

Free viewing of dynamic stimuli by humans and monkeys

David J. Berg

Neuroscience Graduate Program, University of Southern
California, Los Angeles, CA, USA



Susan E. Boehnke

Centre for Neuroscience Studies, Queen's University,
Kingston, ON, Canada



Robert A. Marino

Centre for Neuroscience Studies, Queen's University,
Kingston, ON, Canada



Douglas P. Munoz

Centre for Neuroscience Studies and
Departments of Physiology, Psychology, and Medicine,
Queen's University, Kingston, ON, Canada



Laurent Itti

Department of Computer Science and Neuroscience
Graduate Program, University of Southern California,
Los Angeles, CA, USA



Due to extensive homologies, monkeys provide a sophisticated animal model of human visual attention. However, for electrophysiological recording in behaving animals simplified stimuli and controlled eye position are traditionally used. To validate monkeys as a model for human attention during realistic free viewing, we contrasted human ($n = 5$) and monkey ($n = 5$) gaze behavior using 115 natural and artificial video clips. Monkeys exhibited broader ranges of saccadic endpoints and amplitudes and showed differences in fixation and intersaccadic intervals. We compared tendencies of both species to gaze toward scene elements with similar low-level visual attributes using two computational models—luminance contrast and saliency. Saliency was more predictive of both human and monkey gaze, predicting human saccades better than monkey saccades overall. Quantifying interobserver gaze consistency revealed that while humans were highly consistent, monkeys were more heterogeneous and were best predicted by the saliency model. To address these discrepancies, we further analyzed high-interest gaze targets—those locations simultaneously chosen by at least two monkeys. These were on average very similar to human gaze targets, both in terms of specific locations and saliency values. Although substantial quantitative differences were revealed, strong similarities existed between both species, especially when focusing analysis onto high-interest targets.

Keywords: visual attention, computational model, salience, free viewing, saccades, monkeys

Citation: Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *Journal of Vision*, 9(5):19, 1–15, <http://journalofvision.org/9/5/19/>, doi:10.1167/9.5.19.

Introduction

Monkeys are widely used as animal models for the study of human cognitive processes, such as visual attention, due to the neural homologies between the species. More and more, there is a shift toward studying vision using natural and dynamic stimuli. When the visual system is examined using such stimuli, it responds differently than it does to simple stimuli traditionally used in the laboratory (for reviews, see Felsen & Dan, 2005; Kayser, Körding, & König, 2004; Reinagel, 2001; Simoncelli & Olshausen, 2001). The system also responds differently when monkeys view such stimuli freely (Dragoi & Sur, 2006; Gallant, Connor, & Van Essen, 1998; Vinje & Gallant, 2000). What is not yet known is whether humans and monkeys *behave* similarly under such natural viewing

conditions. This is important because, although there are similarities in the early stages of visual processing, cortical architecture differences exist in parietal and frontal areas related to attention and cognitive processing (Orban, Van Essen, & Vanduffel, 2004).

Computational models (Itti, Koch, & Niebur, 1998; Le Meur, Le Callet, & Barba, 2007; Privitera & Stark, 2000) provide a quantitative framework to assess visual behavior and compare species under complex stimulus conditions. For example, model output for the scene can be investigated at actual saccadic target locations. Simple image statistics (such as local contrast, orientation) and deviations from global image statistics exhibit differences between fixated vs. non-fixated locations (Parkhurst & Niebur, 2003; Reinagel & Zador, 1999), and these statistics are factors in guiding attention. Such experiments have been done with monkeys (Dragoi & Sur, 2006) and humans (Itti, 2005;

Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Tatler, Baddeley, & Gilchrist, 2005) separately. However, viewing behavior has yet to be compared directly using a wide set of complex dynamic natural stimuli (video).

To investigate species correspondence, Einhäuser, Kruse, Hoffmann, and König (2006) compared 2 monkeys and 7 humans who repeatedly viewed static, grayscale natural images. Computational models equally predicted the species gaze shifts, however, differences in viewing strategies were observed when local image contrast was manipulated. Here, we expand on this significantly by comparing human and monkey free-viewing behavior using video clips ranging in semantic content and species relevance. Additionally, the main computational model of viewing behavior, the saliency model, was adapted to better account for the temporal dynamics of video (Itti & Baldi, 2006). We also measured consistency among observers' gaze, which provided context specific predictions of saccadic targets that complement the stimulus-driven predictions of the saliency model.

Our results demonstrate correlations between saliency and both human and monkey visual behaviors; however, marked differences exist between species in eye movement statistics, model correspondence, and interobserver consistency. These differences must be considered when using monkeys as a model of human attention during free viewing. We find that focusing analysis on a subset of high-interest gaze locations—to which two or more monkeys looked simultaneously—can alleviate such differences. We speculate that high-interest locations reveal commonalities between both species, possibly by emphasizing the role of their largely homologous and common low-level visual systems over their likely more different and individualized cognitive systems.

Methods

Subjects

Eye movements during free viewing were recorded from five human (two male) and five monkey (*Macaca Mulatta*, all male) subjects. Human subjects provided informed consent under a protocol approved by the Institutional Review Board of the University of Southern California. Monkeys were used with approval by the Queen's University Animal Care Committee and were in accordance with the Canadian Council on Animal Care policy on the use of laboratory animals and the Policies on the Use of Animals and Humans in Neuroscience Research of the Society for Neuroscience.

Stimulus presentation

Naive subjects (both human and monkey) watched 115 video clips (totaling approx. 27 minutes in duration,

played in random order) that varied in duration and semantic content. The clips were subjectively categorized into six coarse semantic groups (Building/City, Natural, Sports, Indoor, Non-natural, and Monkey-relevant), as shown in Figure 1. Stimuli were collected from television (NTSC source) with a commercial frame grabber (ATI Wonder Pro). Monkey-relevant clips were collected at the Queen's University animal care facility with a consumer grade digital video camera. Frames were acquired and stored at 30 Hz in raw 640×480 RGB555 format and compressed to MPEG-1 movies (640×480 pixels). Stimuli were presented to human subjects, with head stabilized by a chin rest, on a 101.6×57.2 cm LCD TV (Sony Bravia) at a viewing distance of 97.8 cm. This provided a usable field of view of $54.9^\circ \times 32.6^\circ$, which was the largest the video-based human eye-tracker could accommodate. Stimulus presentation was orchestrated using a Linux computer running in house-programmed presentation software (downloadable at <http://iLab.usc.edu/toolkit>) under SCHED_FIFO scheduling to ensure proper frame rate presentation (Finney, 2001; Itti, 2005). Subjects were given minimal instructions—"watch and enjoy the video clips, try to pay attention, but don't worry about small details." Each video presentation was preceded by a fixation point, and the next video began when the subject pressed the space bar.

The exact same stimuli were also presented via the same Linux system to head-restrained monkeys who were seated 60 cm from a Mitsubishi XC2935C CRT monitor (71.5×53.5 cm; 640×480 pixels). This provided a usable field of view of $61.6^\circ \times 48.1^\circ$. Trial initiation was self-paced. Each video presentation was preceded by a fixation point and the next video was initiated when the monkey's eye position remained within a square electronic window with 5° radius of the central fixation point for 300–500 ms. The monkey subjects were not rewarded systematically for doing this task, but most monkey subjects easily learned to fixate in order to initiate the next clip.

Human eye-tracking procedure

Human eye movements were recorded using an infrared-video-based eye-tracker (ISCAN RK-464). Pupil and corneal reflections of the right eye were used to calculate gaze position with an accuracy 1° , sampled at 240 Hz. To calibrate the system, subjects were asked to fixate on a central point and then saccade to one of nine target locations distributed across the screen on a 3×3 grid. This procedure was repeated until each location was visited twice. In subsequent offline analysis, the endpoints of saccades to targets were used to perform an affine transform followed by a thin-plate-spline interpolation (Itti, 2005) on the eye position data obtained in the free-viewing experiment in order to yield accurate estimate of eye position given the geometry of the eye-tracker and



Figure 1. The six categories of scene types. Exemplars are shown from the six categories of scene types: (A) building and city, (B) natural, (C) sports, (D) indoor, (E) non-natural (cartoons, random noise, space), and (F) monkey relevant (monkeys, experimenters, facilities). Each group contains scenes with and without main actors (e.g., empty room vs. talk show). (G) An example of eye movement traces from 4 humans (blue) and 4 monkeys (green) superimposed on a video clip during a relatively stationary 3-s period. Notice that monkeys looked around the screen while humans focused their gaze on the slowly moving car in the background (inset with yellow box).

display. Recalibration was performed every 13 movie clips during the experiment.

