Spatial Vision 10, 433–436.

Butler, R.A., Alexander, H.M., 1955. Daily patterns of visual exploratory behavior in the monkey. J. Comp. Physiol. Psychol. 48, 247–249.

Cabe, P.A., 1976. Transfer of discrimination from solid objects to pictures by pigeons – test of theoretical models of pictorial perception. Percept. Psychophys. 19, 545–550.

Cannon, E.N., Woodward, A.L., Gredeback, G., von Hofsten, C., Turek, C., 2012. Action production influences 12-month-old infants' attention to others' actions. Dev. Sci. 15, 35–42.

Dal Monte, O., Noble, P.L., Costa, V.D., Averbeck, B.B., 2014. Oxytocin enhances attention to the eye region in rhesus monkeys. Front. Neurosci. 8, 41.

Desimone, R., Albright, T.D., Gross, C.G., Bruce, C., 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci. 4, 2051–2062.

Einhauser, W., Kruse, W., Hoffmann, K.-P., Konig, P., 2006. Differences of monkey and human overt attention under natural conditions. Vision Res. 46, 1194–1209.

Emile, N., Barros, M., 2009. Recognition of a 3D snake model and its 2D photographic image by captive black tufted-ear marmosets (*Callithrix penicillata*). Anim. Cogn. 12, 725–732.

Fagot, J., Parron, C., 2010. Picture perception in birds: perspective from primatologists. Comp. Cogn. Behav. Rev. 5, 132–135.

Fagot, J., Thompson, R.K., Parron, C., 2010. How to read a picture: lessons from nonhuman primates. Proc. Natl. Acad. Sci. U. S. A. 107, 519–520.

Gammon, W.D., Singer, J.J., Michels, K.M., 1972. Visual exploratory behavior in the squirrel monkey (*Saimiri sciureus*). Percept. Motor Skills 34, 715–718.

Ghazanfar, A.A., Nielsen, K., Logothetis, N.K., 2006. Eye movements of monkey observers viewing vocalizing conspecifics. Cognition 101, 515–529.

Gothard, K.M., Erickson, C.A., Amaral, D.G., 2004. How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? Anim. Cogn. 7, 25–36.

Gottlieb, J., Oudeyer, P.Y., Lopes, M., Baranes, A., 2013. Information-seeking, curiosity, and attention: computational and neural mechanisms. Trends Cogn. Sci. 17, 585–593.

Gunhold, T., Whiten, A., Bugnyar, T., 2014. Video demonstrations seed alternative problem-solving techniques in wild common marmosets. Biol. Lett. 10, http://dx.doi.org/10.1098/rsbl.2014.0439

Hannula, D.E., Althoff, R.R., Warren, D.E., Riggs, L., Cohen, N.J., Ryan, J.D., 2010. Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. Front. Hum. Neurosci. 4, 166.

Hou, X.D., Harel, J., Koch, C., 2012. Image signature: highlighting sparse salient regions. IEEE Trans. Pattern Anal. Mach. Intell. 34, 194–201.

Humphrey, N.K., 1972. 'Interest' and 'pleasure': two determinants of a monkey's visual preferences. Perception 1, 395–416.

Humphrey, N.K., 1974. Species and individuals in perceptual world of monkeys. Perception 3, 105–114.

Itti, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res. 40, 1489–1506.

Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox -3? Perception 36.

Judge, P.G., Kurdziel, L.B., Wright, R.M., Bohrman, J.A., 2012. Picture recognition of food by macaques (*Macaca silenus*). Anim. Cogn. 15, 313–325.

Jutras, M.J., Buffalo, E.A., 2010. Recognition memory signals in the macaque hippocampus. Proc. Natl. Acad. Sci. U. S. A. 107, 401–406.

Kano, F., Call, J., 2014. Great apes generate goal-based action predictions: an eye-tracking study. Psychol. Sci. 25, 1691–1698.

Kano, F., Hirata, S., Call, J., Tomonaga, M., 2011. The visual strategy specific to humans among hominids: a study using the gap-overlap paradigm. Vision Res. 51, 2348–2355.

Kano, F., Tomonaga, M., 2009. How chimpanzees look at pictures: a comparative eye-tracking study. Proc. R. Soc. B: Biol. Sci. 276, 1949–1955.

Kano, F., Tomonaga, M., 2010. Face scanning in chimpanzees and humans: continuity and discontinuity. Anim. Behav. 79, 227–235.

