

Human-Monkey Gaze Correlations Reveal Convergent and Divergent Patterns of Movie Viewing

Stephen V. Shepherd,^{1,*} Shawn A. Steckenfinger,² Uri Hasson,^{1,2} and Asif A. Ghazanfar^{1,2,3}

¹Neuroscience Institute

²Department of Psychology

³Department of Ecology and Evolutionary Biology
Princeton University, Princeton, NJ 08540, USA

Summary

The neuroanatomical organization of the visual system is largely similar across primate species [1, 2], predicting similar visual behaviors and perceptions. Although responses to trial-by-trial presentation of static images suggest that primates share visual orienting strategies [3–8], these reduced stimuli fail to capture key elements of the naturalistic, dynamic visual world in which we evolved [9, 10]. Here, we compared the gaze behavior of humans and macaques when they viewed three different 3-minute movie clips. We found significant intersubject and interspecies gaze correlations, suggesting that both species attend a common set of events in each scene. Comparing human and monkey gaze behavior with a computational saliency model revealed that interspecies gaze correlations were driven by biologically relevant social stimuli overlooked by low-level saliency models. Additionally, humans, but not monkeys, tended to gaze toward the targets of viewed individual's actions or gaze. Together, these data suggest that human and monkey gaze behavior comprises converging and diverging informational strategies, driven by both scene content and context; they are not fully described by simple low-level visual models.

Results

Brains evolved to guide sensorimotor behavior within an immersive, interactive, ever-changing environment. In the laboratory, however, dynamic and interactive environments are problematic because subjects' instantaneous responses to a stimulus change their perceptual experience. For example, although movie viewing offers (at best) a minimalistic model of real-world interactions, viewers' perceptions crucially drive and depend upon ongoing orienting behaviors. Commercially produced movies nonetheless evoke reliable, selective, time-locked activity in many brain areas [11–13]. Shared perceptual responses to movies depend upon shared gaze behavior, which in turn depends upon shared expectations, goals, and strategy [14–17]; predictably, then, these movies also evoke reproducible gaze behavior [9, 10, 18, 19]. To what extent are these stereotyped experiences and perceptual decisions driven by low-level visual cues, as opposed to higher-order features such as ethologically significant objects, actions, or narrative content? One approach to answering this question is to examine the behavior of a closely related species, such as the macaque,

that shares relevant neural structures involved in gaze control [20]. The gaze control system of the macaque is the best-studied primate model of the nested, iterative, sensorimotor decision loops that make up our natural behavior [21–25] and comprises an important substrate in which to address the evolution of behavior. A second approach is to examine whether gaze behavior can be predicted by neurally inspired computational models of visual saliency. Such models have proven effective at locating areas of interest in static scenes based on low-level visual cues [26, 27]. In the present study, we test the hypothesis that humans and monkeys have adapted shared neural mechanisms to identify, localize, and monitor distinct sets of behaviorally relevant stimuli.

Specifically, we combined behavioral and modeling approaches to compare how humans, monkeys, and computer simulations respond during initial and repeat viewings of movie clips. Clips were taken from three films. One movie featured monkeys in natural environments (the BBC's *The Life of Mammals*), one featured cartoon humans and animals (Disney's *The Jungle Book*), and one featured human social interactions (Charlie Chaplin's *City Lights*). The movie clips were 3 minutes in duration, converted to black and white, and stripped of their soundtrack. Each subject viewed each movie clip multiple times in random sequence. **Figures 1A–1C** show movie frames with superimposed gaze locations (humans in blue; monkeys in green); **Figures 2A–2C** show representative human and monkey gaze traces (see also **Movies S1–S3** available online). We found that the patterns of fixations of humans and monkeys across the movie clips were broadly similar. Scanpaths were significantly correlated across different viewings by humans and monkeys. These correlations were especially pronounced among humans, for whom the average interscanpath correlation (ISC) was almost as high between ($r = 0.39$, permutation test, $p < 0.001$) as within subjects ($r = 0.44$, $p < 0.001$), consistent with past reports [10]. Correlations between monkeys were also significant, but substantially lower than among humans (average same-monkey $r = 0.22$, $p < 0.001$; between-monkey $r = 0.10$, $p < 0.001$); correlations between species were significant and of comparable size to correlations between individual monkeys (average $r = 0.10$, $p < 0.001$) (**Figure 2D**; see also **Figure S1A**). Finally, eye movement speed, like gaze position, was correlated between viewers (average same-human $r = 0.17$, $p < 0.001$; between-human $r = 0.14$, $p < 0.001$; across-species $r = 0.04$, $p < 0.001$; between-monkey $r = 0.04$, $p < 0.001$; same-monkey $r = 0.11$, $p < 0.001$; **Figure 2E**; see also **Figure S1B**). The most likely way for such correlations to arise is if different primates fixate similar locations at similar times; however, because correlation is invariant to shifting and scaling transformations, additional analyses were necessary to confirm this interpretation. We directly analyzed scanpath overlap, counting the percentage of samples in which one scanpath was within 3.5° of the other: Human scanpaths overlapped on average 70% of the time between repetitions and 65% between individuals; monkeys overlapped 33% between repetitions and 27% between individuals; between species, scanpaths overlapped 31% of the time (**Figure 2F**). Together,