Monkey eye-tracking procedure

A stainless steel head post was attached to the skull via an acrylic implant anchored to the skull by stainless steel screws. Eye coils were implanted between the conjunctiva and the sclera of each eye (Judge, Richmond, & Chu, 1980) allowing for precision recording of eye position using the magnetic search coil technique (Robinson, 1963). Surgical methods for preparing animals for head-fixed eye movement recordings have been described previously (Marino, Rodgers, Levy, & Munoz, 2008). Monkeys were seated in a primate chair with their heads restrained for the duration of an experiment (2–4 hours).

Eye position data were digitized at 1000 Hz using data acquisition hardware by Plexon. Concurrently, timestamps of the time of fixation point onset, acquisition of the fixation target by the monkey, and initiation of the clip were recorded.

To calibrate eye position, monkeys performed a step saccade paradigm in which targets at three eccentricities and eight radial orientations from the fixation point were presented in random order. Monkeys were given a liquid reward if they fixated a target within a square electronic window of 4° radius within 800 ms. During calibration, behavioral paradigms and visual displays were controlled by two Dell 8100 computers running UNIX-based real-time data control and presentation systems (Rex 6.1: Hays, Richmond, & Optican, 1982). In order to control for small non-linearities in the field coil, the weighted average of several visits to each target endpoint was later used to

perform an affine transform and thin-plate-spline interpolation on the eye position data collected during free viewing of the video clips.

Quantifying eye movement behavior

In order to quantify viewing behavior, an algorithm was used for both species, which parsed the analog eye position data into saccadic, fixational, and smooth pursuit eye movements. Traditional techniques to separate these various eye movements did not work well with these data, because many of the eye movement patterns elicited during free viewing of dynamic stimuli were non-traditional (e.g., blends of smooth pursuit, optokinetic, and saccadic eye movements). To deal with such idiosyncrasies, standard velocity measurements were combined with a simple windowed Principal Components Analysis (PCA). The eye position data were first smoothed (63 Hz Lowpass Butterworth), and eye positions with velocities greater than 30 deg/s were marked as possible saccades. Within a sliding window, the PCA was computed and the ratio of explained variances (minimum over maximum) for each of the two dimensions was stored. A ratio near zero indicates a straight line, and hence a likely saccade. The results of several different window sizes were linearly combined to produce a robust and smooth estimate. Eye positions with a ratio near zero but with insufficient velocity to be marked as a saccade were labeled as smooth pursuit. The remaining data were marked as fixation. Saccades with short (<80 ms) intervening fixations or smooth pursuits and small differences in saccadic direction (<45°) were assumed to represent readjustments of gaze en route to a target, and so were combined into a single saccadic eye movement toward the final target, rather than two or more separate saccades. Additionally, saccades of <2° in amplitude and <20 ms in duration were removed in order to decrease the false positive rate of saccade parsing and to focus analysis on eye movements that more likely reflected a shift of attention to a new target as opposed to minor gaze adjustments on a current target (Itti, 2005). This saccade parsing algorithm is freely available as part of the stimulus presentation software.

For each subject (human or monkey), clips that contained excessive durations (>30% of clip length) of tracking loss (blinks, loss of signal from search coil, or video-based tracker) or off-screen eye movements (sleeping, inattentive behavior) were excluded from the analysis. The majority of monkey clips was rejected for excessive off-screen eye position (18.6% of the monkey data, 0.7% for humans); 11.8% of the monkey data (1.4% for humans) were discarded for loss of tracking. In monkeys, the implanted search coil still produces a signal when a subject is in a blink, however, strain on the coil due its implanted position (along with other noise factors) will cause some loss of tracking. Due to technical errors, data were not recorded for 17 clips for 1 monkey and 2 clips for another, accounting for 3.3% of the monkey

data. In total, 1.9% of human and 27.3% of monkey eye traces were rejected. Note that the individual rejection percentages do not add to the total percentage rejected due to overlap between clips containing tracking loss and off-screen data. Analysis was consequently performed on different subsets of clips for each observer with the limitation that at least three observers from each species had to have successfully viewed each clip for it to be retained in the analysis.

Implementation of computational models

To assess the visually guided behavior of humans and monkeys, two validated computational models of visual attention (contrast and saliency) and an interobserver consistency metric were used to predict individual eye movements (Figure 2). Models were created and run under Linux using the iLab C++ Neuromorphic Vision Toolkit (Itti, 2004). First, a luminance contrast model (Reinagel & Zador, 1999), defined as the variance of pixel values in 16×16 pixel patches tiling the input image frame (Figure 2, left), is a simple but non-trivial model of attention and serves as a control for the performance of the saliency model. Second, we used the saliency model of visual attention framework (Figure 2, center; Itti & Koch, 2000; Itti et al., 1998). The Itti and Koch model computes salient locations by filtering the movie frames along several feature dimensions (color, intensity, orientation, flicker, and motion). Center-surround operations in each feature channel highlight locations that are different from their surroundings. Finally, the channels are normalized and linearly combined to produce a saliency map, which highlights screen locations likely to attract the attention of human or monkey observers. To process our video clips, we used the latest variant of the saliency model, which uses Bayesian learners to detect locations that are not only salient in space but are also salient (or so-called “surprising”) over time (Itti & Baldi, 2006). This model hence substantially differs from and generalizes other models of stimulus-driven attention (Itti et al., 1998; Le Meur et al., 2007; Privitera & Stark, 2000; Tatler et al., 2005) in that both spatial and temporal events within each feature map that violate locally accumulated “beliefs” about the input cause high output for that location.

The contrast model contains no temporal dynamics and, consequently, would not be expected to outperform the saliency model. Since many simple models would perform significantly above chance, we use the contrast model as a lower bound of performance for any non-trivial model of attention. Additionally, luminance contrast is correlated with many features used in the saliency computation. Comparing the static luminance contrast model with the saliency model gives some insight into the contribution of the dynamic features irrespective of luminance contrast.

To compute a measure of gaze agreement among and between species, an interobserver metric was created separately for each species using a leave-one-out approach

