Monkeys at the Movies: What Evolutionary Cinematics Tells Us about Film

Asif A. Ghazanfar and Stephen V. Shepherd

Abstract: Because the visual neuroanatomy and neurophysiology of monkeys are largely similar to ours, we explore the hypothesis that the same cinematographic techniques that create a visual scene for us likely create one for these close kin. Understanding how monkeys watch movies can illuminate how film exploits the capacities we share with our simian relatives, what capacities are specific to humans, and to what extent human culture exerts an influence on our filmic experience. The article finds that humans and monkeys share a basic capacity to process sensory events on the screen. Both can recognize moving objects and acting individuals, and both prefer looking at motion pictures of social behaviors over static images. It seems clear that some of the same things that make movies "work" for human brains also work for the brains of our nonhuman relatives—excepting two critical features. First, humans appear to integrate sequential events over a much larger time frame than monkeys, giving us a greater attunement to the unfolding narrative. Moreover, humans appear to have special interest in the attention and intentional states of others seen on the screen. These states are shared through deictic cues such as observed gazing, reaching, and pointing. The article concludes that a major difference in how humans and monkeys see movies may be declarative in nature; it recognizes the possibility that movies exist as a means of sharing experience, a skill-set in which the human species has specialized and through which humans have reaped unprecedented rewards, including the art of film.

Keywords: cognitive evolution, episodic memory, eye movements, looming, primate evolution, uncanny valley

The first thing we see is a close-up of a chimpanzee's face: its prominent lip lines pouting, looming closer to lap at the camera lens, then a brief glimpse of startlingly orange eyes. Abruptly, we stare at the grassy ground of the enclosure: hindfeet pace toward us, and with a dark-haired interruption of knee and belly, we are again lifted, bobbing and skewing upward to the jungle gym and an approaching chimpanzee. Another cut and we stare upward at the blue, while a

chimpanzee confidently strides across rope gangways. We roll about, dizzied, suddenly looking down through a netted nest, until we settle on elongated knuckles resting on a wooden strut. Their complement raps solidly against the camera case, then again the whiskered loom and we cut to two chimpanzees close together, grooming in embrace. The curtain descends as one last pair of lips looms toward the smudged lens. The film is over.

A film made by chimpanzees, created as part of the Chimpcam Project, and its accompanying documentary aired by BBC Two (*Natural World: The Chimpcam Project,* 2010), aimed to bridge primate worlds through the medium of film by giving the chimps a camera. Setting aside the fact that the edits—and certainly the score—bear witness to a human hand, the whole notion feels daft. What does a chimpanzee know about film? We would be about as surprised by a chimp blockbuster as we would by a troupe of monkeys typing out Shakespeare. Humans are the only species capable of creating—or arguably, of consuming—a film.

Films affect us powerfully through carefully coordinated images, music, dialogue, lighting, sound, and special effects. Arranged properly, these can elicit deep feelings, help us to better understand our own lives, our place in the world, the lives of those around us, and even how societies and cultures come together. Film is not a simple transcription of the real world. While the real world is spatially and temporally continuous, film is a mix of short shots, from multiple perspectives, with camera jumps between. One strong hypothesis for why film can affect us so vividly is that it is structured to match and exploit our cognitive and perceptual dispositions (Cutting 2005). For example, the "Hollywood style" of filmmaking is one whose main goal is to subordinate all sensory-perceptual aspects of the presentation to the narrative. That is, all manipulations of the camera, lighting, editing, and sets should be transparent to the viewer; they should go unnoticed (Cutting 2005).

Before Hollywood style editing, the most important feature that we *don't* notice is that the movies are composed of discrete frames presented in rapid sequence. Standard film is usually interrupted seventy-two times per second, with an episcotister flashing each of twenty-four frames three times. This brings the flicker rate above the normal human threshold, which for a bright light is about sixty times per second. Thus, continuity of motion is achieved in part by exceeding the limits of the human visual system. Because our visual system evolved to guide three-dimensional behavior based on two-dimensional retinal images, we seem to naturally register these fused frames as a window into a scene, and to interpret objects appearing at steadily changing places across sequential frames as if they move—if not in our immediate environment, then at least in some comprehensible space.

Humans are hardly unique in this respect. Many animal species use their visual systems to navigate the world. Among our primate relatives, only those

lineages that split from our own within the last 23 million years—the Old World primates (Raaum et al. 2005)—reliably share our rich sense of color vision. Because the visual neuroanatomy of Old World monkeys, apes, and humans is largely similar (Kaas and Preuss 1993; Krubitzer and Kaas 1990), we share similar flicker fusion thresholds (De Valois et al. 1974), and, as result, the same cinematographic techniques that create a visual scene for us may also create one for these close kin. To other non-primate species, these visual scenes may be much less compelling: no dog can follow Dorothy in toto into Technicolor, because they lack the color vision for which Technicolor is devised and because they have much higher flicker-fusion thresholds than do primates (Coile, Pollitz, and Smith 1989). Dogs thus require a much faster frame rate to see continuous on-screen motion. This suggests that like humans, Old-World monkeys and apes really *see* movies. But what is it that they see? What can a monkey's responses to film tell us about the way our own view of the world has evolved?

Movies are thought to exploit the properties of our visual system, and monkeys essentially have an identical visual system to ours. Thus, the fundamental question we explore below is: If a closely related species of primate has the same basic visual capacities, then do they respond to movies in the same way? Understanding this question will illuminate to what extent movies exploit pre-existing and evolved capacities shared by other primates versus capacities that are unique to humans. It may also shed light on the importance of culture in shaping our responses to film content. Below, we review most of what we know about how monkeys see moving pictures on a screen. There is a paucity of studies investigating how monkeys or other nonhuman primates watch actual, studio-produced films, so we are forced to piece together what we know from more experimentally controlled approaches involving dynamic faces, social scenes, and objects presented on a two-dimensional screen. Stitched together, these studies reveal what information monkeys can and cannot extract from the medium. We will begin with how different features of film (moving objects, actors and their interactions, and computer animations) are perceived by monkeys. We then review results from the very first study to compare movie-watching between monkeys and humans, followed by an outline of what cognitive capacities we think distinguish human versus monkey movie-watching.

Seeing Moving Objects: Lumière and Looming

Motion is the reason why films exist. One of the first "feature" films to truly exploit this was the fifty-second film by Louis Lumière, *L'Arrivee d'un train a La Ciotat* (The arrival of the train at La Ciotat). The film documents a simple, everyday occurrence: a train pulling into a station. The camera is set very close to the train as it approaches, generating the sensation that the train is com-
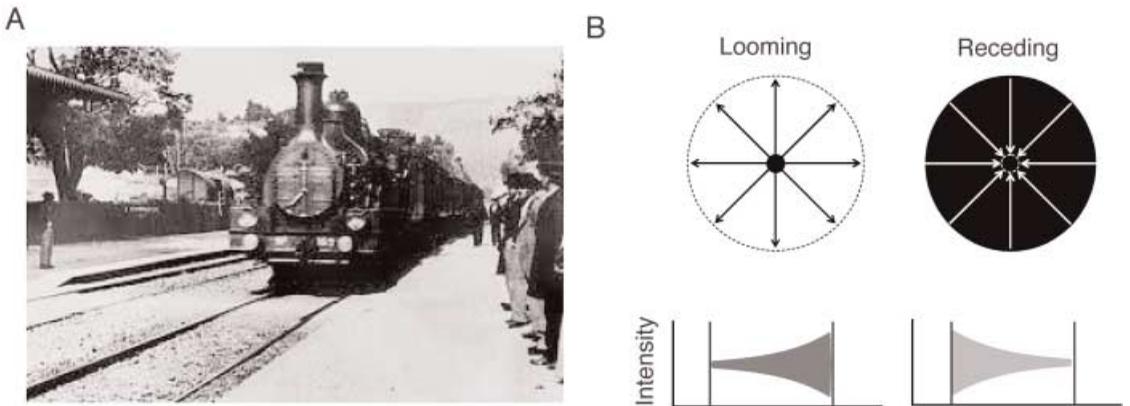
ing directly toward the audience (Figure 1A). The reaction of people watching this film for the first time was, according to legend, one of fear and panic. Members of the audience jumped up from their seats and ran away from the screen toward the back of the room. Released in 1895, it was an immediate hit with thousands of people paying a franc to see it. Over the years, the legend of the audience's reaction has been embellished and the film is the example par excellence of the manipulative power inherent of cinema: the very first movie-goers could not distinguish between reality and film.