*Correspondence: stephen.v.shepherd@gmail.com

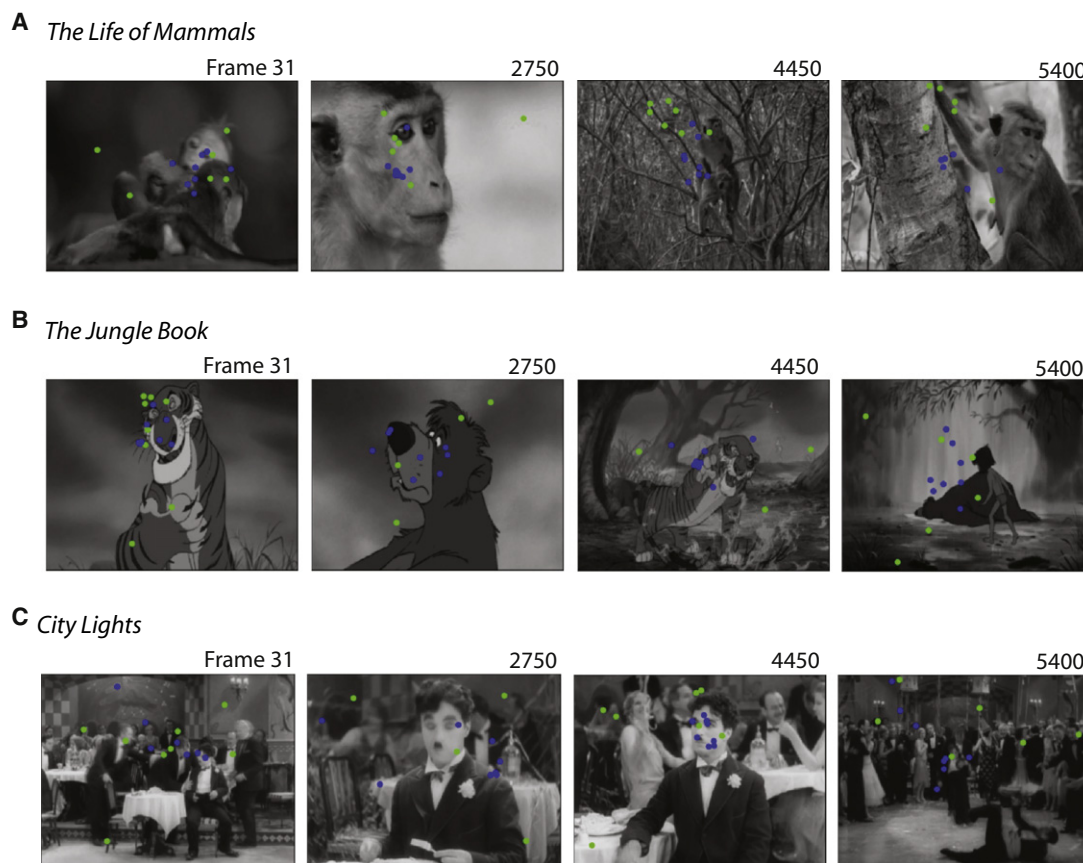


Figure 1. Video Scenes with Superimposed Gaze Coordinates

Example frames from *The Life of Mammals* (A), *The Jungle Book* (B), and *City Lights* (C), shown with superimposed gaze coordinates (monkeys, green; humans, blue).

these findings suggest that humans and monkeys use similar spatiotemporal visual features to guide orienting behavior.

Such similar scanpaths can arise in a variety of ways. One possibility is that shared orienting behaviors are driven solely by salient low-level visual features. At the other extreme, shared orienting behaviors might be driven primarily by high-level behaviorally relevant narrative content. To distinguish between these possibilities, we compared human and monkey scanpaths to artificial scanpaths generated by a well-validated low-level saliency model [26–29]. Indeed, artificial scanpaths did correlate with human and monkey gaze positions in our experiment (Figure 2D, orange bars), but even the best-correlated artificial scanpath ($r = 0.06$, $p < 0.001$; see [Experimental Procedures](#)) played a negligible role in mediating primate ISCs: residual ISCs were just as pronounced after partialing out similarities to artificial scanpaths (average residuals: same-human $r = 0.44$; between-human $r = 0.39$; across-species $r = 0.10$; between-monkey $r = 0.09$; same-monkey $r = 0.22$). Artificial scanpaths were less successful at modeling the timing of attention shifts ($r = 0.01$, $p < 0.05$) and did not predict primate ISCs in gaze shift timing. Although artificial scanpaths overlapped with observed human and monkey scanpaths 28% and 20% of the time, respectively (for the best-performing simulation, see Figure 2F, orange bars), they were strikingly poor at predicting human and monkey overlap: of the 31% of samples in which human and monkey gaze overlapped, only 1 in 15 (2.1% of total) also overlapped the

simulated scanpath (Figure 2F, inset). Application of multidimensional scaling to average normalized interscanpath distances revealed that each video produced a distinct cluster of human, monkey, and simulated scanpaths (Figure S2); for *The Life of Mammals* and the *The Jungle Book* movie clips, human and monkey scanpaths clustered tightly and were separate from artificial scanpaths.

Tracking the standard deviation of gaze coordinates across viewers as a function of time proved to be an effective way of screening for patterns of interactions with the environment. By tracking the standard deviation of human (Figure 3A) and monkey (Figure 3B) gaze coordinates as they watched *The Life of Mammals*, we identified moments at which gaze was significantly clustered (below shaded area) or dispersed (above shaded area). Furthermore, by comparing human and monkey results to one another or to the standard deviation across all primate scanpaths (Figure 3C), we can define scenes of interest in four categories: (1) scenes that significantly dispersed both human and monkey gaze (Figures 3A and 3B, gray boxes above shaded area; example in Figure 3D), (2) scenes that significantly clustered one species while dispersing another (Figures 3A and 3B, orange diamonds; examples in Figure 3E), (3) scenes that significantly clustered both human and monkey gaze in the same place (Figures 3A and 3B, gray boxes below shaded area; example in Figure 3F), and (4) scenes that separately clustered human and monkey gaze at different locations (Figures 3A and 3B,

red circles; example in Figure 3G). Both humans and monkeys generally looked toward faces and toward interacting individuals (Figure 3F), and although both humans and monkeys sometimes scanned the broader scene (Figure 3D), monkeys shifted gaze away from objects of interest more readily and more often. For example, of the three examples of differential clustering shown in Figure 3E, two occurred when monkeys shifted gaze away from regions of interest to scan the background, and one occurred when monkeys, but not humans, quickly scanned newly revealed scenery during a camera pan. Monkeys and humans sometimes made collectively different decisions about where to look, and these differences sometimes reflected differential understanding of movie content: for example, whereas humans used cinematic conventions to track an individual of interest, looking to the character appearing centermost on the screen, monkeys instead tracked the more active member of a pair—even as he jumped offscreen (Figure 3G). To identify crucial visual stimuli omitted from the low-level saliency model, we selected frames on which a species' gaze was strongly clustered but far from the simulated scanpath. We then contrasted image content at three locations on each frame: a location viewed by a monkey, a location viewed by a human, and the location selected by the artificial scanpath (Figure 4). On each frame, these three locations were scored (blindly, in random order) as including an individual's body, hands, face, ears, eyes, or mouth and as being the target of another individual's actions or attention. We found that humans and monkeys gazed toward individuals in a scene significantly more than predicted by low-level visual saliency models. Both species looked particularly often at faces and eyes. Remarkably, humans, but not monkeys, strongly attended objects being manipulated or examined by others.

Although statistically significant, the interspecies ISCs were low in magnitude. One explanation for the low magnitude could be species differences in basic eye movements. For example, although humans and monkeys displayed a strong central bias, monkeys exhibited a broader spatial range of fixations (Figure S3A). Similarly, although the dynamics of eye movement were similar between species, humans exhibited shorter saccades and longer fixations than monkeys (Figure S3B). These data are consistent with previous reports (e.g., [30–32]) and suggest that differences in gaze dynamics during movie viewing contribute to lowered ISCs between species.

Discussion

We found that gaze behavior during movie viewing was significantly correlated across repetitions, individuals, and even species. Gaze behavior during temporally extended video likely depends on species-specific cues, the ability to integrate events over time, and familiarity and fluency with videos. ISCs were substantially stronger among human participants than among monkeys or between humans and monkeys. Computational models of low-level video saliency poorly accounted for behavioral correlations within and between species. Primate scanpaths significantly overlapped, but this overlap seemed not to be mediated by low-level visual saliency: in particular, low-level models missed crucial biological stimuli such as faces and their expressions, bodies and their movements, and (particularly for humans) observed social signals and behavioral cues. ISCs during natural viewing suggest that in the absence of explicit, immediate goals—intrinsic or

instructed—orienting priorities are overwhelmingly similar [9, 10, 31] and focused on faces and social interactions. The importance of such behaviorally relevant visual cues has been supported by findings that primates quickly discriminate animate stimuli [33], facial locations [34, 35], facial expressions (reviewed in [36]), and gaze directions [34, 37, 38] and encode these social variables in neurons governing attention [36, 39–41].