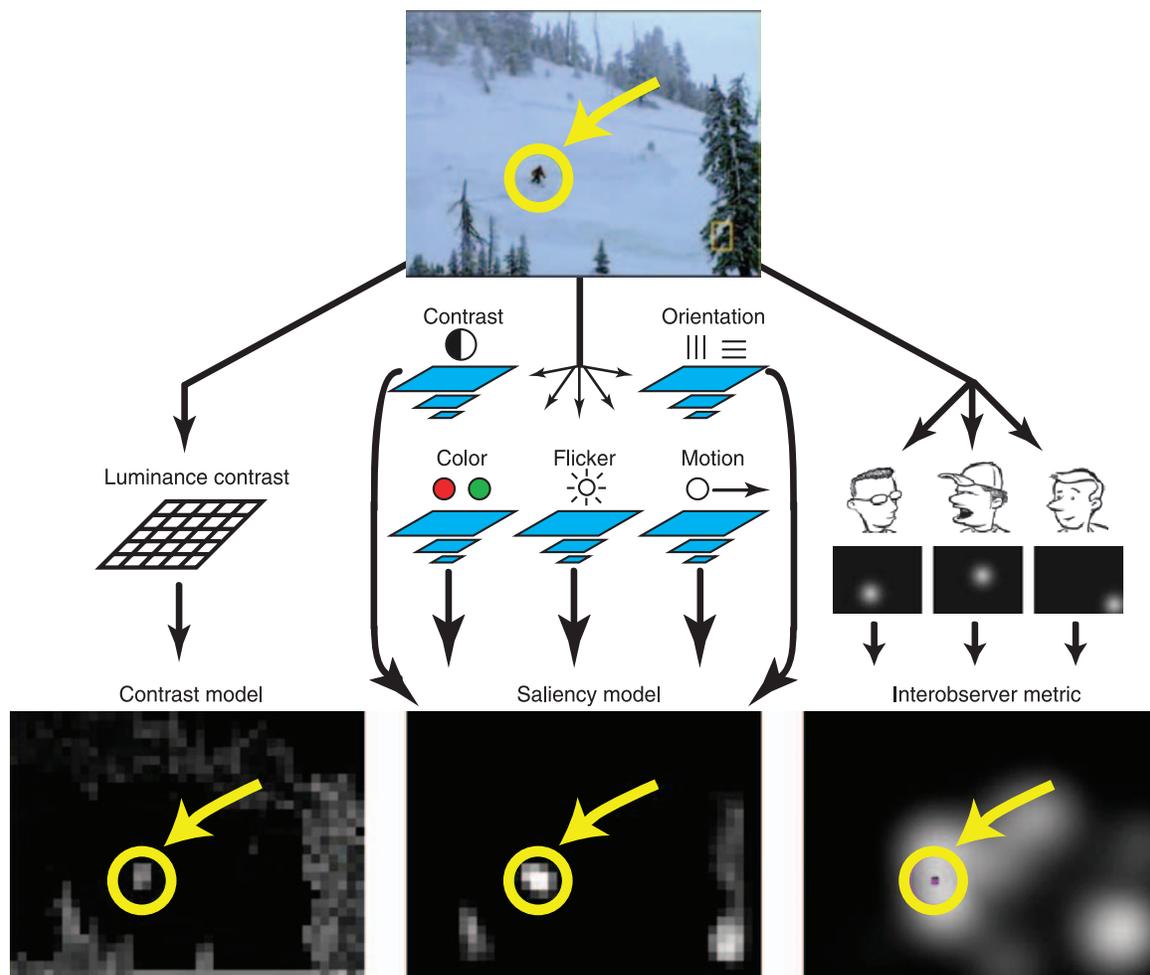


Figure 2. Architecture of the contrast and saliency models, and interobserver agreement metric. (Left) A simple luminance contrast model computed as the variance of luminance values in 16×16 pixel image patches. (Center) The latest implementation of the saliency model (Itti & Baldi, 2006). (Right) An interobserver agreement metric (see [Methods](#) section) created by making a heat map from the pooled eye movements of all observers, except the one under test, on a given movie clip (leave-one-out analysis). The yellow circle indicates the endpoint of a saccadic eye movement. At the start of the saccade, the maximum value within a 48-pixel radius circular aperture was stored along with 100 values chosen randomly from the saccadic endpoint distribution of all clips and subjects except for the one under test. To test for agreement between or among species, the interobserver agreement metric was sampled at the time when the eye landed at its target.

(Figure 2, right). A master map is created by placing Gaussian blobs ($\sigma = 48$ pixels) centered at the instantaneous eye positions of a subset of human or monkey observers. For each subject, a map is created from the eye positions of the 2–4 other subjects in the same species who viewed the clip. A maximum output for this map is achieved when all subjects look at the same item simultaneously. This map represents a combination of stimulus-driven and goal-directed eye movements and has been used as an upper bound for human gaze prediction (Itti & Baldi, 2006).

Comparing eye movements to model and metric output

To compute the performance of each model or metric, the maximum map values in a circular window (3.6° humans,

4.7° monkeys: A 48-pixel window but different viewing distances and screen sizes for each species) around human or monkey saccadic endpoints were compared to 100 map values collected from locations randomly chosen from the distribution of saccadic endpoints from all saccades (in the same species) except those generated in the same clip by the same subject as the sample. This approach is similar to the image-shuffled analysis method used by others for static images (Parkhurst & Niebur, 2003; Reinagel & Zador, 1999; Tatler et al., 2005) and allows for an unbiased measure of model performance despite any accidental correlation between a particular species saccadic endpoint distribution and model output. For a particular subject, at the onset of a saccade we measured the value in each model map at the endpoint of the saccade, i.e., the activity in the map just before the saccade. For the interobserver model, the map value was measured at the time of the

endpoint of the saccade to assess the congruency of gaze locations, either within or between species.

Differences between saliency at human or monkey gaze targets and at the randomly selected locations were quantified using ordinal dominance analysis (Bamber, 1975). Model or metric map values at observers' saccadic endpoints and random locations were first normalized by the maximum value in the map when the saccade occurred (i.e., when the map was sampled). For each model, histograms of values at eye positions and random locations were created. To non-parametrically measure differences between observer and random histograms, a threshold was incremented from 0 to 1, and at each threshold value we tallied the percentage of eye positions and random locations that contained a value greater than the threshold (“hits”). A rotated ordinal dominance curve (similar to a receiver operating characteristic graph) was created with “observer hits” on one axis and “random hits” on the other (Figure 5, inset). The curve summarizes how well a binary decision rule based on thresholding the map values could discriminate signal (map values at observer eye positions) from noise (random map values). The overall performance can be summarized by the area under this curve. This value is calculated and stored for each of the 100 randomly sampled sets. The mean of the 100 ordinal dominance values is taken as the final ordinal dominance estimate. A model that is no more predictive than chance would have equal random and model hits for each threshold, creating a straight line with an ordinal dominance of 0.5. The interobserver metric is assumed to provide the upper bound of predictability, between 0.5 and 1.0 (see Results section), which the best computational models might be expected to approach. Note that an ordinal dominance of 1.0 is not achievable by any model, because there is imperfect agreement among observers, hence it is impossible for a single model to exactly pinpoint the gaze location of each observer.

“High-interest” gaze targets

For some analyses, we defined a subset of saccadic endpoints as “high-interest” gaze targets. These were locations separated by less than 48 pixels (3.6° humans, 4.7° monkeys) that two or more observers of a given species looked at within 150 ms of one another. For monkeys, filtering the 12,826 saccades used for the overall analysis by these criteria resulted in a subset of 1,812 saccades; for humans, filtering the original 12,148 saccades resulted in a subset of 4,142 saccades.

Statistical analysis

Distributions of model and metric output at gaze targets were statistically compared using the permutation

framework (Monte Carlo simulation, 10,000 repetitions). Confidence intervals for model and metric scores were estimated by repeating the ordinal dominance measurement on a randomly selected half of the data, to form a sampling interval. Tests between species or models were carried out using a permutation test, computed by taking all saccades from both groups under test and randomly assigning each saccade to one of the two groups, irrespective of the actual group membership of the saccades. The difference between mean ordinal dominance values for the two randomly assigned groups was computed and stored. The process was repeated to form a sampling interval. The p value represents the probability of observing a value more extreme than the original group assignment (Good, 1994). Statistical analysis of the saccadic endpoint distributions was also carried out in the permutation framework, but the symmetric Kullback–Leibler distance function was used in place of ordinal dominance.

Results

Saccade metrics

Several differences in the saccade metrics of humans and monkeys were observed. Figures 3A and 3B show the smoothed distribution of saccadic endpoints used for analysis. “Hotter” colors represent a higher likelihood that a subject made a gaze shift to that location. Human and monkey saccadic endpoint distributions were significantly different (permutation test, $p \leq 0.0001$), but both species showed the characteristic center bias reported in human experiments using natural photographs (Reinagel & Zador, 1999; Tatler, 2007). This may reflect a physiological bias to return the eyes to the center of the orbits (Paré & Munoz, 2001). Monkeys seemed to explore the spatial extent of the display more thoroughly than humans, who were very center-biased. This difference may be due to a variety of factors including motor differences, cognitive awareness of the main actors and actions that were often near the center of the video, or a general search strategy. The TV channel logo that often appeared in the lower right-hand corner (Figure 1G) also attracted a high number of gaze shifts for both species.

Figures 3C and 3D show the saccadic main sequence for humans and monkeys. The main sequence plots the relationship between saccadic peak velocity and amplitude and is well known to be an exponential function (Bahill, Clark, & Stark, 1975). The shape of this function is thought to reflect the brainstem circuitry controlling saccades and is altered when there is damage in the brainstem circuits or muscles controlling saccades (Ramat, Leigh, Zee, & Optican, 2007). The main sequence data combined across the 5 monkeys were noticeably more variable than the human main sequence. When analyzed

on a log–log scale, a linear regression revealed an R^2 of 0.77 (ANOVA test, $F(1, 15168) = 52002$, $p < 0.0001$) for monkeys compared to an R^2 of 0.96 (ANOVA test, $F(1, 14835) = 342150$, $p < 0.0001$) for humans. Monkeys were much faster for a given amplitude, and regression lines showed monkeys had significantly higher velocity offset (Figure 3). The slope of the line was significantly higher

(Figure 3, although a small magnitude difference) in humans, indicating a steeper relationship between amplitude and peak velocity.