This legend, repeated in nearly every history of film, was dismantled by Loiperdinger (2004), who argues that the moving images projected onto the screen with the Cinématographe Lumière (the Lumière movie projector) could not be mistaken for reality. There are several reasons for this. First, at that time, the speed of the movie projector was too slow to completely eliminate the flickering of light as successive frames were shown. Second, the film was in black and white. Third, the power of the projector's light was only sufficient to project the image on a small screen. Finally, there was no sound accompanying the film. These problematic characteristics of the film and its projection suggest that the audience was not likely to mistake the approaching train for something real. Further, as there are no contemporary accounts of panic reactions among the audience, Loiperdinger (2004) concludes that the oft-repeated anecdote, suggesting the audience felt physically threatened by the train and therefore panicked, must be regarded as a historical fantasy. Actually, the truth of how the Lumières' first audience reacted may lie somewhere in between running in panic and coolly remaining seated.

In addition to trains, the world is full of potentially dangerous approaching objects for humans and other animals. For our ancestors in the primate lineage, a raptor's stoop or pursuit by a dominant individual was a clear and present danger of everyday life. Such dangers remain for primates living in the wild. To survive in such a world, animals must escape or avoid these dynamic sources of danger by rapidly detecting and responding to the signals that specify such *looming* events. Because of their rapid transmission in the environment, visual and auditory signals can each independently be a reliable cue to fast approaching sources. Given their importance for survival, it stands to reason that both humans and other primates should be exquisitely sensitive to looming events, regardless of their exact form. This idea is supported by numerous behavioral studies.

Like the film of an arriving train, other looming events can be simulated and even very simple stimuli can elicit adaptive avoidance responses (Tinbergen 1951). Indeed, in humans, any retinal image that changes in size creates a compelling illusion of motion in depth (Wheatstone 1852). There are many studies suggesting that motion illusions that mimic object-approach (looming) are treated with priority over those that mimic receding objects. For ex-

ample, looming stimuli capture attention better than those that simulate receding objects (Franconeri and Simons 2003). Human infants, even those as young as 2 weeks old, respond with fear when presented with a rapidly expanding circular shadow presented on a screen: widening their eyes, withdrawing their head, and raising their arms (Ball and Tronick 1971; Figure 1B). Rapidly contracting shadows, simulating a receding object rather than an approaching one, do not elicit any such response.

If responding to looming events is adaptive, then it stands to reason that monkeys should respond similarly (Schiff, Caviness, and Gibson 1962). When presented with the same looming shadow used with human infants, both adult and infant monkeys abruptly retreated or "ducked." Their responses usually involved springing to the back of their cage, and sometimes producing alarm calls. Indeed, they responded very much like the human audience did in the legend about *L'Arrivee d'un train a La Ciotat*! Furthermore, the monkeys did not habituate to the looming stimuli nor did they respond in the same way to other types of simulated visual motion events. Taken together, these data suggest that humans, human infants, and monkeys readily respond with defensive reactions to looming events that are simulated on the screen. These responses were elicited by impoverished stimuli consisting of a simple, expanding, black-and-white disk. Thus, although it is not certain exactly how Lumière's audience responded in 1895, it is completely reasonable that an "unrealistic" looming train presented on the screen could have elicited reflexive reactions that are symptoms of panic.

After the invention of moving pictures that exploit the limitations of the primate visual system, the second most important event in film history was the introduction of sound. Presenting realistic sounds in the theater has its own set of problems that need to be solved to accurately convey auditory events in a movie. The solutions often involve exploiting the properties of the auditory system and its relationship to vision. Looming events can illustrate how this is done. Sound in theatres relies on speakers that are in a different lo-

cation than the screen. This means that for events on the screen that generate sound, the sounds an audience hears are actually coming from a different location than from where the event is seen. For auditory looming events, there are three cues that help us identify the speed and distance of a moving object: 1) the arrival time differences between our two ears; 2) the Doppler effect (as a sound source moves forward there is a shortening of the wavelength of waves in front of it and an increase in wavelength of waves behind it); and 3) amplitude (intensity) change (Rosenblum, Carello, and Pastore 1987). Luckily, it turns out that the dominant cue indicating an approaching object is rising intensity (the sound gets louder as the object approaches; Rosenblum, Carello, and Pastore 1987). This means auditory looming events are easy to simulate, and they have many of the same consequences as visual looming (Neuhoff 1998). When human adults are asked to predict the arrival time of an intensifying sound, they repeatedly err on the side of safety, expecting contact before the source actually arrives. The same is not true for falling intensity sounds that mimic receding sound sources. For the same adaptive reasons cited for visual looming, monkeys also have a strong perceptual bias toward rising versus falling intensity auditory signals (Ghazanfar, Neuhoff, and Logothetis 2002).

It is easy to imagine that if *L'Arrivee d'un train a La Ciotat* had been accompanied by a sound track that the audience response would have been even stronger. Loiperdinger (2004) might reasonably argue to the contrary, that the location of the sounds—from speakers on either side of the screen—makes the film even less realistic. However, the primate brain uses temporal coincidence to override spatial disparities between different sensory inputs (i.e., auditory versus visual). The temporal coincidence binds the auditory looming event heard from the speaker and visual looming event seen in two dimensions on the screen together as one event. Our reactions to this bimodal looming event are stronger than to a looming event presented in either modality alone (Cappe et al. 2009). The same is true for monkeys (Maier et al. 2004).

The behavioral data from simulated looming experiments demonstrates that both humans *and* monkeys readily respond in an adaptive manner to moving objects presented on a screen or through a speaker. That is, film and its sound track can effectively elicit responses similar to real world events. Both humans and monkeys appear to understand that movies are scenes, and that the depicted audiovisual elements interact with one another—but we also understand that this scene is somehow self-contained and separate from our own. Because of this strange sense of self-containment, what we take away from movies depends crucially on how these audiovisual elements interact with each other. The action, that is to say, comes from the actors.

Both humans and monkeys appear to understand that movies are scenes, and that the depicted audiovisual elements interact with one another.

Seeing Others on Screen

One of the big emotional draws to cinema is the appeal of watching other people behave—or better yet, misbehave. Humans see actors interact with their surroundings and with each other, see these actions as categorical and consequential, and, in seeing these behaviors, draw conclusions about the depicted individuals. A movie may leave us laughing hysterically or wracked with tears, smiling with pride or morose and lost in thought. We feel hope and judgment, pleasure and catharsis, as if we ourselves were taking part in the events on the screen. In the end, man is the measure of all things, and we understand movies in terms of the human experiences they convey.

Over seventy-five years ago, Heinrich Kluver (1933) suggested that a study of monkeys' responses to motion pictures could provide interesting and valuable information about their emotional and perceptual world. He exposed several species of primates to motion pictures and, surprisingly, did not find much evidence that they were even attending to the film, let alone responding to its content. Since that time, however, many studies have more carefully examined monkeys' responses to movies and have shown that, on the contrary, monkeys can recognize other monkeys on the screen. Here, "movie" actually refers to a live video feed projected onto a two-dimensional screen, not a film constructed with shots, cuts, and scenes. Butler (1961) showed that monkeys much prefer to look at motion pictures than static pictures, and that movies with well-articulated figures were preferred over ones with amorphous, out-of-focus scenes. This suggests that motion pictures capture more attention than motionless ones, but that motion alone is insufficient: content matters. More recently, Rosenblum and colleagues demonstrated that monkeys will *pay* (in the form of foregoing a food treat) to watch movies of a monkey group (Andrews, Bhat, and Rosenblum 1995). In this experiment, monkeys used a joystick to move a cursor over one button on the screen that would result in a food reward or over another button that would result in a short movie. Monkeys more often chose to watch the movie showing a group of monkeys. After they tired of watching one group of monkeys, they would pay more again to watch another group (Andrews and Rosenblum 2001).

Do monkeys have emotional reactions to who is on the screen and what they are doing? Absolutely. Young juvenile monkeys will respond with affiliative facial expressions upon seeing an unfamiliar large, dominant male monkey producing a threatening gesture on screen (Plimpton, Swartz, and Rosenblum 1981). They did not do so when the same monkey on the video did not produce a threat. Along the same lines, how a monkey might respond to another monkey seen on a screen depends on its personality (Capitanio 2002). Monkeys differ in their sociability: some monkeys often approach other individuals, produce affiliative gestures, and participate in mutual grooming; other monkeys do so rarely. When watching videos of other monkeys, low-

sociable monkeys tended to sit and stare. They yawned often and did not re-act to what was seen. In contrast, high-sociable monkeys tended to react as though the agent on the screen was real. They tended to look away from the screen and were, in general, much more active during the movie presentation.

It is one thing to engage an agent on a screen as though communicating with it and quite another to understand that two depicted monkeys commu-nicate with each other. Monkeys seem to understand that animals interact and that these interactions have consequences. There is evidence that, after a certain amount of training, monkeys understand the third-party relationships between monkeys presented via video and with whom they are otherwise un-familiar. Most monkeys live in linear dominance hierarchies, and knowing dominance relationships among other monkeys is important for navigating their social worlds. They can recognize depicted individuals and their rela-tionships (Ghazanfar and Santos 2004), and can interpret their interac-tions in terms of species-typical behaviors, identifying the dominant party in a behavioral exchange and generalizing across various types of dominance behavior (Bovet and Washburn 2003). Using videos, monkeys can be trained to pick out the dominant individual in a pair after one type of interaction with a subordinate (Bovet and Washburn 2003; Paxton et al. 2010); moreover, they spontaneously pick out the same dominant individual in new, previ-ously unseen videos depicting different types of interactions with subordi-nates. This suggests that monkeys can use videos to learn about individuals and remember this information across what are, in effect, different movies scenes.