Although ISCs between species and between monkeys were significant, they were lower than among humans. Several accounts may explain why correlations between monkeys were less pronounced than between humans, as has previously been reported for still images [42]. Like Berg et al. [30], we found that humans and monkeys have similar gaze behavior but differ in the degree of central bias, in the duration and regularity of fixation periods, and in the amplitude of saccades (Figure S3). This may suggest species-specific visual strategies, with monkeys fixating for short and stereotyped intervals separated by large saccades and humans fixating for more prolonged and variable periods. Such differences might facilitate relatively fast threat and resource detection by monkeys and are also consistent with the finding that monkeys abbreviate fixations toward high-risk social targets, such as high-ranking male faces [43]. Alternatively, monkeys may fail to orient systematically in response to video content because they fail to attend toward or understand the meaning of videos. Our findings echo reports that monkeys poorly integrate and generalize concepts from laboratory experiences [44, 45] and choose to watch videos only after accruing adequate experience with the medium [46]. Most humans are familiar with video broadcast, and this familiarity likely both shapes viewer expectations and increases viewer interest; likewise, cinematographers craft movies to entertain humans, not monkeys. Indeed, human gaze anticipates areas of interest even when viewing novel movie scenes [31]. If monkeys were inadequately engaged by video, it was not solely due to anthropocentric visual content: Humans and monkeys had similar responses to the three videos, independent of ecological relevance. For both species, *The Jungle Book*—a children's cartoon—strongly and consistently captured gaze, whereas *City Lights*—a visually crowded comedy—did so weakly (see Figure S1).

Our world is not static, and subtle perceptual behaviors, such as orienting, transform incoming sensation. However, ISCs suggest that across primates, complex and dynamic stimuli nonetheless may evoke consistent cognitive and behavioral responses. Human gaze is nearly as consistent across individuals as across repetitions; furthermore, significant correlations are evident between monkeys and between monkeys and humans. These findings extend the pioneering experiments of Buswell [47] and Yarbus [17] to natural temporal sequences: Although our data cannot reveal covert orienting decisions, they strongly suggest that primates attend similar features and shift attention at similar times. Weaker correlations in monkeys than in humans may be due to species differences in vigilance or fluency. Importantly, ISCs within and between species were greater than could be explained by low-level saliency models: In particular, primates respond to biologically relevant features including animate objects and faces. Finally, we found that humans, but not monkeys, strongly attended the foci of other individuals' attention and activity. This tendency is provocative and suggests a synchronizing force at work in human social evolution. Primates tend to passively orient in similar directions, making observed

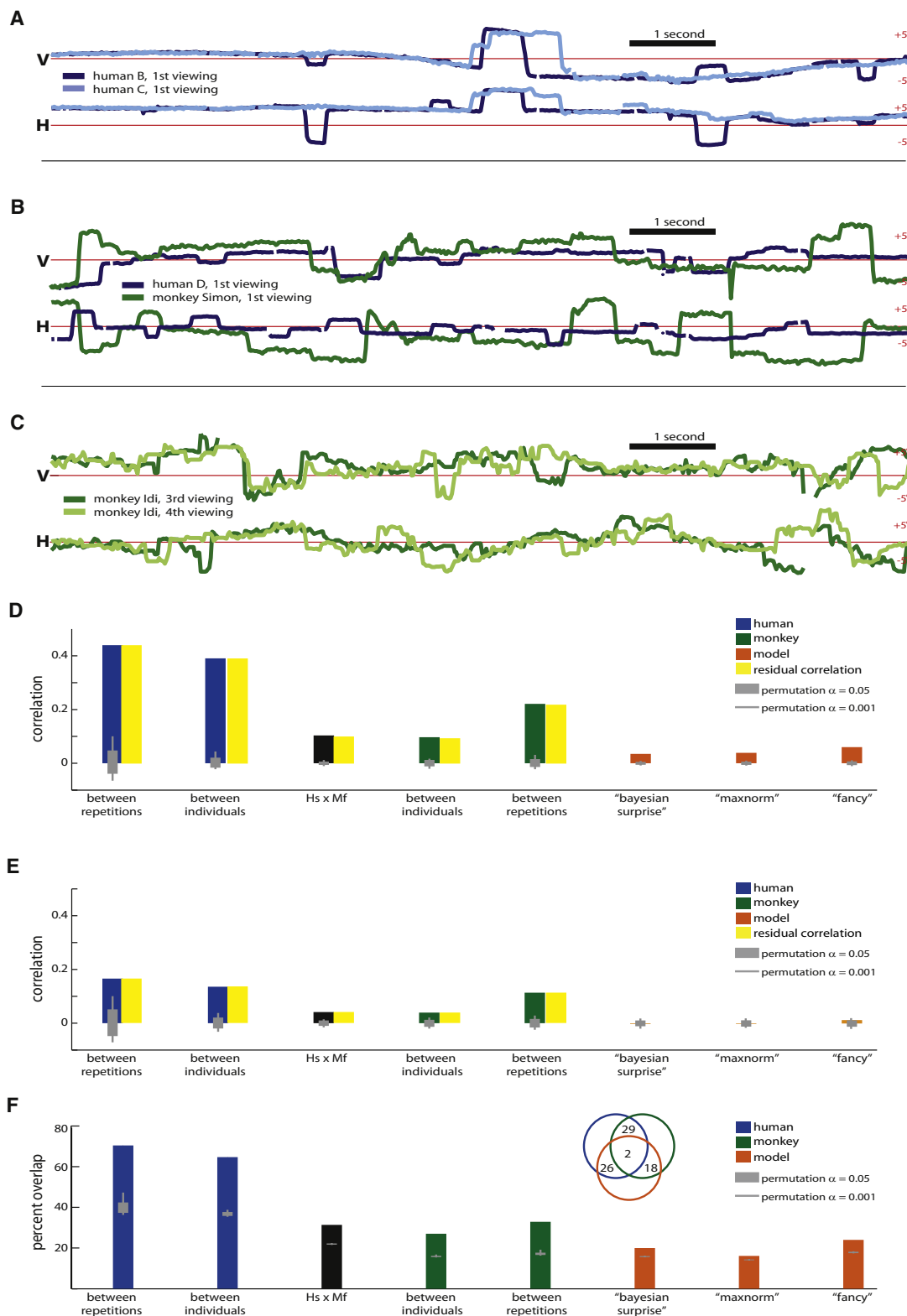


Figure 2. Correlation in Gaze across Scanpaths

(A–C) Correlations are evident across humans, monkeys, and between species. Here, scanpaths are split into vertical and horizontal coordinates and plotted for two humans viewing *City Lights* (local correlation $r = 0.66$) (A), a human and a monkey viewing *The Jungle Book* (local correlation $r = 0.08$) (B), and one monkey during repeated viewings of *The Life of Mammals* (local correlation $r = 0.36$) (C). The local correlations are typical of the interhuman, interspecies, and intramonkey scanpaths, respectively. Significant correlations existed between primate scanpaths produced in response to the same video clip. (D and E) Both spatial position (D) and eye movement speed (E) were correlated across primate scanpaths, whether produced by the same individual or a different individual and whether produced by a human (blue) or a monkey (green). Gray bars indicate the permutation baseline for $\alpha = 0.05$ (thick bars)

gaze a useful indicator of important environmental features, in turn incentivizing gaze following [40, 48], which further correlates our collective behavior. Interscanpath correlation may thus prove important not just because of what it tells us about the evolution of the visual system, but because of what it reveals about the evolution of primate societies.