Figure 4 compares saccadic amplitude, fixation duration, and intersaccadic interval distributions for monkeys (green bars) and humans (blue bars). The probability distribution of saccadic amplitudes differed significantly,

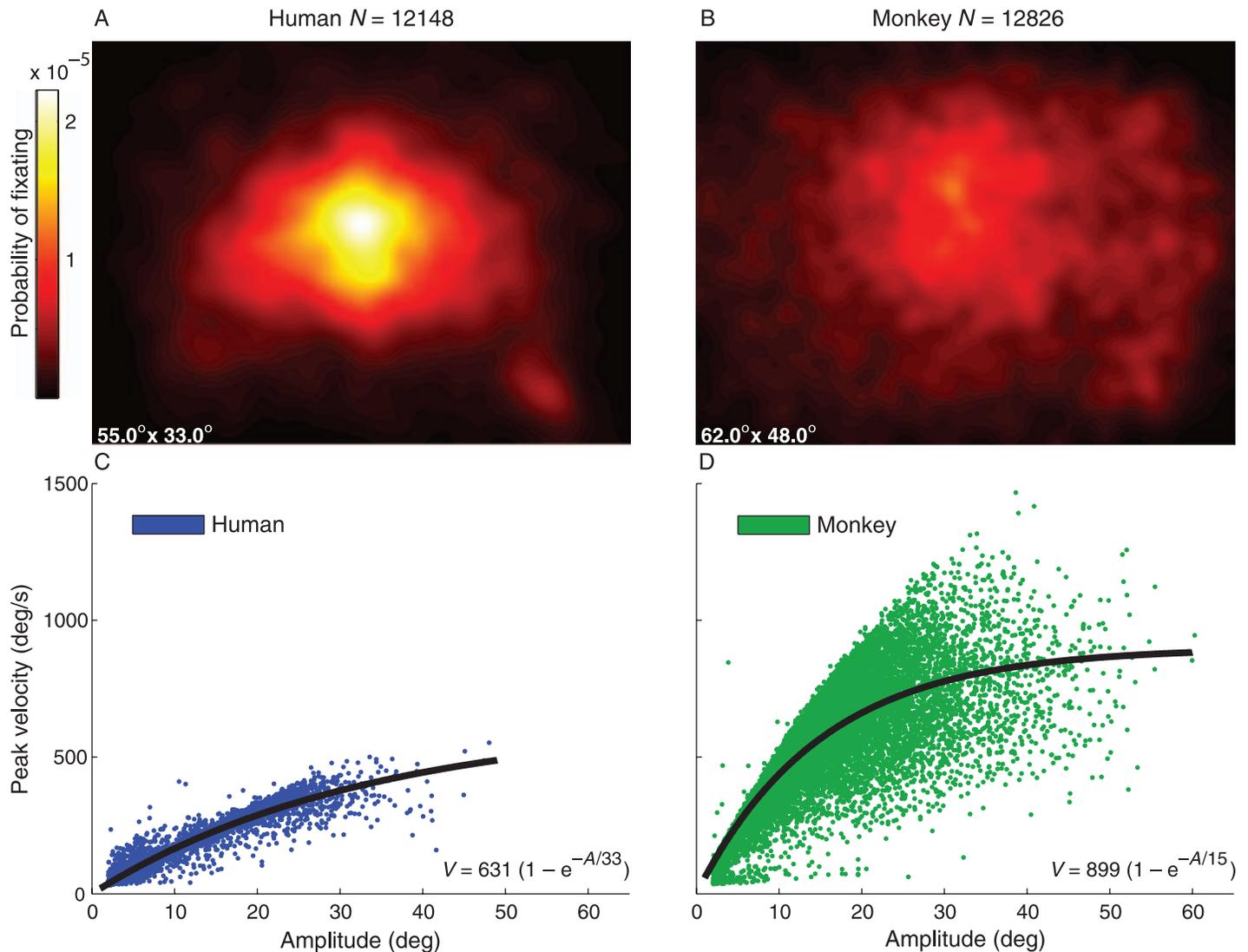


Figure 3. Saccade Metrics: Endpoint distributions and main sequences. (A–B) Saccadic endpoint distributions for the 12,138 human and 12,832 monkey saccades (computed after removing noisy data and clips with fewer than three observers, resulting in less data for monkeys) used for comparison with the contrast and saliency models and the interobserver agreement metric. Points were smoothed by convolving each map with a Gaussian kernel ($\sigma = 1.5^\circ$). “Hotter” colors represent a higher likelihood that a human or monkey gaze shift landed at that screen location. Distributions were significantly different at $p \leq 0.0001$, using the Kullback–Leibler distance function between distributions in a permutation test (see Methods section). (C–D) Main sequence for all saccades (14,837 from humans and 15,170 from monkeys, before removing clips with fewer than three observers) recorded from humans (blue) and monkeys (green). The main sequence was computed before combining multi-step saccadic eye movements into a single saccade, yielding separate entries for each component of the multi-step saccade. Main sequences for humans and monkeys were significantly different (ANOVA test, $F(2,30003) = 58024.55$, $p < 0.0001$), testing for coincident regression lines on a log–log scale. Significant differences were observed for both the slope (ANOVA test, $F(1,30003) = 1703.29$, $p < 0.0001$) and velocity offset (ANOVA test, $F(1,30003) = 21805.25$, $p < 0.0001$) components of the main sequence. Black lines fitted to the data were computed by minimizing $V = a(1 - e^{-A/s})$, where V and A are saccadic velocities and amplitudes, respectively; a and s are the model parameters representing maximum amplitude and slope of the lines.

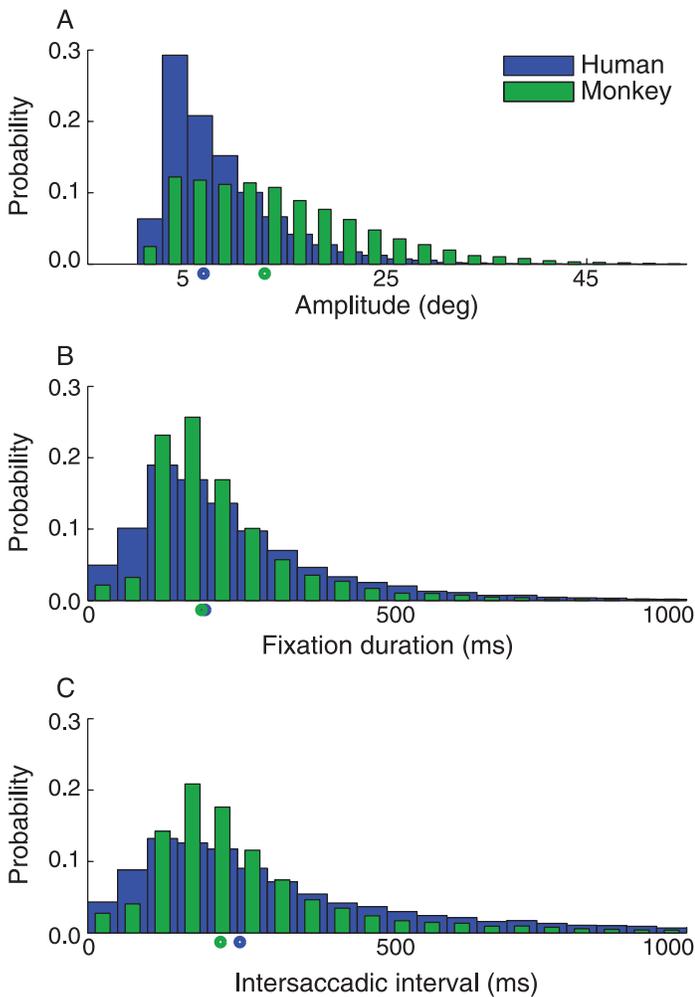


Figure 4. Saccade Metrics: Distributions of saccade amplitude, fixation durations, and intersaccadic intervals. Probability histograms for (A) saccadic amplitude, (B) fixation duration after a saccade, and (C) intersaccadic interval (which may include smooth pursuit) for humans (blue) and monkeys (green) calculated before combining multi-step saccades into a single saccade. For display purposes only, the green bars are half the width of the blue bars, which represent the actual interval for both. The time axes are truncated at 1000 ms. Amplitude (Two-tailed Kolmogorov–Smirnov, $D = 0.34$, $n_1 = 14837$, $n_2 = 15170$, $p < 0.0001$), fixation duration (Two-tailed Kolmogorov–Smirnov, $D = 0.12$, $n_1 = 14837$, $n_2 = 15170$, $p < 0.0001$), and intersaccadic interval (Two-tailed Kolmogorov–Smirnov, $D = 0.13$, $n_1 = 14837$, $n_2 = 15170$, $p < 0.0001$) histograms were significantly different. Green and blue circles represent the median scores for each species.

in that monkeys had a broader distribution and a greater median (Figure 4A). This could in part have been because monkey subjects had a slightly wider field of view; however, when amplitudes were replotted on a normalized axis, the same qualitative results were obtained (not shown). The probability distributions for fixation durations and intersaccadic intervals also significantly differed between species with humans having slightly longer

median durations (Figures 4B and 4C). Monkey fixation and intersaccadic interval distributions were narrower, which possibly indicates a stereotyped fixation pattern (e.g., fixate for 250 ms and then saccade to new place). In contrast, human fixation durations and intersaccadic intervals were spread over a wide range of values.