Though monkeys are not sophisticated imitators—and though they share less well than your average preschooler—they do seem to empathize with an-imals seen onscreen. Like humans, monkeys can "catch" yawns, and do so from videotaped conspecifics (Paukner and Anderson 2006); similarly, mon-keys are more likely to scratch after seeing other monkeys attend an itch (Nakayama 2004). Moreover, monkeys seem to experience vicariously at least some of the emotions they see depicted in film. In a series of experiments per-formed a half century ago, during the rule of behaviorism, Robert Miller and colleagues trained monkeys to pull a lever. One variant of this experiment is particularly enlightening (Mirsky, Miller, and Murphy 1958). Subjects were trained to pull a lever to avoid a shock whenever they saw a particular mon-key on a screen. After they learned to do this, their training was "extinguished" by showing the same monkey but without ever delivering any shocks. Now we get to the interesting part: Once the conditioned response was so thoroughly extinguished that three days had passed with no lever pull, they showed sub-jects another monkey on the screen—which they shocked. The subjects im-mediately pulled the lever to stop the shock, despite the fact they were not threatened with shock themselves. It is possible that the monkeys were act-

ing altruistically—when subjects pressed the lever, both the view of the monkey (and the shock thereto) disappeared. More likely, the monkeys felt some emotional contagion from the observed monkeys—seeing a fearful monkey, the subject monkeys became slightly frightened themselves, and this reawakened the latent conditioned response. This indicates that, to at least a small extent, monkeys "empathize" with the monkeys they see onscreen.

Computer Animation and the Uncanny Valley

Humans have a very reliable response to simulated humans: if they appear to be too realistic, audiences are creeped out. This is known as the "uncanny valley" response. Movies can transport us to different worlds, where the characters are like us in some ways but different in others, or they can be wholly foreign. Originally, monsters and aliens were made using costumed or heavily made-up humans or stop-motion animation involving models. In *King Kong* (1933), Kong was created using rubber and rabbit fur combined with stop-motion animation; the original *Godzilla* (1954) was a human in a suit destroying a scale-model of Tokyo. In these worlds, characters are made to do the humanly impossible or act in an alternative reality. With the advent of computer-generated imagery (CGI), we have inexorably moved from adding some scenes with special effects—like the "Genesis effect" in *Star Trek II: The Wrath of Khan* (1982)—to movie scenes depicting full-blown new worlds and characters all in purely digital formats (Cook 2004). The stained-glass knight in *Young Sherlock Holmes* (1985) was the first computer-generated character (Cook 2004), and since then many movies feature better and more fully realized synthetic agents. One might think that it is only a matter of time that real human actors will be replaced altogether with computer-generated ones. There is one major problem, however. It is nearly impossible to create with computer animation a realistic-looking human agent without creeping out the audience.

> *It is nearly impossible to create with computer animation a realistic-looking human agent without creeping out the audience.*

It is natural to assume that as computer-generated synthetic agents come closer to resembling humans, they will be more likely to elicit behavioral responses similar to those elicited by real humans. However, this is only true up to a point. Increased realism does not necessarily lead to increased acceptance. If agents become too realistic, people find them emotionally unsettling. This feeling of eeriness is known as the "uncanny valley" effect and is symptomatic of entities that elicit the concept of a human but do not meet all the requirements for being one. The label is derived from a hypothetical curve proposed by the roboticist Mori (1970), in which agents having a low resemblance to humans are judged as familiar with a positive emotional valence, but as agents' appearances become increasingly humanlike, the once positive familiarity and emotional valence falls precipitously, dropping into a basin of

negative valence (Figure 2A). Although the effect can be elicited by still images, Mori further predicted that movement would intensify the uncanny valley effect. Support for the uncanny valley has been reported anecdotally in the mass media for many years (e.g., reports of strong negative audience responses to computer-animated films such as *The Polar Express* [2004] and *The Final Fantasy: The Spirits Within* [2001]), but there is now good empirical support for the uncanny valley based on controlled perceptual experiments (MacDorman et al. 2009; Seyama and Nagayama 2007).

Despite the widespread acknowledgment of the uncanny valley as a valid psychological phenomenon (Geller 2008; Kahn et al. 2007; MacDorman and Ishiguro 2006; Walters et al. 2008), there are no clear explanations for it. There are many hypotheses, and a good number of them invoke evolved mechanisms (MacDorman et al. 2009; MacDorman and Ishiguro 2006). For example, one explanation for the uncanny valley is that it is the outcome of a mechanism for pathogen avoidance. In this scenario, humans evolved a disgust response to diseased-looking humans (Rozin and Fallon 1987), and the more human a synthetic agent looks, the stronger the aversion to perceived visual defects that presumably indicate the increased likelihood of a communicable disease. Another idea posits that realistic synthetic agents engage our face processing mechanisms, but fail to meet our evolved standards for facial aesthetics. That is, features such as vitality, skin quality, and facial proportions that can enhance facial attractiveness (Jones, Little, and Perrett 2004; Rhodes and Tremewan 1996) may be deficient in synthetic agents and so elicit the uncanny valley effect. Thus, a computer-generated agent with pale skin might appear either unhealthily anemic or merely extremely unattractive. Regardless of the specific underlying mechanism, in all cases of the uncanny valley effect the appearance of the synthetic agent somehow falls outside the spectrum of our expectations—expectations built from everyday experiences with real human beings.

Now we can ask the following question: if realistic computer-generated avatars fail to properly exploit or trick the human visual system, must we assume monkeys, who share the same visual capacities and limitations, should also be susceptible to the uncanny valley response? We tested whether this was true (Steckenfinger and Ghazanfar 2009). Like human studies (MacDorman et al. 2009; Seyama and Nagayama 2007), we compared monkey preferences to differently rendered individuals: actual monkey faces, realistic synthetic monkey faces, and unrealistic synthetic monkey faces (Figure 2B). The facial expressions for all three included a "coo" face, a "scream" face, and a "neutral" face. Just as humans tend to look longer at attractive versus unattractive faces (Maner et al. 2007), monkey preferences were measured using both the number of fixations made on each face as well as the duration of fixation (Humphrey 1974). If monkeys are prone to the uncanny valley effect, they
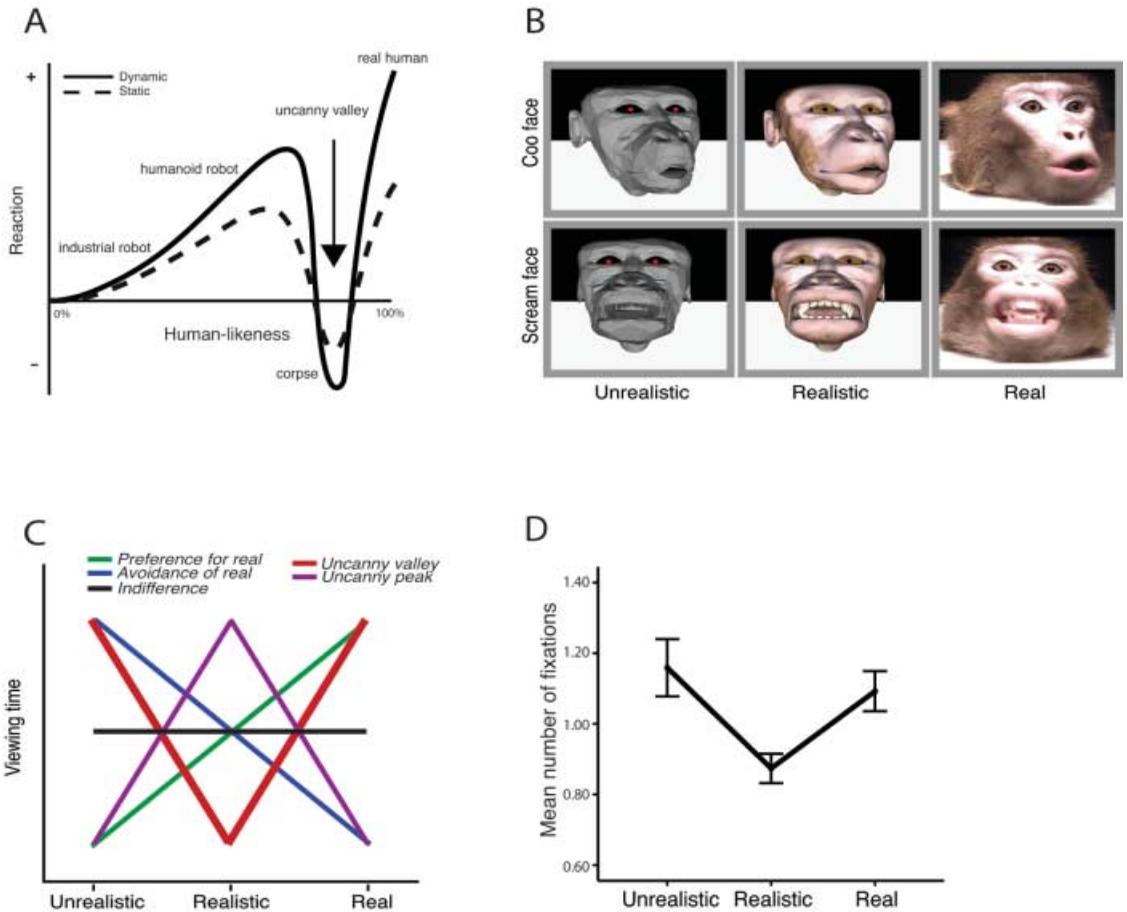
Figure 2. Uncanny valley in monkeys. A. Hypothetical uncanny valley curve. B. Monkey versions of unrealistic synthetic agents, realistic synthetic agents, and real agents. C. Possible outcomes for how long monkeys might view each of the agents. D. Looking time data (as measured by the number of fixations) of monkeys viewing the agents in panel B.