Experimental Procedures

Participants

Two adult male long-tailed macaques, *Macaca fascicularis*, and four adult male humans participated in the study. Nonhuman participants were born in captivity and socially housed indoors; all nonhuman experimental procedures were in compliance with the local authorities, National Institutes of Health guidelines, and Institutional Animal Care and Use Committee standards for the care and use of laboratory animals. Human participants provided informed consent under a protocol authorized by the Institutional Review Board of Princeton University and were debriefed at the conclusion of the session. In addition to humans and monkeys, data sets collected from an artificially intelligent agent were derived from the iLab Neuromorphic Vision C++ Toolkit [29] (<http://ilab.usc.edu/toolkit/downloads-virtualbox.shtml>, downloaded May 26, 2009).

Stimulus Presentation

The three visual stimuli consisted of silent, grayscale, 3-minute digital video clips taken from *City Lights* (1931), *The Jungle Book* (1967), or *The Life of Mammals: Social Climbers* (2003). Charlie Chaplin's film *City Lights* has been used in earlier fMRI experiments [10] and features humans in indoor environments. *The Jungle Book*, a cartoon, includes simplified caricatures of human and animal stimuli. Finally, the scene from *The Life of Mammals* features macaques in their natural habitat.

Human stimuli were presented on a 60 Hz, 17-inch LCD monitor operating at 1024 × 768 resolution at a distance of 85 cm. This provided a 22° × 18° field of view of the monitor. The 770 × 584 videos, centrally located, subtended approximately 17° × 14°. Human eye data were captured with a Tobii X120 Eye Tracker (www.tobii.com) at 120 Hz. Prior to each session, participants completed a five-point calibration. Following calibration, participants completed a nine-point calibration check three times consecutively and then viewed the three videos separated by 30 s intervals of blank screen. A 90 s break preceded another three consecutive nine-point calibration checks; followed by video presentation again in random order separated by 30 s blanks. A final nine-point calibration check concluded the session. To pass each point in the calibration check, the system was required to report sustained gaze within 2.5° of a 0.4° fixation target.

Monkey subjects sat in a primate chair fixed 74 cm away from a 17-inch LCD or CRT monitor operating at 60 Hz and 1024 × 768 resolution and were restrained via head prosthesis. This provided a 25° × 20° view of the monitor. All video stimuli were located centrally and occupied an area of 770 × 584 pixels; this subtended a visual angle of 19° × 16°. Monkey eye data were captured with an ASL Eye-Trac 6000 (www.asleyetracking.com) with either an ASL R6 remote optics camera operating at 60 Hz (LCD rig) or an ASL high-speed optics camera operating at 120 Hz (CRT rig). Prior to each session, monkeys completed a nine-point calibration. In the second session only, nine-point calibration checks confirmed gaze tracking accuracy as described above. The sequence of calibration checks and video playback was identical to that for humans, with the exception that monkeys were rewarded with juice during calibration checks and were randomly given juice throughout the course of the videos.

Quantifying Eye Movement Behavior

Eye data were downsampled in MATLAB (www.mathworks.com) to 60 Hz (10,740 data points), and all offscreen fixations and signal loss were recorded

as “not a number.” In total, this filtering rejected 16.8% of the monkey eye traces and 11.5% of the human eye traces. To facilitate low-level gaze analysis, we grouped eye data into saccades or fixations by using a velocity-based criterion. Fixations were defined as eye movements in which the total velocity did not exceed 20°/s. Fixations shorter than 100 ms (six samples) were discarded and integrated into the surrounding saccade, and fixations separated by 17 ms (one sample) or saccades smaller than 2° were merged into single fixation events.

Saliency Map and Simulated-Gaze Generation

To model orienting responses to low-level visual stimuli, we generated artificial scanpaths toward each video with the Sun VirtualBox (www.virtualbox.org) implementation of the iLab C++ Neuromorphic Vision Toolkit [29]. Details regarding the development of this toolkit have been published elsewhere [26–30, 49].

Analysis

For each pair of scanpaths, a general correlation was obtained by averaging the *r* values obtained from the series of horizontal and of vertical coordinates. To compare the similarity in gaze shift timing between pairs of scanpaths, we first smoothed the spatial position across time by using a Gaussian kernel 100 ms (six samples) in standard deviation and then calculated correlations in the absolute value of the first derivative. Additionally, we performed an analysis of scanpath overlap, which we operationalized as the percentage of time points for which paired scanpath coordinates were within 3.5° of one another (2.5° error radius + 1° foveal radius). Finally, to detect high-dimensional scanpath features that may have varied across scanpaths, we performed a multidimensional scaling (MDS) of interscanpath distance, normalized by dividing out the average shuffled interscanpath distance. MDS maps high-dimensional data to a low-dimensional surface in which map proximity correlates with similarity and was implemented with the MATLAB command *mdscale*.

To determine the significance of interscanpath correlations, we needed to correct for sample-to-sample correlations within scanpaths. To do this, we established baselines via a consecutivity-preserving time-shuffling permutation procedure. Instead of randomly sampling each data point individually, we took the entire sequence of time points, randomly flipped the direction, and rotated the indices so as to randomize timestamps while adding a single temporal discontinuity where the last sample looped back to the first. We then recalculated the statistic to be tested with the newly permuted data and repeated. This population comprises the “chance” baseline against which our observations can be compared: If our observations lie outside the 2.5th and 97.5th percentile, for example, then it is significant at a two-tailed α level of 0.05. All permutation values reported here used this procedure unless otherwise indicated. (As a precaution, we also performed these analyses without randomly reversing the temporal order of samples; results were not significantly altered.)

Because low-level visual features may have influenced human and monkey gaze in similar ways, we compared primate scanpaths to artificially generated gaze sequences (described above). Specifically, we correlated human and monkey gaze behavior with the behavior of artificially generated simulated eye movements and recalculated interscanpath correlations after partialing out the artificial scanpath with MATLAB's partial correlation function. We likewise compared scanpath overlap between human and monkey scanpaths and simulated scanpaths, and—to establish whether overlap in primate scanpaths was mediated by low-level visual features—measured the three-way overlap between human scanpaths, monkey scanpaths, and the best-performing simulation.

To determine those factors that consistently influenced human and monkey gaze but were missed by the low-level saliency model, we selected frames on which gaze was strongly clustered (the standard deviations of the gaze locations were in the bottom 5% observed for that species) but where the artificial scanpath was unusually far from gaze (more than a standard

or 0.001 (thin bars): all primate gaze interscanpath correlations (ISCs) were significant with $\alpha < 0.001$. Artificial scanpaths produced by a low-level visual saliency model (orange) were significantly correlated with primate scanpaths in spatial position ($\alpha < 0.001$) and gaze shift timing ($\alpha < 0.05$); however, residual interprimate correlations (yellow) were essentially unchanged despite partialing out shared similarities to artificial scanpaths.

(F) Finally, to confirm that behavioral correlations were driven by visual fixation priorities, we compared the percentage of gaze samples that overlapped ($\pm 3.5^\circ$) across different scanpaths. The pattern of results was identical to the pattern observed for intersubject correlation. Furthermore, we found that samples that overlapped between humans and monkeys rarely overlapped with the best-performing artificial scanpath (2% of samples overlapped between humans, monkeys, and artificial scanpaths, a small fraction of the 31% of samples that overlapped between humans and monkeys; see inset). These data rule out the hypothesis that gaze correlations are driven primarily by low-level visual features, at least as characterized by well-established neuromorphic computational saliency models [27, 29].