Model predictions of gaze shift endpoints

To further quantify species differences, we used a computational model of saliency-based visual attention. In previous human experiments, this model has revealed that observers gaze more frequently toward the salient “hot-spots” computed by the model in both static images and dynamic scenes (Itti, 2005; Itti, 2006; Itti & Baldi, 2006; Parkhurst et al., 2002; Peters et al., 2005). The model takes as input an image or video clip frame and outputs a saliency map that gives a prediction of the screen locations likely to attract attention. The specific implementation details of this model have been described previously (Itti & Baldi, 2006; Itti et al., 1998).

We measured the amount of computed saliency for each video frame at the endpoints of saccadic eye movements in both species (see Methods section), to assess the extent to which humans and monkeys exhibited similar computations of saliency (perhaps represented in monkey LIP, Goldberg, Bisley, Powell, & Gottlieb, 2006) and strategies for deploying gaze toward salient locations. To quantify the chance-corrected performance of the saliency model, values at gaze targets were compared to values at gaze targets taken at random from other video clips, giving an ordinal dominance score (see Methods section). Measurements from the contrast model and interobserver agreement metric were similarly chance-adjusted. Figure 5A shows the comparison of human and monkey ordinal dominance scores for different models and metrics, and Figure 5B shows a summary of the statistical analysis. All models and metrics predicted human and monkey gaze targets significantly better than chance (permutation test, $p \leq 0.0001$), and saliency predicted human and monkey gaze behavior significantly better than the baseline-control contrast model (permutation test, $p \leq 0.0001$). This finding validated the use of the saliency model as a good predictor of visually guided attentive behavior in both humans and monkeys.

Interestingly, we found that saliency correlated with human behavior significantly better than monkey behavior, over all clips combined (permutation test, $p \leq 0.0001$). Differences in the likelihood to deploy attention to salient items should be minimized when using monkeys as a model for human attention during free viewing. The saliency differences were, however, small in magnitude compared to the difference in interobserver agreement (Figure 5). Comparing saliency scores with interobserver agreement may provide insight into a way to reconcile such differences. Although saliency was a strong predictor of human visually guided behavior, the stimulus-driven nature

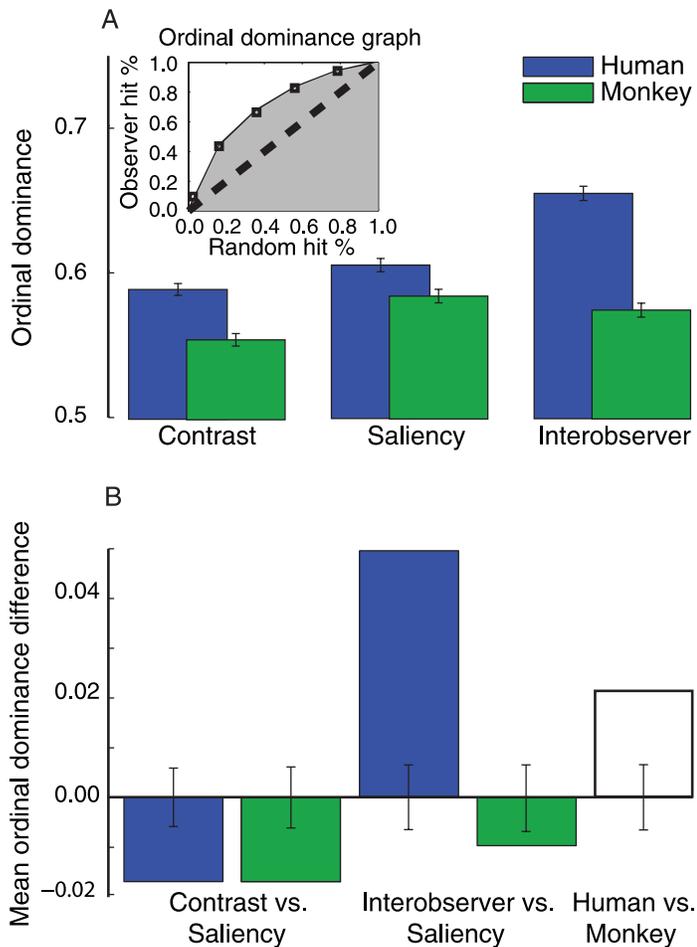


Figure 5. Model and metric scores at human and monkey saccadic endpoints. (A) Comparison of the contrast and saliency model, and interobserver agreement metric values at human (blue) and monkey (green) saccadic endpoint locations with values at randomly selected eye positions. Overall, human and monkey gaze shifts were predicted (permutation test, $p \leq 0.0001$) by all models and metrics greater than chance levels (ordinal dominance of 0.5). Error bars show the 95% confidence interval on the ordinal dominance estimate (see [Methods](#) section). (B) This figure summarizes the statistical differences between species and models as obtained through permutation tests (see [Methods](#) section). Blue (human), green (monkey), and white (human–monkey) bars show the magnitude of the test statistic (mean ordinal dominance difference) obtained between pairs labeled on the x-axis. Values greater than 0 indicate that the first model or species in the pair had a larger ordinal dominance score. Black bars represent the 95% confidence interval of the test statistics sampling distribution. (Left) Saliency performed better than the baseline-control contrast model for both humans and monkeys (permutation test, $p \leq 0.0001$). (Center) Interobserver agreement was more predictive than saliency for humans (permutation test, $p \leq 0.0001$), however, interobserver agreement was less predictive than saliency for monkeys (permutation test, $p = 0.0027$). (Right) The human saliency ordinal dominance score was significantly higher than the monkey score (permutation test, $p \leq 0.0001$).

of the model limited its predictive power. The interobserver agreement metric captured aspects of stimulus-driven (saliency) and top-down (context specific) attentional allocation, the latter of which has also been shown to be a significant factor in guiding human gaze shifts in natural scenes (De Graef, De Troy, & Dydewalle, 1992; Neider & Zelinski, 2006; Noton & Stark, 1971; Oliva, Torralba, Castelhana, & Henderson, 2003; Yarus, 1967). The interobserver agreement metric was the best predictor of human saccadic targets (permutation test, $p \leq 0.0001$). Interestingly, this trend did not hold for monkeys and the interobserver agreement metric was significantly less correlated with monkey gaze shifts than the saliency model (permutation test, $p = 0.0027$). That is, the computational saliency model better predicted where one monkey might look than was predicted from the gaze patterns of two to four other monkeys. Any top-down information present in the monkey interobserver agreement metric was insufficient to increase predictability of gaze patterns over a purely stimulus-driven model. Monkey top-down attentional allocation may be completely inconsistent among observers (e.g., [Figure 1G](#)), leaving saliency to be the best predictor of visually guided attentive behavior.

[Figure 6](#) shows a scatter plot of median normalized (not chance corrected) monkey vs. human saliency values at all saccadic endpoints that occurred during each entire clip. This clip-by-clip analysis revealed that saliency values from monkeys and humans were significantly correlated ([Figure 6](#)). The best fitting line (solid black) had significantly lower slope than the unity line (dashed black), indicating that monkeys' saliency scores varied less than those of humans from clip to clip, and clips that contained higher saliency values for humans contained on average slightly lower saliency values for monkeys. The y-offset, however, was not different from 0 ([Figure 6](#)), indicating that there was no systematic bias, or baseline shift, in human or monkey raw saliency scores. The majority of the regression line falls below the unity line; hence, on average the saliency scores were lower for monkeys, as was already the case with our aggregate analysis ([Figure 5](#)). Individual clip content affected deployment of gaze to salient locations for humans and monkeys in a comparable way; however, monkeys may have had a tendency to be less modulated by clip content. This likely reflects differences in semantic understanding of the clips between the two species.