should prefer to look at both unrealistic synthetic faces and actual faces more than at the uncannily "realistic" faces (Steckenfinger and Ghazanfar 2009).

Five behavioral outcomes were possible in this experiment (Figure 2C). The monkey subjects could exhibit a preference for actual faces (green line), an avoidance of actual faces (or a preference for the unrealistic faces; blue line), an uncanny valley effect (a decreased preference for "realistic" faces relative to the two other types; red line), an uncanny peak (a preference for "realistic" faces; purple line), and finally, no preference at all (or general lack of interest, black line). Surprisingly, given these numerous possible outcomes, the monkeys consistently exhibited the uncanny valley effect (Figure 2D), preferring to look at the unrealistic synthetic and actual faces more than at the "realistic" faces. Indeed, all five of our monkey subjects displayed this pattern of looking preference.

In summary, out of five possible patterns of looking preferences toward faces with different levels of realism, monkeys exhibit one pattern consistently: they preferred to look at unrealistic synthetic faces and actual faces more than at "realistic" synthetic faces (Steckenfinger and Ghazanfar 2009). The visual behavior of monkeys evinces an uncanny valley just the same as human visual behavior does (MacDorman et al. 2009; Seyama and Nagayama 2007). Thus, these data demonstrate that the uncanny valley effect is not unique to humans and that evolutionary hypotheses regarding its origins are tenable. The same processes that make the human visual system acutely sensitive to defects in computer-generated avatars are also operational in monkeys. What cannot be discerned in our experiment is whether the monkeys experience disgust or fear (or aversion at all) when they avoid looking at the uncanny faces.

As in humans (MacDorman and Ishiguro 2006), if monkey face perception exhibits an uncanny valley it may be because their brains are, to a greater degree than with unrealistic synthetic agents, processing the realistic synthetic agents as conspecifics—conspecifics that elicit, and fail to live up to, certain visual expectations. Importantly, it is not the increased realism that elicits the uncanny valley effect, but rather that this increased realism lowers the tolerance for abnormalities (Green et al. 2008). Understanding the features that exceed this tolerance threshold is a topic of intense investigation, particularly for those interested in computer animation (Geller 2008; Walters et al. 2008).

One likely possibility is that the computer-animated faces simply cannot capture all the rapid and subtle movements of the face, and thus violate our strongly felt social expectations (MacDorman and Ishiguro 2006). How such expectations are constructed is not known, but it is likely driven by social experience in both monkeys and humans (Lewkowicz and Ghazanfar 2009). To test this hypothesis in humans, we showed all possible pairs of a human face, realistic avatar face, and an unrealistic avatar face to groups of 6-, 8-, 10-, and 12-month-old infants across different experiments and recorded visual preferences (Lewkowicz and Ghazanfar 2011). Consistent with the developmental hypothesis, we found that the uncanny valley effect first begins to emerge at twelve months of age, suggesting that perceptual experience with the human face prototype is critical to its emergence. The same developmental process is likely to occur in monkeys.

The Eyes of the Beholder: How Monkeys and Humans Watch Movies
The story that we have presented so far is that humans and monkeys share a basic capacity to process sensory events on the screen: they can recognize moving objects and acting individuals. They prefer looking at motion pictures of social behaviors over static images, and will even pay (forgoing food) to do so. Even subtly nuanced properties of human responses to movies (the un-

canny valley) are shared with monkeys. It seems clear that some of the same things that make movies "work" for human brains also work for the brains of our nonhuman relatives.

Brains evolved to guide sensorimotor behavior within an immersive, inter-active, ever-changing environment. Movie-viewing offers a kind of minimal model of these real-world interactions, in that viewers' perceptions crucially drive and depend upon ongoing orienting behaviors. This is because each eye movement changes the pattern of light falling on the retina, and much of the visual brain is retinotopic in character. Our eye movements while we watch a film constrain what we notice, and the things we notice are used to decide where next to look. Each time we watch a movie, there is a legitimate sense in which we are *interacting* with it. This suggests that understanding an audience's experience is more art than science. There are a million ways to watch a movie; how could we possibly understand them? But just as keen cine-matographers can skillfully manipulate their audiences, scientists can care-fully analyze gaze behavior to learn how movies get watched. Such studies reveal that we humans tend to watch the same movies in the same way (Goldstein, Woods, and Peli 2007; Hasson, Landesman, et al. 2008; Hasson, Yang, et al. 2008; Nakano et al. 2009).

We compared how monkeys, humans, and a computational model re-sponded to watching identical movie clips multiple times as measured by gaze patterns (or scanpaths) (Shepherd et al. 2010a,b). The model had proven effective at locating areas of interest in static scenes based on low-level visual cues such as lighting, contrast and motion (Berg et al. 2009; Itti, Dhavale, and Pighin 2003; Itti and Koch 2000). The use of the model allowed us to deter-mine to what extent similarities between human and monkey gaze patterns were driven solely by low-level cues versus higher-order features such as be-haviorally relevant objects, actions, or narrative content. Clips were taken from three films. One movie featured monkeys in natural environments (the BBC's *Life of Mammals* [2002]), one featured cartoon humans and animals (Disney's *The Jungle Book* [1967]), and one featured human social interactions (Chaplin's *City Lights* [1931]). The movie clips were 3-minutes in duration, converted to black-and-white, and stripped of any sound. Each subject viewed each movie clip multiple times in random sequence (see the video of this investigation, http://www.vimeo.com/7736755). Comparing humans and monkey eye move-ments, we found that gaze behavior during movie viewing was significantly correlated between repetitions, individuals, and species. Individuals tended to look at the same spot on the screen at the same time as other individuals, whether human or monkey (Figure 3A). Gaze sequences were thus signifi-cantly correlated across different viewings by humans and monkeys. These correlations were especially pronounced among humans, consistent with past reports (Hasson, Landesman, et al. 2008): as far as eye movements go, we

are as similar to our neighbors in the theatre as to ourselves in a repeat show-ing. Correlations between monkeys were also significant, if substantially lower than among humans, and correlations between species were comparable (Figure 3A). In the end, humans directed the highest-sensitivity part of their eyes toward the same image 70 percent of the time between repetitions and 65 percent between individuals. By contrast, monkeys did so 33 percent be-tween repetitions and 27 percent between individuals—a smaller but still sig-nificant amount; and finally, monkeys and humans overlapped 31 percent of the time (Figure 3B). These findings suggest that humans and monkeys use similar spatiotemporal visual features to guide orienting behavior, but, more