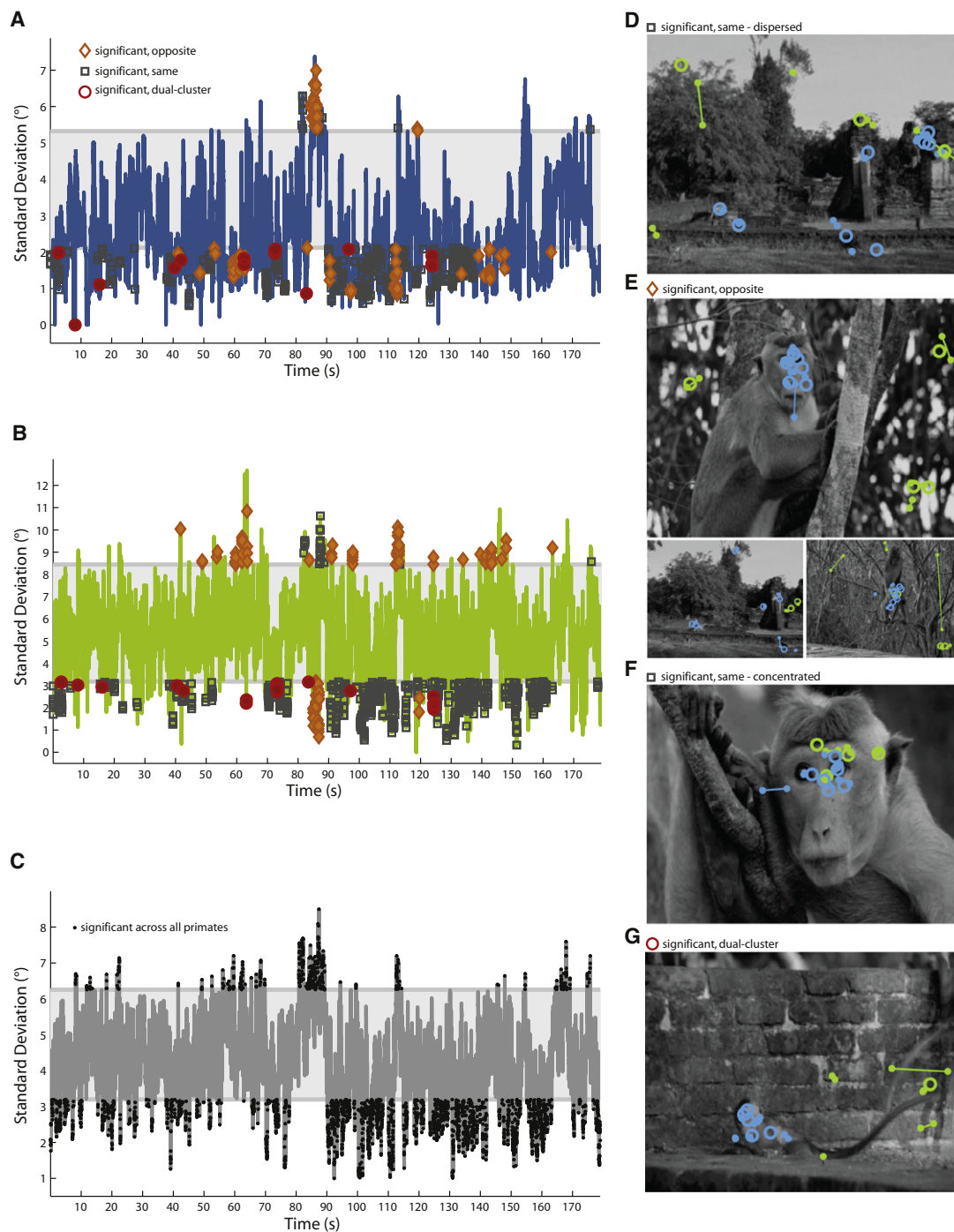


Figure 3. Comparison of Gaze Reliability across Time, by Species

(A–C) The standard deviation of simultaneously recorded gaze coordinates from humans (A), monkeys (B), and both species (C) can, in conjunction, differentiate the visual strategies of monkeys or humans.

(D–G) Different scenes from *The Life of Mammals* significantly dispersed viewer attention (D, a long shot featuring a number of monkeys) or gathered it (E, social stimuli captured sustained human interest, whereas newly visible scenery during a pan was quickly surveyed by monkeys). Faces often captured the attention of all primates (F), whereas dyadic social interactions sometimes produced separate gaze clusters for humans and macaques (G).

deviation above average). When multiple frames were identified within the same 0.5 s period, only the first was accepted. We then scored these frames for image content at three locations—the artificial scanpath, a random human scanpath, and a random monkey scanpath—in a random order unknown to the scorer. Image content at a given location was described as including a social agent or the target of an agent’s action or gaze.

Additionally, fixations on social agents that fell on faces or hands were tallied; facial fixations were likewise tallied based upon fixations on ears, eyes, or mouth. Throughout scoring, the observer was blind as to whether they were scoring an artificial, human, or monkey scanpath. Finally, the significance of differential image content at gaze-selected versus model-selected regions was determined by χ^2 test.

Gaze Behavior of Man, Monkey, and Machine

7

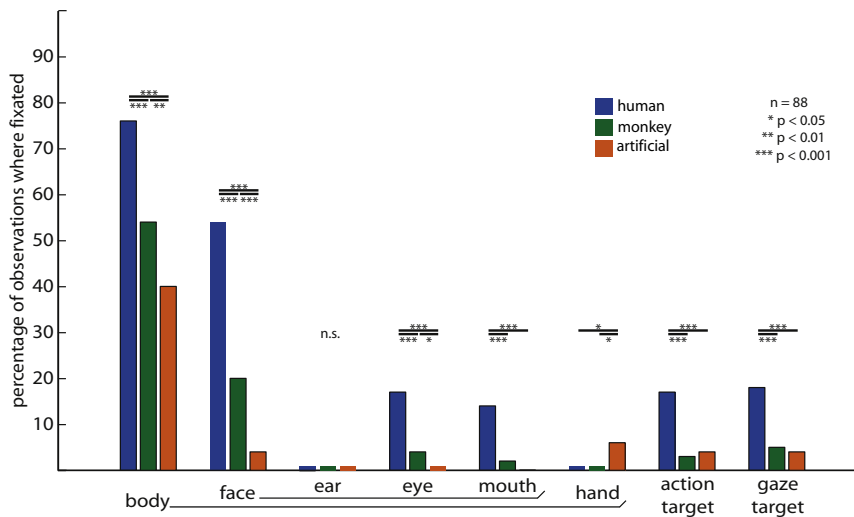


Figure 4. Higher-Level Social Cues Contribute to Observed Gaze Correlations

We selected video frames in which observed gaze was highly clustered at a location far from that chosen by the low-level saliency model; examined scene content at the human, monkey, and simulated-gaze coordinates; and tallied the percentage of frames in which scene content comprised biologically relevant stimuli. Relative to image regions selected by artificial scanpaths, regions that were consistently viewed by humans and monkeys often featured social agents and the targets of their actions or attentions. Humans and monkeys fixated bodies and faces significantly more often than predicted by low-level simulations, especially the eye and (for humans) mouth regions. Humans, but not monkeys, fixated socially cued regions—for example, things being reached or gazed toward—significantly more often than predicted by the model. This effect was not mediated just by hand motion, which attracted gaze significantly more often for the simulation than for actual humans or monkeys.

Supplemental Information

Supplemental Information includes three figures, Supplemental Experimental Procedures, and three movies and can be found with this article online at doi:10.1016/j.cub.2010.02.032.