Khani, A., Rainer, G., 2012. Recognition memory in tree shrew (Tupaia belangeri) after repeated familiarization sessions. Behav. Processes 90 (3), 364–371, http://dx.doi.org/10.1016/j.beproc.2012.03.019

Koehler, K., Guo, F., Zhang, S., Eckstein, M.P., 2014. What do saliency models predict. J. Vision 14, 14.

Logothetis, N.K., 2000. Object recognition: holistic representations in the monkey brain. Spatial Vision 13, 165–178.

Koch, C., Ullman, S., 1985. Shifts in selective visual-attention – towards the underlying neural circuitry. Hum. Neurobiol. 4, 219–227.

O'Hara, M., Huber, L., Gajdon, G.K., 2015. The advantage of objects over images in discrimination and reversal learning by kea, Nestor notabilis. Anim. Behav. 101, 51–60.

Pannasch, S., Helmert, J.R., Roth, K., Herbold, A.K., Walter, H., 2008. Visual fixation durations and saccade amplitudes: shifting relationship in a variety of conditions. J. Eye Mov. Res. 2 (2), 1–19.

Parron, C., Call, J., Fagot, J., 2008. Behavioural responses to photographs by pictorially naive baboons (*Papio anubis*), gorillas (*Gorilla gorilla*) and chimpanzees (*Pan troglodytes*). Behav. Processes 78, 351–357.

Rabedeau, R., Miles, R.C., 1959. Response decrement in visual exploratory behavior. J. Comp. Physiol. Psychol. 52, 364–367.

Railton, R., Foster, T.M., Temple, W., 2014. Object/picture recognition in hens. Behav. Processes 104, 53–64.

Reger, M.L., Hovda, D.A., Giza, C.C., 2009. Ontogeny of rat recognition memory measured by the novel object recognition task. Dev. Psychobiol. 51, 672–678.

Schutz, A.C., Braun, D.I., Gegenfurtner, K.R., 2011. Eye movements and perception: a selective review. J. Vision 11, http://dx.doi.org/10.1167/11.5.9

Machado, C.J., Nelson, E.E., 2011. Eye-tracking with nonhuman primates is now more accessible than ever before. Am. J. Primatol. 73, 562–569.

Sigala, R., Logothetis, N.K., Rainer, G., 2011. Own-species bias in the representations of monkey and human face categories in the primate temporal lobe. J. Neurophysiol. 105, 2740–2752.

Snow, J.C., Skiba, R.M., Coleman, T.L., Berryhill, M.E., 2014. Real-world objects are more memorable than photographs of objects. Front. Hum. Neurosci. 8, http://dx.doi.org/10.3389/fnhum.2014.00837

Snyder, K.A., Blank, M.P., Marsolek, C.J., 2008. What form of memory underlies novelty preferences. Psychon. Bull. Rev. 15, 315–321.

Southgate, V., Senju, A., Csibra, G., 2007. Action anticipation through attribution of false belief by 2-year-olds. Psychol. Sci. 18, 587–592.

Spetch, M.L., Friedman, A., 2006. Pigeons see correspondence between objects and their pictures. Psychol. Sci. 17, 966–972.

Tanaka, K., 1996. Inferotemporal cortex and object vision. Annu. Rev. Neurosci. 19, 109–139.

Tatler, B.W., Hayhoe, M.M., Land, M.F., Ballard, D.H., 2011. Eye guidance in natural vision: reinterpreting salience. J. Vision 11, http://dx.doi.org/10.1167/11.5.5

Truppa, V., Spinozzi, G., Stegagno, T., Fagot, J., 2009. Picture processing in tufted capuchin monkeys (*Cebus apella*). Behav. Processes 82, 140–152.

Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols. Proc. Eye Tracking Res. Appl. Symp., 71–78.

Watanabe, S., 1997. Visual discrimination of real objects and pictures in pigeons. Anim. Learn. Behav. 25, 185–192.

Wilkinson, A., Mueller-Paul, J., Huber, L., 2013. Picture-object recognition in the tortoise *Chelonoidis carbonaria*. Anim. Cogn. 16, 99–107.

Wilson, F.A., Goldman-Rakic, P.S., 1994. Viewing preferences of rhesus monkeys related to memory for complex pictures, colours and faces. Behav. Brain Res. 60, 79–89.

Zimmerman, R.R., Hochberg, J., 1970. Responses of infant monkeys to pictorial representations of a learned visual discrimination. Psychon. Sci. 18, 307–308.