We defined a subset of clips ([Figure 1F](#)) as monkey relevant. These clips contained scenes from the monkeys' daily environment (e.g., their housing, familiar monkeys and humans, facilities) and represented a contextual control to ensure monkeys attended to familiar natural scenes similarly to novel ones. The points in the scatter plot for monkey-relevant clips ([Figure 6](#), green triangles) were in the same distribution as those for other clips. Only considering these monkey-relevant clips, a significant linear correlation was found (Pearson correlation, $r(13) = 0.72$, $p = 0.005$). This line was not significantly different

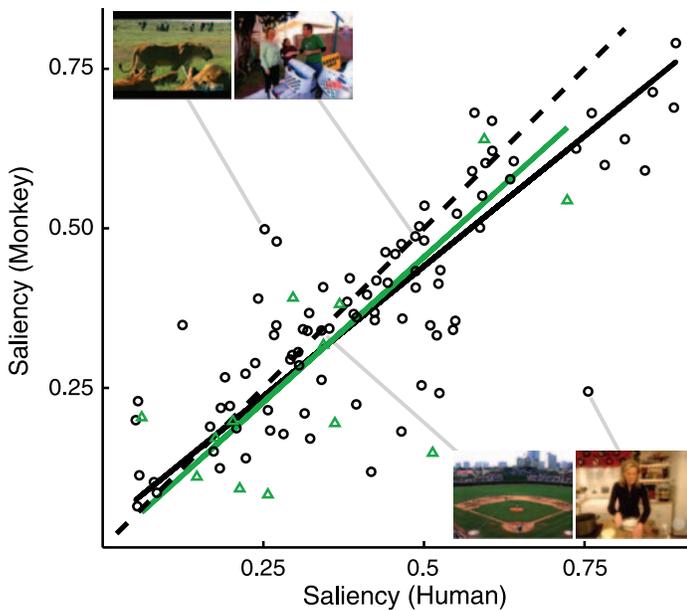


Figure 6. Correlation between saliency values at human and monkey eye positions. The scatter plot shows median saliency values considering all saccadic endpoints in a given video clip for monkeys vs. humans. Each point represents the median of raw (not chance corrected) saliency values for each video clip, with green triangles indicating clips that would be relevant to a monkey as described in the [Methods](#) section. Human and monkey scores were well correlated (Pearson correlation, $r(98) = 0.80$, $p < 0.0001$). Analysis of coefficients obtained by major axis regression (Sokal & Rohlf, 1995) revealed that the best fitting line ($y = 0.82x + 0.032$, solid black) was significantly different from unity (dotted black) in slope (F -test, $F(1,98) = 7.26$, $p = 0.0083$) but not y -offset (t -test, $t(98) = 0.089$, $p = 0.38$). The regression line for monkey-relevant clips ($y = 0.91x - 0.00021$, solid green) was not significantly different from the regression line for all other clips (chi-square test, $\chi^2(1, N = 100) = 0.2$, $p = 0.65$), computed by testing for coincident lines. Hypothesis testing was performed according to Warton, Wright, Falster, and Westoby (2006). The sample frames in the upper left and lower right corners are from videos where one species had a considerably higher saliency score than the other. The two adjacent frames are from the two videos where human and monkey scores were most similar.

from that calculated for all other clips (Figure 6). Taken together, this analysis indicates that monkeys were visually attentive to the video clips in a similar fashion to humans, at least as far as saliency is concerned, although from this analysis we cannot know if they looked at similar spatial locations at the same time, only that they looked at similarly salient items.

“High-interest” gaze locations

We wondered if the relatively poor predictability of monkey behavior by the saliency model and interobserver

agreement metric might be due to idiosyncratic search strategies and/or cognitive systems by monkeys, which may or may not have been related to the video content. To remove idiosyncratic gaze shifts from the analysis, we determined a subset of “high-interest” gaze targets—those locations that attracted the attention of two or more observers toward the same location at the same time (see [Methods](#) section). Saliency and interobserver agreement metrics were then reanalyzed based on this subset for each species. Figure 7A shows the effect of filtering by high-interest gaze targets on an “interspecies agreement” metric. This metric represents the correlation between monkey saccadic target locations and those target locations selected by humans. This metric was computed by testing monkey saccadic endpoints against the same human-derived interobserver metric that was used for human interobserver agreement analysis. The interspecies agreement metric allowed us to directly measure the extent to which monkey gaze target locations were also looked at by humans. The lowest score the interobserver agreement metric obtained for humans was when all human saccades were analyzed together (Figures 5A and 7A, lower black line). This can serve as a lower bound for our interspecies agreement metric; as to be a good model of human visual behavior, monkeys should be as consistent with human gaze targets as humans are with one another. A useful upper bound for this metric is obtained by recalculating the interobserver agreement metric for saccadic target locations where at least two humans agreed to look (Figure 7A, upper black line). We expect the best models of human visual behavior (animal or computational) to approach this level of correlation with humans, as it means the model is often selecting the strong attractors of attention—those scene locations that on average attracted the attention of multiple human observers.

When all monkey saccades were considered, the interspecies ordinal dominance score was lower than the score obtained from the human interobserver agreement metric (permutation test, $p \leq 0.0001$). That is, monkey saccadic target selection was less consistent with human target selection than humans were with one another. However, the interspecies ordinal dominance score dramatically increased (permutation test, $p \leq 0.0001$) when analysis was limited to monkey saccades made toward monkey high-interest targets. In fact, the interspecies score for these high-interest monkey saccades fell above our human-derived lower bound (permutation test, $p \leq 0.0001$) but below our human-derived upper bound (permutation test, $p \leq 0.0001$). This demonstrates a high correlation between locations where humans and monkeys looked when analysis of monkey saccades was restricted to high-interest locations.

Figure 7B compares human and monkey saliency ordinal dominance scores for all gaze targets and high-interest gaze targets. As was shown in Figure 5, when all saccades were considered the monkey saliency ordinal

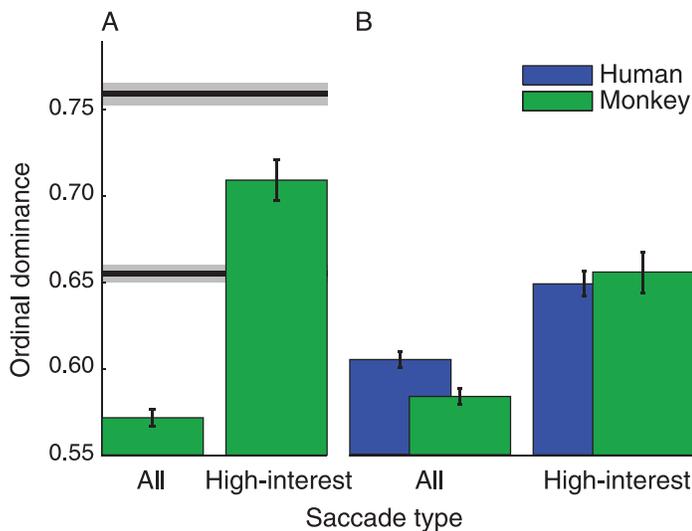


Figure 7. Analysis at high-interest gaze locations. To test the agreement in saccadic target selection between humans and monkeys, the human interobserver metric was used to predict the gaze locations of monkeys (interspecies agreement metric). (A) Ordinal dominance scores for the interspecies agreement metric for all monkey saccadic endpoints, and a subset of “high-interest” saccadic targets, that multiple monkeys looked at simultaneously. When only high-interest targets were considered, monkey saccadic endpoints were closer to human gaze locations (permutation test, $p \leq 0.0001$). To serve as a reference, the lower black line is the mean ordinal dominance score of the human interobserver agreement metric. The upper black line is the mean ordinal dominance score of the human interobserver agreement metric when only locations where two or more humans agreed to look were considered. Shaded regions represent the 95% confidence intervals of these estimates. When all monkey gaze targets were considered, the interspecies agreement metric scored lower than the human interobserver agreement metric (permutation test, $p \leq 0.0001$). However, when only high-interest gaze targets were considered, the interspecies ordinal dominance score fell between the lower and upper bounds derived from our human interobserver metric (permutation test, $p \leq 0.0001$). (B) Saliency ordinal dominance scores for all gaze endpoints and a subset of high-interest gaze locations for humans and monkeys. The ordinal dominance scores for all saccades (Figure 5) are replotted as a reference. When all monkey gaze targets were considered, the monkey saliency ordinal dominance score was lower than the human score (permutation test, $p \leq 0.0001$). For the subset of high-interest gaze targets, where two or more monkeys agreed, the ordinal dominance score was increased (permutation test, $p \leq 0.0001$) and indistinguishable from the human high-interest gaze targets (permutation test, $p = 0.16$), putting the monkeys in the range of human predictability.