Acknowledgments

This work was supported by Princeton University Training Grant in Quantitative Neuroscience NRSA T32 MH065214-1 (S.V.S.), National Science Foundation CAREER Award BCS-0547760 (A.A.G.), and Autism Speaks (A.A.G.).

Received: December 9, 2009
 Revised: January 29, 2010
 Accepted: February 1, 2010
 Published online: March 18, 2010

References

1. Kaas, J.H., and Preuss, T.M. (1993). Archontan affinities as reflected in the visual system. In *Mammalian Phylogeny*, F.S. Szalay, M.J. Novacek, and M.C. McKenna, eds. (New York: Springer-Verlag), pp. 115–128.
2. Krubitzer, L.A., and Kaas, J.H. (1990). Cortical connections of MT in four species of primates: Areal, modular, and retinotopic patterns. *Vis. Neurosci.* *5*, 165–204.
3. Ghazanfar, A.A., Nielsen, K., and Logothetis, N.K. (2006). Eye movements of monkey observers viewing vocalizing conspecifics. *Cognition* *101*, 515–529.
4. Gothard, K.M., Brooks, K.N., and Peterson, M.A. (2009). Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim. Cogn.* *12*, 155–167.
5. Gothard, K.M., Erickson, C.A., and Amaral, D.G. (2004). How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Anim. Cogn.* *7*, 25–36.
6. Guo, K. (2007). Initial fixation placement in face images is driven by top-down guidance. *Exp. Brain Res.* *181*, 673–677.
7. Guo, K., Robertson, R.G., Mahmoodi, S., Tadmor, Y., and Young, M.P. (2003). How do monkeys view faces?—A study of eye movements. *Exp. Brain Res.* *150*, 363–374.
8. Nahm, F.K.D., Perret, A., Amaral, D.G., and Albright, T.D. (1997). How do monkeys look at faces? *J. Cogn. Neurosci.* *9*, 611–623.
9. Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D.J. (2008). Neurocinematics: The neuroscience of film. *Projections* *2*, 1–26.
10. Hasson, U., Yang, E., Vallines, I., Heeger, D.J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* *28*, 2539–2550.
11. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Inter-subject synchronization of cortical activity during natural vision. *Science* *303*, 1634–1640.
12. Bartels, A., and Zeki, S. (2005). The chronoarchitecture of the cerebral cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *360*, 733–750.
13. Hasson, U., Malach, R., and Heeger, D.J. (2010). Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* *14*, 40–48.
14. Henderson, J.M. (2003). Human gaze control during real-world scene perception. *Trends Cogn. Sci.* *7*, 498–504.
15. Land, M.F., and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Res.* *41*, 3559–3565.
16. Shinoda, H., Hayhoe, M.M., and Shrivastava, A. (2001). What controls attention in natural environments? *Vision Res.* *41*, 3535–3545.
17. Yarbus, A.L. (1967). *Eye Movements and Vision* (New York: Plenum Press).
18. Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., and Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proc. Biol. Sci.* *276*, 3635–3644.
19. Goldstein, R.B., Woods, R.L., and Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Comput. Biol. Med.* *37*, 957–964.
20. Ghazanfar, A.A., and Santos, L.R. (2004). Primate brains in the wild: The sensory bases for social interactions. *Nat. Rev. Neurosci.* *5*, 603–616.
21. Ikeda, T., and Hikosaka, O. (2003). Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron* *39*, 693–700.
22. Hikosaka, O. (2007). Basal ganglia mechanisms of reward-oriented eye movement. *Ann. N Y Acad. Sci.* *1104*, 229–249.
23. Leon, M.I., and Shadlen, M.N. (1999). Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* *24*, 415–425.
24. Platt, M.L., and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* *400*, 233–238.
25. Sugrue, L.P., Corrado, G.S., and Newsome, W.T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science* *304*, 1782–1787.
26. Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* *40*, 1489–1506.
27. Itti, L., Dhavale, N., and Pighin, F. (2003). Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proc. SPIE 48th Annual International Symposium on Optical Science and Technology*, Volume 5200, B. Bosacchi, D.B. Fogel, and J.C. Bezdek, eds. (Bellingham, WA: SPIE Press), pp. 64–78.
28. Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* *20*, 1254–1259.
29. Itti, L. (2004). The iLab neuromorphic vision C++ toolkit: Free tools for the next generation of vision algorithms. *Neuromorphic Eng.* *1*, 10.

30. Berg, D.J., Boehnke, S.E., Marino, R.A., Munoz, D.P., and Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *J. Vis.* **9**, 1–15.
31. Tosi, V., Mecacci, L., and Pasquali, E. (1997). Scanning eye movements made when viewing film: Preliminary observations. *Int. J. Neurosci.* **92**, 47–52.
32. Brasel, S.A., and Gips, J. (2008). Points of view: Where do we look when we watch TV? *Perception* **37**, 1890–1894.
33. Kirchner, H., and Thorpe, S.J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Res.* **46**, 1762–1776.
34. Fletcher-Watson, S., Findlay, J.M., Leekam, S.R., and Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception* **37**, 571–583.
35. Cerf, M., Harel, J., Einhaeuser, W., and Koch, C. (2007). Predicting human gaze using low-level saliency combined with face detection. *Adv. Neural Inf. Process. Syst.* **20**, 241–248.
36. Vuilleumier, P. (2002). Facial expression and selective attention. *Curr. Opin. Psychiatry* **15**, 291–300.
37. Deaner, R.O., and Platt, M.L. (2003). Reflexive social attention in monkeys and humans. *Curr. Biol.* **13**, 1609–1613.
38. Friesen, C.K., and Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon. Bull. Rev.* **5**, 490–495.
39. Klein, J.T., Deaner, R.O., and Platt, M.L. (2008). Neural correlates of social target value in macaque parietal cortex. *Curr. Biol.* **18**, 419–424.
40. Shepherd, S.V. (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Front. Integr. Neurosci.*, in press. 10.3389/fnint.2010.00005.
41. Shepherd, S.V., Klein, J.T., Deaner, R.O., and Platt, M.L. (2009). Mirroring of attention by neurons in macaque parietal cortex. *Proc. Natl. Acad. Sci. USA* **106**, 9489–9494.
42. Einhäuser, W., Kruse, W., Hoffmann, K.P., and König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Res.* **46**, 1194–1209.
43. Deaner, R.O., Khera, A.V., and Platt, M.L. (2005). Monkeys pay per view: Adaptive valuation of social images by rhesus macaques. *Curr. Biol.* **15**, 543–548.
44. Nielsen, K.J., Logothetis, N.K., and Rainer, G. (2006). Discrimination strategies of humans and rhesus monkeys for complex visual displays. *Curr. Biol.* **16**, 814–820.
45. Nielsen, K.J., Logothetis, N.K., and Rainer, G. (2008). Object features used by humans and monkeys to identify rotated shapes. *J. Vis.* **8**, 1–15.
46. Humphrey, N.K., and Keeble, G.R. (1976). How monkeys acquire a new way of seeing. *Perception* **5**, 51–56.
47. Buswell, G.T. (1935). *How People Look at Pictures: A Study of the Psychology of Perception in Art* (Chicago: University of Chicago Press).
48. Klein, J.T., Shepherd, S.V., and Platt, M.L. (2009). Social attention and the brain. *Curr. Biol.* **19**, R958–R962.
49. Itti, L., and Baldi, P. (2006). Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems*, Volume 19, B. Schölkopf, J. Platt, and T. Hofmann, eds. (Cambridge, MA: MIT Press), pp. 547–554.