Figure 3. Human and monkey eye gaze correlations while viewing the same movie clips. A. Spatial positions 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 p = 0.05 (thick) or 0.001 (thin): all primate gaze inter-subject correlations were significant with p < .001. Artificial scanpaths produced by a low-level visual saliency model (orange) were significantly correlated with primate scanpaths in spatial position (p < 0.001) and gaze shift timing (p < 0.05); however, residual inter-primate correlations (yellow bars) were essentially unchanged despite partialling out shared similarities to artificial scanpaths. B. To confirm that behavioral correlations were driven by visual fixation priorities, we compared the percentage of gaze samples that overlapped (±3.5°) across different scanpaths. The pattern of results was identical to the pattern observed for inter-subject correlation. Furthermore, we found that samples that overlapped between humans and monkeys rarely overlapped with the best-performing artificial scanpath (2 of 31%, see inset). These data rule out the hypothesis that gaze correlations were driven primarily by low-level visual features, at least as characterized by well-established neuromorphic computational saliency models.
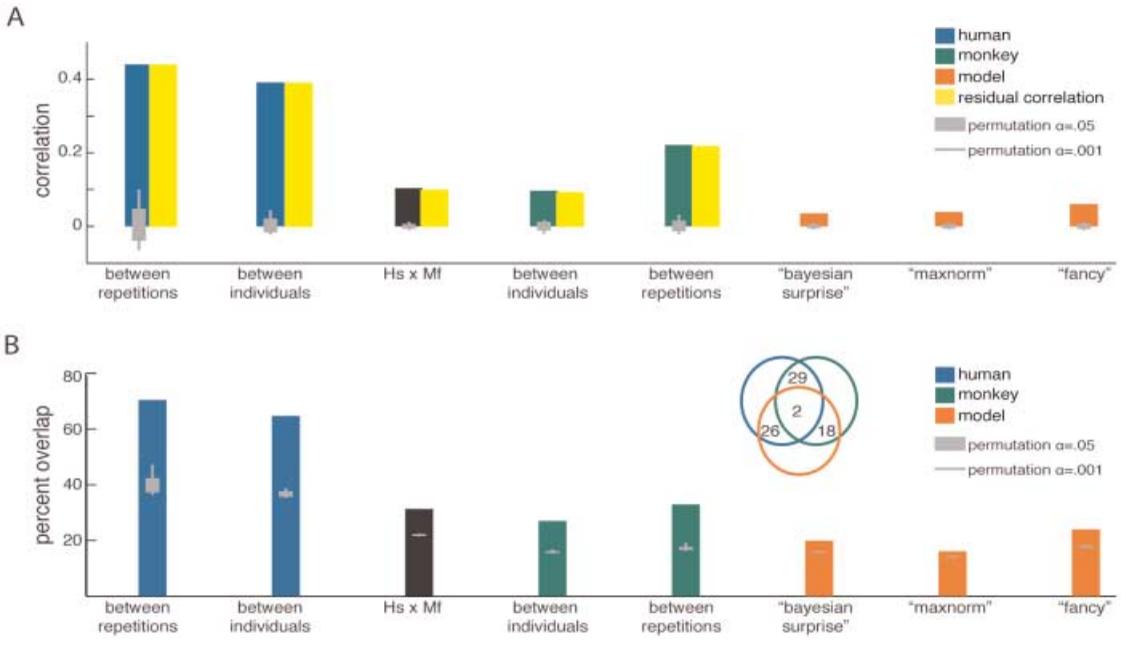
important, they suggest a strategy for comparing what types of visual features are most important to us.

As suggested above, 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 a high-level, behaviorally relevant narrative. To distinguish between these possibilities, we compared human and monkey scanpaths to artificial scanpaths generated by a well-validated low-level saliency model (Itti 2004; Itti, Dhavale, and Pghin 2003; Itti and Koch 2000; Itti, Koch, and Niebur 1998). Indeed, artificial scanpaths did correlate with human and monkey gaze positions in our experiment (Figure 3A), but even the best-correlated artificial scanpath failed to explain primate similarities: residual correlations were just as pronounced after partialling out similarities to artificial scanpaths (yellow bars in Figure 3A). In total, although the best artificial sequence overlapped with observed human and monkey gaze—28 percent and 20 percent of the time, respectively—it was strikingly poor at predicting human and monkey overlap. Of the 31 percent of samples in which human and monkey gaze overlapped, only 1 in 15 (2.1% of total samples) also overlapped the simulated scanpath (Figure 3B, inset).

The fact that computational models of low-level visual saliency accounted poorly for behavioral correlations suggests that our shared attention shifts are not driven solely by low-level similarities in the visual system, but also by shared sensitivity to complex, behaviorally relevant stimuli. Most notably, low-level models underestimated the importance of crucial biological stimuli such as faces and their expressions, bodies and their movements, and (particularly for humans) observed social signals and behavioral cues. Gaze correlations during natural viewing suggest that in the absence of explicit, immediate goals—intrinsic or instructed—orienting priorities are overwhelmingly similar (Hasson, Landesman, et al. 2008; Hasson, Yang, et al. 2008; Tosi, Mecacci, and Pasquali 1997) and focused on faces and social interactions.

Although statistically significant, inter-species correlations were low in magnitude. To explore why, we measured the extent to which a species' gaze overlapped or dispersed as a function of time. When gaze choices significantly overlapped, this suggested that the fixated feature had special importance to a species (Shepherd et al. 2010b). When gaze choices significantly diverged, this suggested that diffuse vigilance dominated focal attention toward any particular scene feature. We could thus identify scenes in which both humans and monkeys strongly attended to the same feature (Figure 4A), scenes in which both humans and monkeys were diffusely vigilant (Figure 4B), scenes in which humans and monkeys strongly attended different features (Figure 4C), and scenes in which one species' attention was captured while the other's was diffused (Figure 4D). Both humans and monkeys generally looked toward
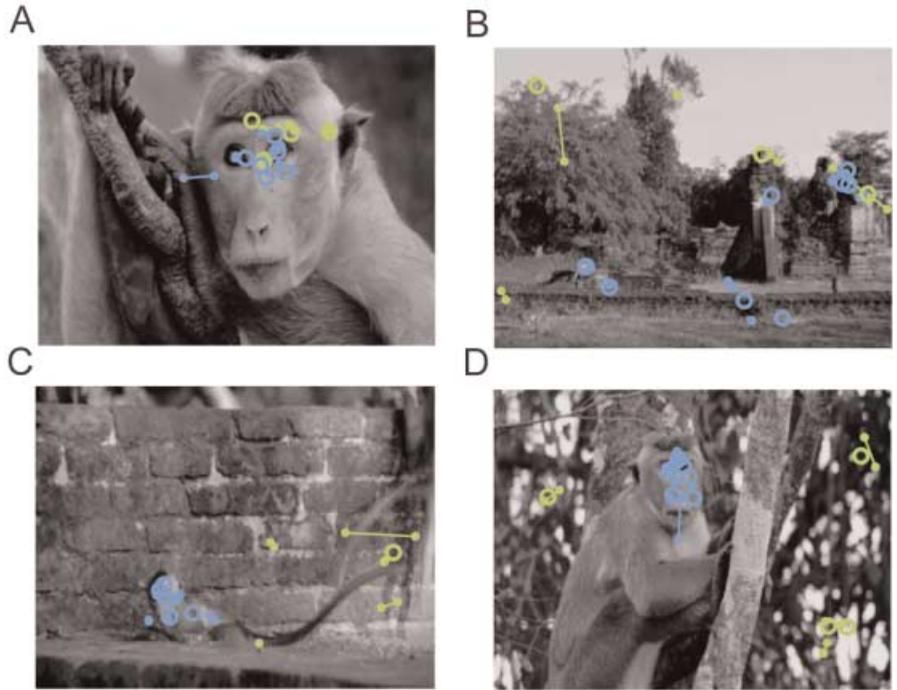
Figure 4. Comparison of gaze reliability across time, by species. Measuring the frame-by-frame grouping or divergence of scanpaths can suggest similarities and differences between human and monkey gaze toward video sequences, as illustrated in these stills from the Life of Mammals. Faces often captured the attention of all primates (A), while long shots with many different features of interest dispersed it (B). By contrast, dyadic interactions sometimes provoked consistently different macaque and human gaze selections (in C, one monkey is more active but is leaving the field of view). Finally, some features that reliably gathered attention from one species did not hold the attention of the other (D): macaques looked away more quickly than humans from social stimuli, but more avidly explored scenery appearing during a camera pan. (Green, Monkeys; Blue, Humans.)

faces and toward interacting individuals (Figure 4A); and although both humans and monkeys sometimes scanned the broader scene (Figure 4B), monkeys shifted gaze away from objects of interest more readily and more often (Figure 4D). Monkeys and humans sometimes made collectively different decisions about where to look, and these differences sometimes reflected differential understanding of movie content: for example, 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 off-screen (Figure 4C)!

There appear to be two important differences in the ways monkeys and humans watch movies, and both suggest crucial evolutionary changes in how humans view the world. First, humans appear to pay attention to related details in a movie for much longer than monkeys do, suggesting that humans integrate events over time in a fundamentally different way. Second, humans

appear to look at the focus of actor's attention and intentions to a much greater extent than do monkeys.