dominance score was significantly lower than the human score, indicating that the saliency model predicted human saccades better than monkey saccades. However, at high-interest gaze targets, the ordinal dominance scores were significantly higher for humans and monkeys (permutation test, $p \leq 0.0001$), indicating that the saliency model was a

better predictor of high-interest gaze targets than of low-interest ones (e.g., when the five observers looked at five different locations) for both species. Note that increasing the number of humans who agreed on a saccadic target to three did not significantly increase the saliency ordinal dominance score (not shown). Thus, in our analysis, gaze locations where two human observers agreed can serve as an upper bound for human gaze predictability. Increasing the number of agreeing monkeys beyond two seemed to increase the ordinal dominance scores linearly (not shown), but more data would be required for hypothesis testing. Interestingly, the saliency ordinal dominance score for monkey high-interest saccadic targets was greater than the human score for all saccades (permutation test, $p \leq 0.0001$) and was indistinguishable from the score for human high-interest gaze targets (permutation test, $p = 0.16$). That is, scene items that drew the attention of multiple monkeys (high-interest gaze targets) contained similar chance-corrected saliency values than those locations that attracted the gaze of multiple humans.

Discussion

The present study objectively compared, for the first time, human and monkey visually attentive behaviors during free viewing of natural dynamic (video) stimuli. In addition to examining saccadic eye movement metrics, several models of visual attention were employed to provide objective metrics by which to compare human and monkey viewing behaviors. We found significant differences between human and monkey gaze shifts during free viewing. In summary, monkeys generated faster saccades, which spanned a greater range of the screen and were separated by shorter fixation durations. Although both species shifted gaze to locations that were deemed salient by the saliency model, humans were more likely to do so. The gaze locations of other humans were the best predictors of human behavior, but this was not true of monkeys. The saliency model predicted monkey gaze shifts better than the combined gaze behavior of other monkeys. These differences, however, could be minimized if we only examined high-interest gaze locations—those that at least two monkeys jointly attended. When the saccades were filtered in this way, monkey behavior became more human like, almost indistinguishable in terms of gaze location and saliency values. This filtering technique focuses analysis on common attractors of attention between species, possibly by emphasizing the role of the shared low-level saccadic selection processes over the more idiosyncratic cognitive processes. High-interest targets minimize differences between the species, providing a method to make the best use of monkeys as a model of human visual behavior under free-viewing conditions.

Monkey–human differences in eye movement metrics

Eye movement metrics under free viewing of video stimuli were found to be quite different between monkeys and humans. Monkeys were less center-biased and made saccades with larger amplitudes on average. This may suggest that monkeys were less interested in the video actions and actors, which tended to be filmed near the center. Monkeys may have had less cognitive understanding of the scenes, and/or they were more interested in exploring the screen, possibly in search of actions/locations that could have resulted in reward.

At a more mechanical level, monkeys differed from humans in features of their saccadic main sequence (saccadic velocity vs. amplitude). Monkeys made much faster saccades for a given amplitude compared to humans, confirming what has been found by Harris, Wallman, and Scudder (1990). The main sequences under free-viewing conditions were comparable to those obtained in previous studies using laboratory stimuli with humans (Bahill, Brockenbrough, & Troost, 1981; Bahill et al., 1975; Becker & Fuchs, 1969; Boghen, Troost, Daroff, Dell’Osso, & Birkett, 1974) and monkeys (Quaia, Paré, Wurtz, & Optican, 2000; Van Gisbergen, Robinson, & Gielen, 1981) separately. Our data tended to have slower peak velocities, particularly in humans; however, velocities still fell within the normal range defined by Boghen et al. (1974). Differences in our data may be a feature of free viewing, or idiosyncratic to our subjects and methodology.

Discrepancies between species could be partly accounted for by differences in neural connectivity from the retina through the oculomotor system to the eye muscles, and possibly by differences in the motor plant, e.g., smaller viscous reactive forces in monkeys because they have a smaller eyeball. These plant differences probably reflect little on the processes involved in the deployment of visual attention. However, some discrepancies (e.g., intersaccadic intervals, saccadic endpoint distributions) may stem from different scanning strategies employed and should be accounted for when comparing species.

Monkey–human differences in model correspondence and interobserver agreement

More relevant to understanding visual attention is an examination of image properties at human and monkey gaze positions. To objectively compare species, we examined how computational models predicted saccadic targets of humans and monkeys. We used a model that measures static luminance contrast, which has been shown to be an attractor of gaze in humans and monkeys watching grayscale images (Einhäuser et al., 2006), and a saliency model, which has been shown to capture aspects of stimulus-driven eye movements in humans

viewing images (Peters et al., 2005) and videos (Itti, 2005; Itti & Baldi, 2006). The contrast model, although it does not contain temporal dynamics, serves as a baseline to measure the performance of the saliency model, for even simple models of attention will predict behavior significantly above chance (random sampling). Both models predicted gaze shifts of both species above chance, but the saliency model performed better, as expected. Validation of the saliency model with monkeys suggests that the species may possess similar computations of saliency during free viewing, and the model captures aspects of these mechanisms shared among primates. This is encouraging as it validates investigation of the neural substrates of such computations in monkeys.

Interestingly, the computational models predicted human gaze shifts better than monkey gaze shifts. This was surprising, as we had expected monkeys would be more saliency-driven than humans, due to their impoverished knowledge of the clips content (e.g., one video clip shows the earth viewed from space, likely a foreign concept to our monkeys). Our finding was also in contrast to results from Einhäuser et al. (2006) who found monkeys and humans to be equally saliency-driven to grayscale images. However, inconsistency in gaze target selection among monkey observers relative to humans provided some insight into these discrepancies.

Human attention has been described as a combination of stimulus-driven (bottom-up) and contextually driven or goal-directed (top-down) factors (Itti & Koch, 2001; Treisman & Gelade, 1980), and monkey attention is likely controlled by similar mechanisms (Fecteau & Munoz, 2006). The interobserver agreement metric contains elements of both factors while the saliency algorithm captures aspects of bottom-up processing *only*. As expected, for humans, the interobserver agreement metric provided the best prediction of gaze deployment. It has been known since Henderson, Weeks, and Hollingworth (1999) and Loftus and Mackworth (1978) that gaze density among observers is increased over scene regions containing semantically inconsistent or highly informative objects. Hence, the gaze consistency among our humans likely reflects their shared notion of semantically informative regions in the clips. Monkey gaze, however, was best predicted by the saliency model. This suggests that monkeys made many idiosyncratic eye movements, possibly related to each monkey’s unique interpretation of the scene, the goal of the experiment, or inattentiveness to the stimuli. Monkeys may have been engaged by the clips but shared less top-down knowledge of how to follow the main actions compared with humans. Alternatively, it may be that as a result of their training, monkeys were in part examining the screen looking to “unlock the task” or find a screen location or action that would lead to a reward. Such a search strategy is supported by the stereotyped fixational pattern (more narrow distribution of intersaccadic intervals). In either case, since their top-down interpretation seems inconsistent, saliency-based

computations may serve as the lowest common denominator in deploying gaze in natural scenes for monkey observers.

High-interest image locations minimizes monkey–human differences

Perhaps the most relevant question to consider, given the observed differences, is to what degree monkeys looked at the same places that humans looked. To address this, we focused analysis on high-interest targets, those locations that were gazed at by two or more monkeys simultaneously. This effectively forced consistency on our monkey data by filtering out some idiosyncratic eye movements that may have been due to differences in top-down scene interpretation or general attentiveness to the stimuli. An interspecies agreement metric revealed that when all saccade data were used, monkey saccadic targets were not as consistent with humans, as humans were with each other. In other words, monkeys did not often look where humans looked. This is not unexpected, as monkeys were inconsistent with each other. However, when the analysis was repeated using only the subset of monkey high-interest saccadic targets, those targets were dramatically closer to locations where, on average, humans looked (Figure 7A). High-interest gaze targets for monkeys became consistent with human visual behavior and were within the expected range of human interobserver agreement scores. These saccadic targets may focus our monkey analysis on scene locations that were of common interest to both species, narrowing the gap between human and monkey visual behaviors during free viewing of dynamic scenes.