Supplemental Information

Human-Monkey Gaze Correlations

Reveal Convergent and Divergent

Patterns of Movie Viewing

Stephen V. Shepherd, Shawn A. Steckenfinger, Uri Hasson, and Asif A. Ghazanfar

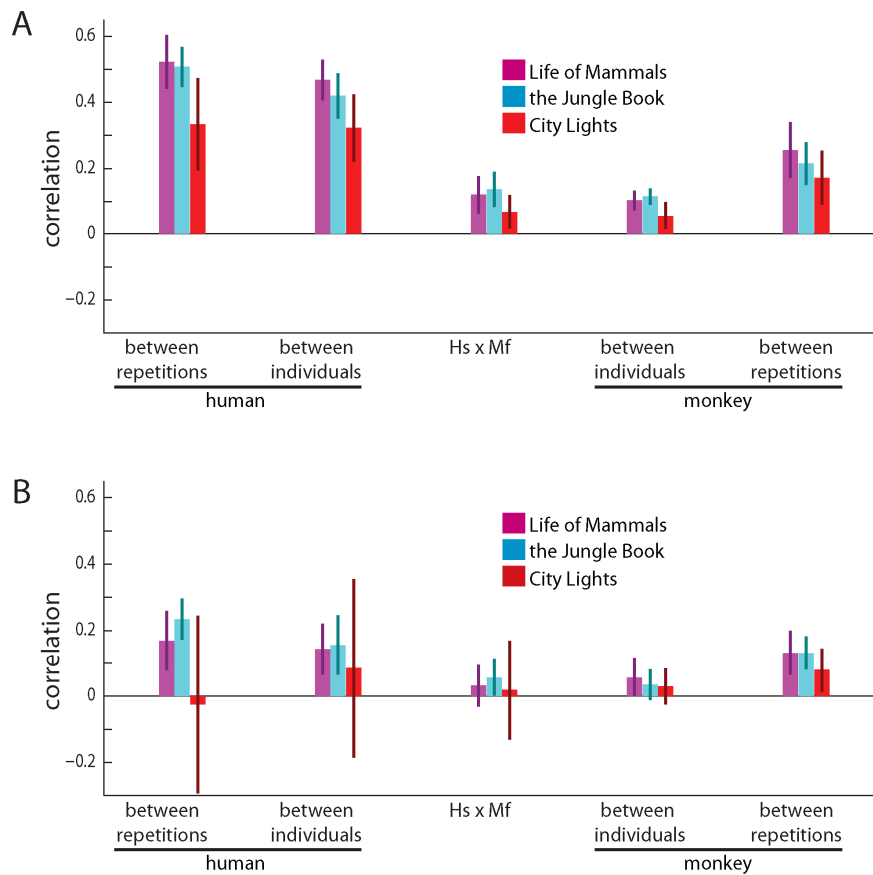


Figure S1. Gaze ISC across Primates by Movie Clip Source

Across all three movies, spatial position (A) and eye movement speed (B) tended to correlate across primate scanpaths, whether produced by the same individual or a different individual and whether produced by a human (left) or a monkey (right). For both species, ISCs were higher and more consistent for clips from *The Life of Mammals* (magenta) and *The Jungle Book* (cyan) than from *City Lights* (red). Thick bars illustrate the average r value across pairwise correlations; thin bars, the standard deviation.

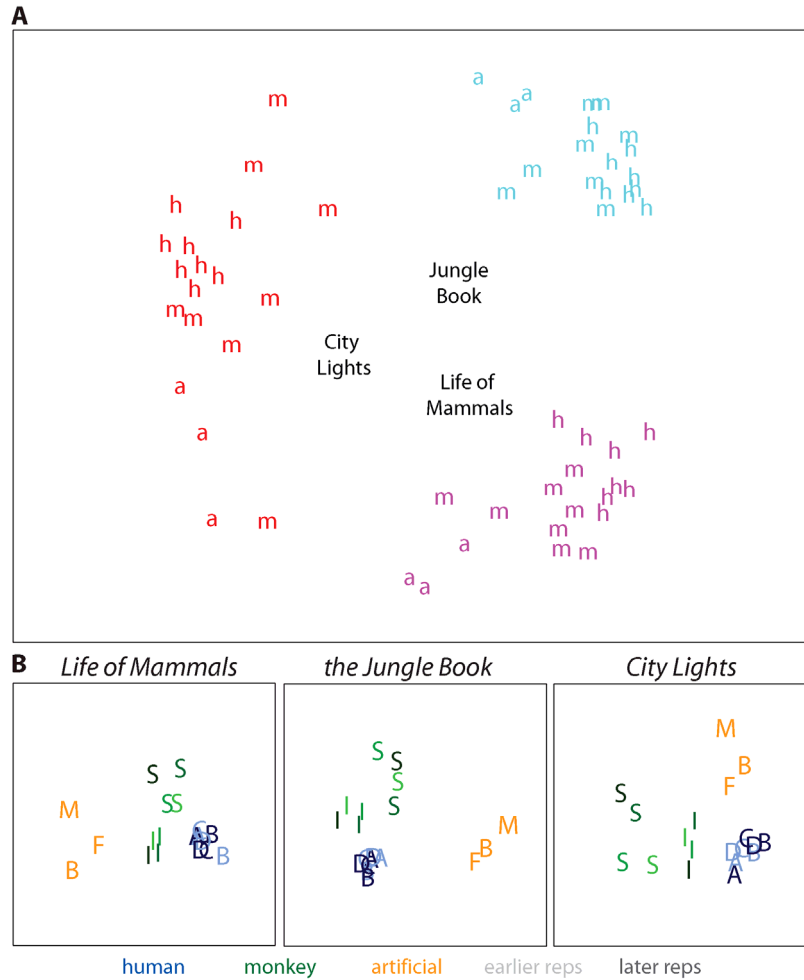


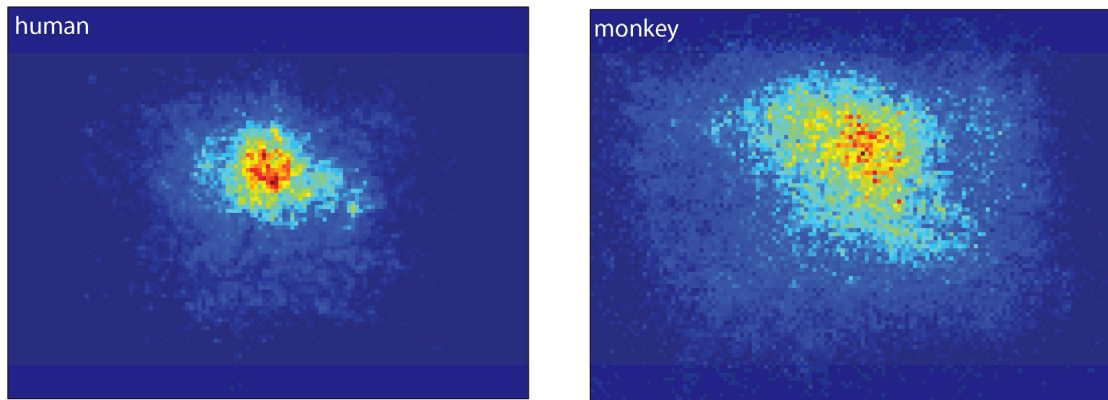
Figure S2. Interscanpath Distances Reflect Stimulus and Subject Identity

Scanpaths are highly multidimensional stimuli in which meaningful differences can be difficult to visualize. To simply represent differences between scanpaths, we performed a multidimensional scaling (MDS) of normalized scanpath distances. This procedure results in a two dimensional map of observations in which more similar observations are grouped and more dissimilar separated.