Gaze Fixation and Scene Integration

One reason for the low magnitude of inter-species eye gaze correlations (Figure 3A) could be due to species differences in basic eye movements. Both humans and monkeys concentrate their attention toward the center of the screen, no doubt in part because cinematographers tend to put the most important images in this central region (Shepherd et al. 2010b), but also because humans have learned to expect this to be the case—a convention uncomfortably threatened, for example, by the faux-amateur cinematography of *The Blair Witch Project* (1999). Consistent with these differences, monkeys gaze toward the peripheral regions of the screen much more often than do humans. More telling, however, is a striking difference in fixation duration between humans and monkeys. When we look at a visual scene—even, for example, when reading—we do not smoothly scan the features of the scene but fixate one at a time, sequentially (Dodge and Cline 1901). While watching movies, humans organized their fixations in a very different way than did monkeys (Shepherd et al. 2010b). Sequential fixations were likely to be grouped on similar regions of the screen, producing shorter saccades than were observed in monkeys, for whom each fixation was relatively independent. Moreover, humans generally maintained fixations far longer than monkeys did—sometimes maintaining attention at approximately the same point on a screen for several seconds.

These differences suggested that humans were integrating information across repeated fixations, selecting in each fixation a portion of the screen that seemed relevant to their current experience of the narrative. In contrast, fixations by monkeys often appeared nearly independent of one another, as if monkeys were reacting moment-by-moment instead of assembling and testing a narrative explanation for the scene before them. This idea is hard to test precisely, but with a few assumptions it is possible to investigate the notion. We can assume, for example, that *conceptually* related features are often also *spatially* related—for example, because they physically co-locate, are parts of the same object, or are parts of interacting objects. We can then measure the distance between gaze location at one point in time and gaze location at a slightly later point in the same scanpath. We do this two ways: first, we contrast the autocorrelation of each scanpath with the cross-correlation of scanpaths from different viewings by the same subject of the same movie. This allows us to see how much viewers' choices of where to look constrained their subsequent fixations: we find that for humans, these moment-to-moment correlations were significant for almost 10 seconds, while for monkeys, they fell to baseline within well under a second (Figure 5A). Because these effects might simply have been driven by humans' tendency toward extended fixa-

tions (Shepherd et al. 2010b), we also explicitly looked at the distances between distinct fixations in a sequence—immediately adjacent, once removed, twice removed, and so on—relative to the baseline distance between randomly chosen fixations. Differences are pronounced. Even though human fixations are longer than monkey fixations—meaning the scene has potentially changed more—human gaze targets are significantly spatially clustered over seven to ten sequential fixations (Figure 5B). By contrast, monkeys significantly cluster over only three sequential fixations. Only for immediately adjacent fixations are human and monkey fixations similarly related. This implies that monkeys, like humans, integrate sequential fixations into some sort of narrative—but unlike humans, they have a severely limited ability to interrelate events occurring at different times.

These data thus speak to a larger debate—and one that's far from settled—regarding the difference between human and nonhuman understanding of life events. This debate hinges on specific type of mnemonic cognition

These data thus speak to a larger debate—and one that's far from settled—regarding the difference between human and nonhuman understanding of life events. This debate hinges on specific type of mnemonic cognition called "episodic memory."

called "episodic memory": though a precise definition is elusive, episodic memory refers to the ability to recall one's past experiences in the manner in which they unfolded. (Other forms of memory, for example, might involve remembering a factoid but not where it had been learned, or finding that one can easily ride a bicycle despite not having been on one for decades.) Episodic memory is so tightly intertwined with our sense of self—and with our use of language—that it has proved difficult to adapt to a nonhuman model (Martin-Ordas et al. 2010; Suddendorf and Corballis 2010). Is episodic memory a sense of time passing? A label of what, when, and where events occurred? A self-awareness keyed to our personal history? We prefer the idea that it is a memory for narratives: for sequences of interactions among individuals, objects, and events within an environment. Our comparison of human and monkey gaze behavior suggests that many animals may share a basic ability to associate events in time, but humans can do so more easily and over a far larger range of times than can monkeys. Evidence from humans suggests that remembering our past is a kind of "mental time travel," implying that we do not passively store and recall concrete chains of events, but actively reconstruct them through some process of generalization and reassembly (Suddendorf and Corballis 2010).This understanding of episodic memory as a kind of mental time travel implies an ability to imagine events that have not yet happened; events happening to us as they happened to someone else; or even events as they could have been, but were not. Such an understanding is essential to the human experience of movies: as an integrated sequence of events happening in their own self-contained time (just as they occur in their own
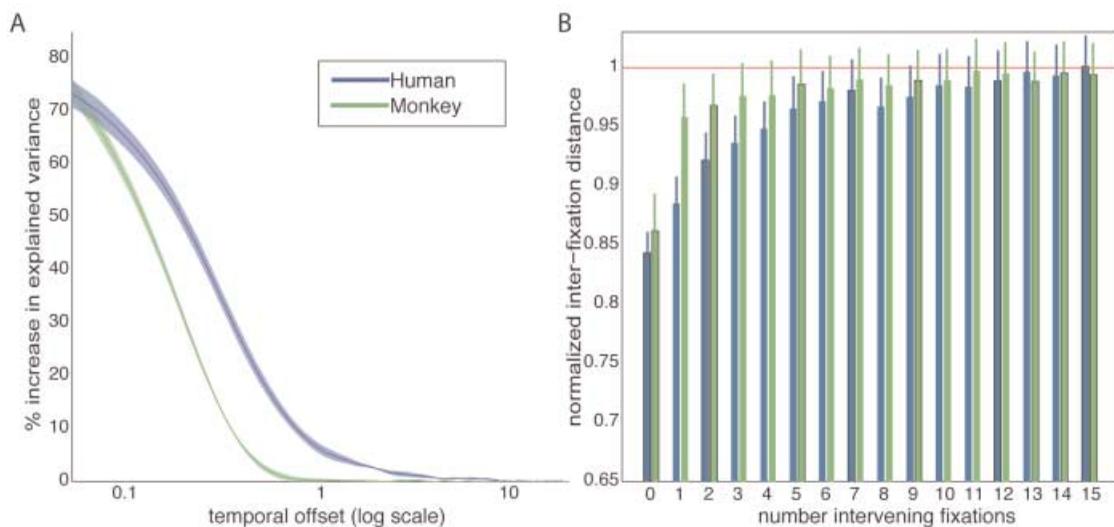
Figure 5. Integration of scene content by gaze. A. Cross-correlations of human viewings of the same movie during the same session, compared to monkey viewings of the same movie during the same session. Gaze position during the same viewing predict subsequent gaze toward nearby (generally conceptually related) targets; however, for humans this falls off much more slowly (almost 10 seconds) than macaques (less than 1 second). B. The tendency shown in A is also evident across successive fixations, which remain closer than expected across many intervening fixations (no more than 9) for humans but few (no more than 2) for monkeys.

self-contained space) but with which we can nonetheless engage, and from which we can nonetheless learn. There is scant evidence that nonhuman primates can engage in such cognition, but that may be in part because controlled studies are so difficult to design—and moreover, there have been some hints that monkeys *can* learn from might-have-beens (Hayden, Pearson, and Platt 2009; Lee, McGreevy, and Barraclough 2005) and from the observation of others' experiences (Subiaul et al. 2004). How, then, do monkeys look toward the sequence of actions that comprise a film's narrative?

The Crucial Difference between Monkeys and Humans

To identify high-level visual features that attracted attention but that were omitted from the low-level saliency model (Figure 3), we selected movie frames on which a species' gaze was strongly clustered (Shepherd et al. 2010b). We then contrasted image content at three locations on each frame: at a location viewed by a monkey, a location viewed by a human, and the location selected by the artificial scanpath (Figure 6). 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 often than is predicted by low-level visual saliency
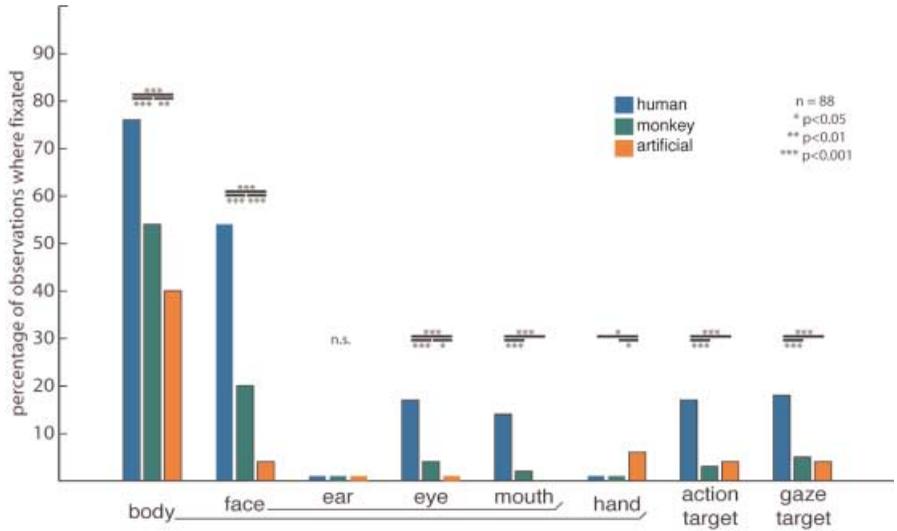
Figure 6. Higher-level social cues contribute to observed gaze correlations. We selected video frames in which observed gaze was highly clustered at a location far from that chosen by the low-level saliency model, examined scene content at the human, monkey, and simulated gaze coordinates, and tallied 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 than 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.

models. Both species looked particularly often at faces and eyes. These results were expected, as faces and bodies are important stimuli for many terrestrial animals, and especially for primates, which communicate through facial expressions (Ghazanfar and Santos 2004).