Interestingly, those same high-interest targets that correlated well with human behavior were also highly salient; in fact, indistinguishable from human high-interest gaze targets in terms of their chance-corrected saliency scores. Highly salient items, as predicted by our model, may have simultaneously attracted the attention of multiple monkey observers. Since the monkey high-interest targets are also close to human gaze targets, this may indicate that saliency was the common factor in driving human and monkey attention to those locations. Analysis of monkey high-interest saccades minimized species differences both in terms of specific saccadic targets and saliency model agreement. This analysis may emphasize the shared bottom-up attentional processes among humans and monkeys, filtering out the more individualized cognitive processes.

This result may be particularly relevant when using monkeys in experiments requiring neural recording or imaging during free viewing of dynamic or natural scenes. Restricting analysis of neural responses to stimuli that attracted the gaze of at least two monkeys would ensure that the monkeys' behavior would be as consistent as

possible with human behavior under such conditions. While doing so eliminates a significant portion of the data, more data can be collected more easily under free viewing compared with traditional single-trial methods. This technique may emphasize common attentional mechanisms between species, thus making the best use of our animal model to generate results meaningful to human behavior and cognition.

Acknowledgments

This work was supported by NSF (CRCNS), Canadian Institutes of Health Research, Canada Research Chair Program, NGA, DARPA, and ARO. The authors affirm that the views expressed herein are solely their own and do not represent the views of the United States government or any agency thereof.

Commercial relationships: none.

Corresponding author: Laurent Itti.

Email: itti@usc.edu.

Address: 3641 Watt Way, HNB-30A, Los Angeles, CA 90089-2520, USA.

References

- Bahill, A. T., Brockenbrough, A., & Troost, B. T. (1981). Variability and development of a normative data base for saccadic eye movements. *Investigative Ophthalmology & Visual Science*, *21*, 116–125. [[PubMed](#)] [[Article](#)]
- Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*, 191.
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, *12*, 387–415.
- Becker, W., & Fuchs, A. F. (1969). Further properties of the human saccadic system: Eye movements and correction saccades with and without visual fixation points. *Vision Research*, *9*, 1247–1258. [[PubMed](#)]
- Boghen, D., Troost, B. T., Daroff, R. B., Dell'Osso, L. F., & Birkett, J. E. (1974). Velocity characteristics of normal human saccades. *Investigative Ophthalmology*, *13*, 619–623. [[PubMed](#)] [[Article](#)]
- De Graef, P., De Troy, A., & Dydewalle, G. (1992). Local and global contextual constraints on the identification of objects in scenes. *Canadian Journal of Psychology*, *46*, 489–508. [[PubMed](#)]

- Dragoi, V., & Sur, M. (2006). Image structure at the center of gaze during free viewing. *Journal of Cognitive Neuroscience*, *18*, 737–748. [[PubMed](#)]
- Einhäuser, W., Kruse, W., Hoffmann, K. P., & König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Research*, *46*, 1194–1209. [[PubMed](#)]
- Fecteau, J. H., & Munoz, D. P. (2006). Saliency, relevance, and firing: A priority map for target selection. *Trends in Cognitive Sciences*, *10*, 382–390. [[PubMed](#)]
- Felsen, G., & Dan, Y. (2005). A natural approach to studying vision. *Nature Neuroscience*, *8*, 1643–1646. [[PubMed](#)]
- Finney, S. A. (2001). Real-time data collection in linux: A case study. *Behavior Research Methods, Instruments, & Computers*, *33*, 167–173. [[PubMed](#)]
- Gallant, J. L., Connor, C. E., & Van Essen, D. C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport*, *9*, 2153–2158. [[PubMed](#)]
- Goldberg, M. E., Bisley, J. W., Powell, K. D., & Gottlieb, J. (2006). Saccades, saliency and attention: The role of the lateral intraparietal area in visual behavior. *Progress in Brain Research*, *155*, 157–175. [[PubMed](#)]
- Good, P. I. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer-Verlag.
- Harris, C. M., Wallman, J., & Scudder, C. A. (1990). Fourier analysis of saccades in monkeys and humans. *Journal of Neurophysiology*, *63*, 877–886. [[PubMed](#)]
- Hays, A. V. J., Richmond, B. J., & Optican, L. M. (1982). Unix-based multiple-process system for real-time data acquisition and control. *Wescon Conference Record*, *2*, 1–10.
- Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.
- Itti, L. (2004). The iLab neuromorphic vision C++ toolkit: Free tools for the next generation of vision algorithms. *The Neuromorphic Engineer*, *1*, 10. [[Article](#)]
- Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, *12*, 1093. [[Article](#)]
- Itti, L. (2006). Quantitative modelling of perceptual saliency at human eye position. *Visual Cognition*, *14*, 959. [[Article](#)]
- Itti, L., & Baldi, P. (2006). Bayesian surprise attracts human attention. *Advances in Neural Information Processing Systems*, *19*, 547–554. [[Article](#)]
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506. [[PubMed](#)]
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews, Neuroscience*, *2*, 194–203. [[PubMed](#)]
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254. [[Article](#)]
- Judge, S. J., Richmond, B. J., & Chu, F. C. (1980). Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Research*, *20*, 535–538. [[PubMed](#)]
- Kayser, C., Körding, K. P., & König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current Opinion in Neurobiology*, *14*, 468–473. [[PubMed](#)]
- Le Meur, O., Le Callet, P., & Barba, D. (2007). Predicting visual fixations on video based on low-level visual features. *Vision Research*, *47*, 2483–2498. [[PubMed](#)]
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal Experimental Psychology: Human Perception and Performance*, *4*, 565–572. [[PubMed](#)]
- Marino, R. A., Rodgers, C. K., Levy, R., & Munoz, D. P. (2008). Spatial relationships of visuomotor transformations in the superior colliculus map. *Journal of Neurophysiology*, *100*, 2564–2576. [[PubMed](#)]
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*, 614–621. [[PubMed](#)]
- Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, *171*, 308–311. [[PubMed](#)]
- Oliva, A., Torralba, A., Castelano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *Proceedings of the 2003 International Conference on Image Processing (Cat no. 03CH37429)*, *1*, 253.
- Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, *8*, 315–324. [[PubMed](#)]
- Paré, M., & Munoz, D. P. (2001). Expression of a re-centering bias in saccade regulation by superior colliculus neurons. *Experimental Brain Research*, *137*, 354–368. [[PubMed](#)]
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of saliency in the allocation of overt visual attention. *Vision Research*, *42*, 107–123. [[PubMed](#)]

- Parkhurst, D. J., & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16*, 125–154. [[PubMed](#)]
- Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research, 45*, 2397–2416. [[PubMed](#)]
- Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 970.
- Quaia, C., Paré, M., Wurtz, R. H., & Optican, L. M. (2000). Extent of compensation for variations in monkey saccadic eye movements. *Experimental Brain Research, 132*, 39–51. [[PubMed](#)]
- Ramat, S., Leigh, R. J., Zee, D. S., & Optican, L. M. (2007). What clinical disorders tell us about the neural control of saccadic eye movements. *Brain, 130*, 10–35. [[PubMed](#)] [[Article](#)]
- Reinagel, P. (2001). How do visual neurons respond in the real world? *Current Opinion in Neurobiology, 11*, 437–442. [[PubMed](#)]
- Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network, 10*, 341–350. [[PubMed](#)]
- Robinson, D. A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Transactions on Biomedical Engineering, 10*, 137–145. [[PubMed](#)]
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*, 1193–1216. [[PubMed](#)]
- Sokal, R. R., & Rohlf, F. J. (1995). *Biometry: The principles and practices of statistics in biological research*. New York: W. H. Freedman and Company.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17, <http://journalofvision.org/7/14/4/>, doi:10.1167/7.14.4. [[PubMed](#)] [[Article](#)]
- Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45*, 643–659. [[PubMed](#)]
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136. [[PubMed](#)]
- Van Gisbergen, J. A., Robinson, D. A., & Gielen, S. (1981). A quantitative analysis of generation of saccadic eye movements by burst neurons. *Journal of Neurophysiology, 45*, 417–442. [[PubMed](#)]
- Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287*, 1273–1276. [[PubMed](#)]
- Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews of the Cambridge Philosophical Society, 81*, 259–291. [[PubMed](#)]
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.