(A) MDS of scanpaths evoked by humans (h), monkeys (m), and simulated agents (a) in response to *City Lights* (red), *The Jungle Book* (cyan), and *The Life of Mammals* (magenta). Scanpaths were most strongly influenced by video stimulus, then by simulation or species of origin.

(B) Separate MDS of monkey (green), human (blue), and simulated (orange) scanpaths over each movie, in which letter indicates source identity and shading reflects viewing order (lightest first). In each movie, monkey and human scanpaths are well-separated from artificial scanpaths, whether created using “Bayesian surprise” (B), “fancy” (F), or “maxnorm” (M) normalizations (see Methods). This separation indicates that simulated scanpaths poorly captured common high-dimensional features of human and monkey gaze behavior. Furthermore, because primate clusters neither converge nor diverge with repetition, it appears that repeated exposures neither produced more standardized nor more idiosyncratic gaze behavior: Repetition had no obvious effect on scanpath similarity.

A



B

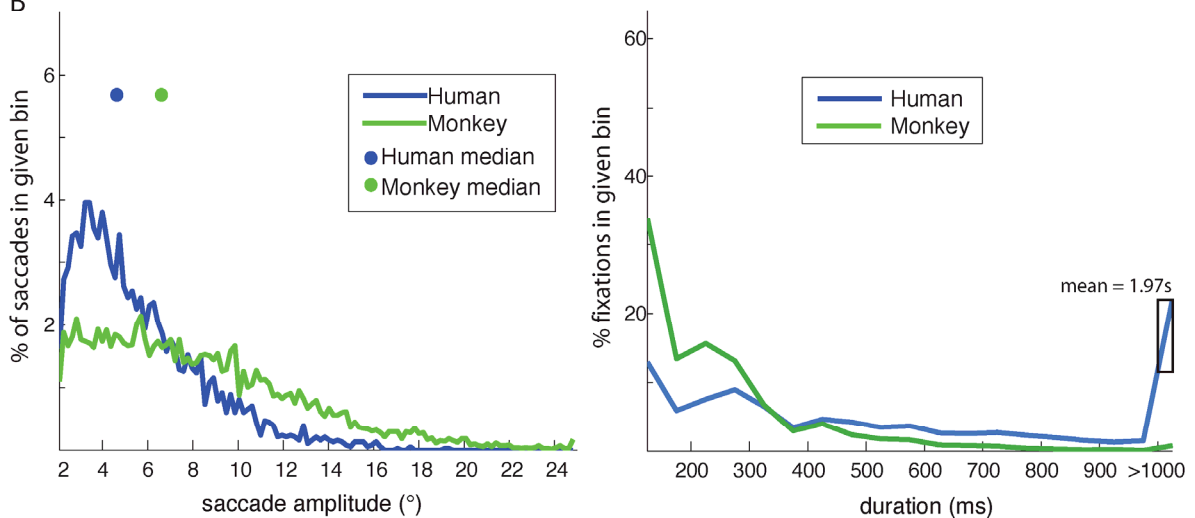


Figure S3. Low-Level Gaze Metrics

(A) Average primate gaze distributions for humans and monkeys over all frames and videos indicate a strong tendency to fixate near the center of the screen.

(B) Left panel: saccade amplitude distribution for monkeys (green) and humans (blue), plotted as the percent falling in each 6 pixel ($\approx 0.15^\circ$) bin, with saccade speed threshold $20^\circ/\text{s}$ and data cleaning as described in the text. Right panel: fixation duration distribution for monkeys (green) and humans (blue), plotted as the percent falling in each 50 ms bin; 20% of human fixations lasted over 1 s, with a mean of 1970 ms. Humans exhibited a greater central bias, made fewer and shorter saccades, and regularly fixated the screen for long periods with minimal eye movement. Because of the low spatiotemporal precision of our gaze tracking system, and because smooth pursuits may confound the discrimination of saccades from fixations, these data may somewhat underestimate the frequency of small eye movements and register adjacent fixations as single events.

Supplemental Experimental Procedures

Stimulus Presentation

We extracted videos under the fair use doctrine of United States copyright law using DVD Rip Master Pro (www.mcfunsoft.com), and converted to silent gray-scale videos using Adobe Premiere Pro 3.0 (www.adobe.com). Videos were encoded in the Xvid (www.xvid.org) codec at 30 Hz; video frames occupied an area of 770x584 pixels. All stimuli were presented with the Neurobehavioral Systems Presentation 12.2 (www.neurobs.com). Human participants took part in one session (two repetitions of each video, total), while the monkeys took part in two sessions separated by several weeks (four repetitions, total). Video stimuli were extracted from:

Chaplin, C. (Producer & Director). (1931). *City Lights* [Motion picture]. Century City, CA: United Artists.

Disney, W. (Producer), & Reitherman, W. (Director). (1967). *The Jungle Book* [Motion picture]. Burbank, CA: Walt Disney Home Video.

Salisbury, M. (Producer), & Attenborough, D. (Host). (2003). *The Life of Mammals: Social Climbers* [Motion picture]. Bristol, UK: British Broadcasting Corporation (BBC).

Quantifying Eye Movement Behavior

Raw monkey eye movement data were extracted with EYENAL; a software tool provided by ASL for use with the ASL Eye-Tracking system. Raw human eye movement data were recorded via the VisionSpace 1.0 extension linking Tobii and NBS Presentation (<http://www.vision-space.at>). To allow the gaze tracking signal to stabilize, data for the first 30 frames (1 s) of each video were discarded. Eye data were then loaded in MatLab (www.mathworks.com) for further analysis.

Saliency Map and Simulated-Gaze Generation

We applied the saliency map model using publicly available software using available documentation. We used the intensity, orientation, flicker, and motion feature channels under three normalization schemes (“surprise”, “fancy”, and “maxnorm”) to produce three alternate sets of scale-four saliency maps and simulated gaze coordinates; all other settings were left at default. Because the “fancy” scheme produced scanpaths that best matched observed primate behaviors, the analyses reported herein used “fancy” scanpaths. The “surprise” maps and scanpath were generated using the command:

```
ezvision --T, --movie, --in=/file/path.avi, --out=png:/file/path/##, --vc-chans=IOFM, -  
-maxnorm-type=Surprise, --vc-type=Surp, --gabor-intens=20.0, --direction-sqrt, --  
display-map-factor=1e11, --vcx-outfac=5.0e-9, --display-eye=yes, --display-larger-  
markers, --pixperdeg=40.0, --fovea-radius=40, --save-vcx-output=yes
```

while the “fancy” and “maxnorm” maps and scanpaths were generated by substituting "Fancy" or "Maxnorm" for "NORMALIZATION" below:

```
ezvision --T, --movie, --in=/file/path.avi, --out=png:/file/path/##, --vc-chans=IOFM, -  
-maxnorm-type=NORMALIZATION, --display-eye=yes, --display-larger-markers, --  
pixperdeg=40.0, --fovea-radius=40, --save-vcx-output=yes
```

Simulated gaze coordinates were then extracted from output frames using custom MatLab code.