Remarkably, however, humans (and not monkeys) strongly attended objects being manipulated or examined by others (Shepherd et al. 2010b). To be honest, this difference surprised us. It is well known that monkeys follow gaze both in naturalistic and laboratory settings; in fact, many animal species have some sensitivity to gaze, suggesting that this may have arisen early in evolution (Shepherd 2010). Nonetheless, humans appear to have a special interest in the attention and intention of others. As a species whose success is founded in no small part on our ability to communicate and coordinate behavior, these joint attention abilities appear to be a crucial part of healthy human development. Gaze following by humans strongly predicts the acquisition of language (Brooks and Meltzoff 2008), and its deficit is associated with disorders such as autism (Nation and Penny 2008). Perhaps most significantly, humans actively share our attention and intentions with others through the use of

pointing gestures. For example, humans (perhaps uniquely among primates) point *declaratively,* which is to say, for no clear reason but to share an experience (Tomasello 2008).

This suggests that a major difference in how humans and monkeys see movies may be declarative in nature: we recognize the possibility that movies exist as a means of sharing experience, a skill-set in which our species has specialized and through which we have reaped unprecedented rewards. Movies are more than a continuous depiction of the scene in front of our head. It is more, we hypothesize, like the narrative we derive from the detailed observation and analysis of sequential events and actions, and which we share with one another through gesture, through language—and indeed, through art. Humans and monkeys share similar visual processing systems and similar mechanisms for integrating light and sound. We probably share some rudimentary ability to remember our narrative past and to learn from could-have-beens and from the experiences of others. There is a sense, after all, in which both humans and monkeys have tales. But only humans trade in the experience of telling them and hearing them told, and only humans have evolved the skill and developed the technology to bring these stories to life as sentences, as images, as sound, and, through the glare and glamour of their unification, as film.

Acknowledgments

Asif A. Ghazanfar is an associate professor of Neuroscience and Psychology, and an associated faculty member in Ecology & Evolutionary Biology at Princeton University. He is a recipient of the Human Frontiers Science Program Fellowship (2000); National Science Foundation CAREER Award (2006); and James S. McDonnell Scholar Award (2010). His research is at the interface of neuroscience, ethology, developmental biology, morphology, and evolution; his comparative approach includes studying a variety of primate species, including macaque monkeys, marmosets, and humans.

Stephen V. Shepherd is a postdoctoral fellow at the Neuroscience Institute and Department of Psychology at Princeton University. His research focuses on the neural mechanisms for communication and coordination between individuals.

References

Andrews, Michael W., Minakshi C. Bhat, and Leonard A. Rosenblum. 1995. "Acquisition and Long-Term Patterning of Joystick Selection of Food-Pellet vs Social-Video Reward by Bonnet Macaques." *Learning and Motivation* 26 (4): 370–379.

Andrews, Michael W., and Leonard A. Rosenblum. 2001. "Effects of Change in Social Content of Video Rewards on Response Patterns of Bonnet Macaques." *Learning and Motivation* 32 (4): 401–408.

Ball, William, and Edward Tronick. 1971. "Infant Responses to Impending Collision: Optical and Real." *Science* 171 (3973): 818–820.

Berg, David J., Susan E. Boehnke, Robert A. Marino, Douglas P. Munoz, and Laurent Itti. 2009. "Free Viewing of Dynamic Stimuli by Humans and Monkeys." *Journal of Vision* 9 (5): 1–15.

Bovet, Dalilia, and David A. Washburn. 2003. "Rhesus Macaques (Macaca Mulatta) Categorize Unknown Conspecifics According to Their Dominance Relations." *Journal of Comparative Psychology* 117 (4): 400–405.

Brooks, Rechele, and Andrew N. Meltzoff. 2008. "Infant Gaze Following and Pointing Predict Accelerated Vocabulary Growth through Two Years of Age: A Longitudinal, Growth Curve Modeling Study." *Journal of Child Language* 35 (1): 207–220.

Butler, Robert A. 1961. "The Responsiveness of Rhesus Monkeys to Motion Pictures." *Journal of Genetic Psychology* 98 (2): 239–245.

Capitanio, John P. 2002. "Sociability and Responses to Video Playbacks in Adult Male Rhesus Monkeys (Macaca Mulatta)." *Primates* 43 (3): 169–177.

Cappe, Celine, Gregor Thut, Vincenzo Romei, and Micah M. Murray. 2009. "Selective Integration of Auditory-Visual Looming Cues by Humans." *Neuropsychologia* 47 (4): 1045–1052.

Coile, D.Carolin, Celia H. Pollitz, and James C. Smith. 1989. "Behavioral Determination of Critical Flicker Fusion in Dogs." *Physiology & Behavior* 45 (6): 1087–1092.

Cook, David A. 2004. *A History of Narrative Film*. New York: W. W. Norton.

Cutting, James E. 2005. "Perceiving Scenes in Film and in the World." Pp. 9–27 in *Moving Image Theory: Ecological Considerations,* eds. J. D. Anderson and B. F. Anderson. Carbondale: University of Southern Illinois Press.

De Valois, Russell L., Herman C. Morgan, Martha C. Polson, William R. Mead, and Elaine M. Hull. 1974. "Psychophysical Studies of Monkey Vision—I. Macaque Luminosity and Color Vision Tests." *Vision Research* 14 (1): 53–67.

Dodge, Raymond, and Thomas Sparks Cline. 1901. "The Angle Velocity of Eye Movements." *Psychological Review* 8 (2): 145–157.

Franconeri, Steven L., and Daniel J. Simons. 2003. "Moving and Looming Stimuli Capture Attention." *Perception & Psychophysics* 65 (7): 999–1010.

Geller, T. 2008. "Overcoming the Uncanny Valley." *IEEE Computer Graphics and Applications* 28 (4): 11–17.

Ghazanfar, Asif A., John G. Neuhoff, and Nikos K. Logothetis. 2002. "Auditory Looming Perception in Rhesus Monkeys." *Proceedings of the National Academy of Sciences of the United States of America* 99 (24): 15755–15757.

Ghazanfar, Asif A., and Laurie R. Santos. 2004. "Primate Brains in the Wild: The Sensory Bases for Social Interactions." *Nature Reviews Neuroscience* 5 (8): 603–616.

Goldstein, Robert B., Robert L. Woods, and Eli Peli. 2007. "Where People Look When Watching Movies: Do All Viewers Look at the Same Place?" *Computers in Biology and Medicine* 37 (7): 957–964.

Green, Robert D., Karl F. MacDorman, Chin-Chang Ho, and Sandosh Vasudevana. 2008. "Sensitivity to the Proportions of Faces That Vary in Human Likeness." *Computers in Human Behavior* 24 (5): 2456–2474.

Hasson, Uri, Ohad Landesman, Barbara Knappmeyer, Ignacio Vallines, Nava Rubin, and David J. Heeger. 2008. "Neurocinematics: The Neuroscience of Film." *Projections: The Journal for Movies and Mind* 2 (1): 1–26.

Hasson, Uri, U., Eunice Yang, Ignacio Vallines, David J. Heeger, and Nava Rubin. 2008. "A Hierarchy of Temporal Receptive Windows in Human Cortex." *Journal of Neuroscience* 28 (10): 2539–2550.

Hayden, Benjamin Y., John M. Pearson, and Michael L. Platt. 2009. "Fictive Reward Signals in the Anterior Cingulate Cortex." *Science* 324 (5929): 948–950.

Humphrey, N. K. 1974. "Species and Individuals in Perceptual World of Monkeys." *Perception* 3 (1): 105–114.

Itti, Laurent. 2004. "The iLab Neuromorphic Vision C++ Toolkit: Free Tools for the Next Generation of Vision Algorithms." *Neuromorphic Engineer* 1 (1): 10.

Itti, Laurent, Nitin Dhavale, and Frederic Pighin. 2003. "Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention." Pp. 64–78 in *Proceedings of the SPIE 48th Annual International Symposium on Optical Science and Technology,* eds. B. Bosacchi, D. B. Fogel, and J. C. Bezdek. Bellingham, WA: SPIE Press.

Itti, Laurent, and Christof Koch. 2000. "A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention." *Vision Research* 40 (10–12): 1489–1506.

Itti, Laurent, Christof Koch, and Ernst Niebur. 1998. "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11): 1254–1259.

Jones, Benedict C., Anthony C. Little, D. Michael Burt, and David I. Perrett. 2004. "When Facial Attractiveness Is Only Skin Deep." *Perception* 33 (5): 569–576.

Kaas, Jon H., and Tod M. Preuss. 1993. "Archontan Affinities as Reflected in the Visual System." Pp. 115–128 in *Mammalian Phylogeny: Placentals,* eds. Frederick S. Szalay, Michael J. Novacek, and Malcolm C. McKenna. New York: Springer-Verlag.

Kahn, Peter H, Hiroshi Ishiguro, Batya Friedman, Takayuki Kanda, Nathan G. Freier, Rachel L. Severson, and Jessica Miller. 2007. "What Is Human? Toward Psychological Benchmarks in the Field of Human-Robot Interaction. " *Interaction Studies* 8 (3): 363–390.

Kluver, Heinrich. 1933. *Behavior Mechanisms in Monkeys.* Chicago: University of Chicago Press.

Krubitzer, Leah A., and Jon H. Kaas. 1990. "Cortical Connections of Mt in 4 Species of Primates—Areal, Modular, and Retinotopic Patterns." *Visual Neuroscience* 5 (2):165–204.

Lee, Daeyeol, Benjamin P. McGreevy, and Dominic J. Barraclough. 2005. Learning and Decision Making in Monkeys During a Rock-Paper-Scissors Game." *Cognitive Brain Research* 25 (2): 416–430.

Lewkowicz, David J., and Asif A. Ghazanfar. 2009. "The Emergence of Multisensory Systems Through Perceptual Narrowing." *Trends in Cognitive Science.* 13(11): 470–478.

———. 2011. "The Development of the Uncanny Valley in Infants." *Developmental Psychobiology.* In press. doi: 10.1002/dev.20583.

Loiperdinger, Martin. 2004. "Lumiere's Arrival of The Train: Cinema's Founding Myth." *Moving Image* 4 (1): 89–118.

MacDorman, Karl F., Robert D. Green, Chin-Chang Ho, and Clinton T. Koch. 2009. "Too Real for Comfort? Uncanny Responses to Computer Generated Faces." *Computers in Human Behavior* 25 (3): 695–710.

MacDorman, Karl F., and Hiroshi Ishiguro. 2006. "The Uncanny Advantage of Using Androids in Cognitive and Social Science Research." *Interaction Studies* 7 (3): 297–337.

Maier, Joost X., John G. Neuhoff, Nikos K. Logothetis, and Asif A. Ghazanfar. 2004. "Multisensory Integration of Looming Signals by Rhesus Monkeys." *Neuron* 43 (2): 177–181.

Maner, Jon K., Matthew T. Galliot, D. Aaron Rouby, and Saul L. Miller. 2007. "Can't Take My Eyes Off You: Attentional Adhesion to Mates and Rivals." *Journal of Personality & Social Psychology* 93 (3): 389–401.

Martin-Ordas, Gema, Daniel Haun, Fernando Colmenares, and Josep Call. 2010. "Keeping Track of Time: Evidence for Episodic-Like Memory in Great Apes." *Animal Cognition* 13 (2): 331–340.

Mirsky, I. Arthur, R. E. Miller, and John V. Murphy. 1958. "The Communication of Affect in Rhesus Monkeys: I. An Experimental Method." *Journal of the American Psychoanalytic Association* 6 (3): 433–441.

Mori, M. 1970. "Bukimi No Tani [The Uncanny Valley]." *Energy* 7 (4): 33–35.

Nakano, Tanami, Yoshiharu Yamamoto, Keiichi Kitajo, Toshimitsu Takahashi, and Shigeru Kitazawa. 2009. "Synchronization of Spontaneous Eyeblinks while Viewing Video Stories." *Proceedings of the Royal Society: Biological Sciences* 276 (1673): 3635–3644.

Nakayama, Katsura. 2004. "Observing Conspecifics Scratching Induces a Contagion of Scratching in Japanese Monkeys (Macaca Fuscata)." *Journal of Comparative Psychology* 118 (1): 20–24.

Nation, Kate, and Sophia Penny. 2008. "Sensitivity to Eye Gaze in Autism: Is It Normal? Is It Automatic? Is It Social?" *Development and Psychopathology* 20 (1): 79–97.

Neuhoff, John G. 1998. "A Perceptual Bias for Rising Tones." *Nature* 395 (6698): 123–124.

Paukner, Annika, and James R. Anderson. 2006. "Video-Induced Yawning in Stumptail Macaques (Macaca Arctoides)." *Biology Letters* 2 (1): 36–38.

Paxton, Regina, Benjamin M. Basile, Ikuma Adachi, Wendy A. Suzuki, Mark E. Wilson, and Robert R. Hampton. 2010. "Rhesus Monkeys (Macaca Mulatta) Rapidly Learn to Select Dominant Individuals in Videos of Artificial Social Interactions between Unfamiliar Conspecifics." *Journal of Comparative Psychology* 124 (4): 395–401.

Plimpton, Edward H., Karyl B. Swartz, and Leonard A. Rosenblum. 1981. "Responses of Juvenile Bonnet Macaques to Social Stimuli Presented Through Color Videotapes." *Developmental Psychobiology* 14 (2): 109–115.

Raaum, Ryan L., Kirstin N. Sterner, Colleen M. Noviello, Caro-Beth Stewart, and Todd R. Disotell. 2005. "Catarrhine Primate Divergence Dates Estimated from Complete Mitochondrial Genomes: Concordance with Fossil and Nuclear DNA Evidence." *Journal of Human Evolution* 48 (3): 237–257.

Rhodes, Gillian, and Tanya Tremewan. 1996. "Averageness, Exaggeration, and Facial Attractiveness." *Psychological Science* 7 (2): 105–110.

Rosenblum, Lawrence D., Claudia Carello, and Richard E. Pastore. 1987. "Relative Effectiveness of 3 Stimulus Variables for Locating a Moving Sound Source." *Perception* 16 (2): 175–186.

Rozin, Paul, and April E. Fallon. 1987. "A Perspective on Disgust." *Psychological Review* 94 (1): 23–41.

Schiff, Willim, James A. Caviness, and James J. Gibson. 1962. "Persistent Fear Responses in Rhesus Monkeys to the Optical Stimulus of 'Looming.'" *Science* 136 (3520): 982–983.

Seyama, Jun'ichiro, and Ruth S. Nagayama. 2007. "The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces." *Presence* 16 (4): 337–351.

Shepherd, Stephen V. 2010. "Following Gaze: Gaze-Following Behavior as a Window into Social Cognition." *Frontiers in Integrative Neuroscience* 4 (5): 1–13.

Shepherd, Stephen V., Shawn A. Steckenfinger, Uri Hasson, and Asif A. Ghazanfar. 2010a. "Gaze of Man, Monkey, & Machine." http://www.vimeo.com/7736755 (accessed 23 June 2011).

———. 2010b. "Human-Monkey Gaze Correlations Reveal Convergent and Divergent Patterns of Movie Viewing." *Current Biology* 20 (7): 649–656.

Steckenfinger, Shawn A., and Asif A. Ghazanfar. 2009. "Monkey Visual Behavior Falls into the Uncanny Valley." *Proceedings of the National Academy of Sciences* 106 (43): 18362–18366.

Subiaul, Francys, Jessica F. Cantlon, Ralph L. Holloway, and Herbert S. Terrace. 2004. "Cognitive Imitation in Rhesus Macaques." *Science* 305 (5682): 407–410.

Suddendorf, Thomas, and Michael C. Corballis. 2010. "Behavioural Evidence for Mental Time Travel in Nonhuman Animals." *Behavioural Brain Research* 215 (2): 292–298.

Tinbergen, Nicolass. 1951. *A Study of Instinct*. London: Oxford University Press.

Tomasello, Michael. 2008. *Origins of Human Communication*. Cambridge: MIT Press.

Tosi, Virgilio, Luciano Mecacci, and Elio Pasquali. 1997. "Scanning Eye Movements Made When Viewing Film: Preliminary Observations." *International Journal of Neuroscience* 92 (1–2): 47–52.

Walters, Michael L., Dag S. Syrdal, Kerstin Dautenhaun, René te Boekhorst, and K.L. Koay. 2008. "Avoiding the Uncanny Valley: Robot Appearance, Personality and Consistency of Behavior in an Attention-Seeking Home Scenario for a Robot Companion." *Autonomous Robots* 24 (2): 159–178.

Wheatstone, Charles. 1852. "On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision (Continued)." The Bakerian Lecture—Contributions to the Physiology of Vision: Part II. *Philosophical Transactions of the Royal Society of London* 142: 1–18.

Filmography

Attenborough, David. 2002. *The Life of Mammals* (Documentary). UK.

BBC Two. 2010. *Natural World: The Chimpcam Project* (Documentary). UK.

Chaplin, Charles. 1931. *City Lights*. USA.

Cooper, Merian C., and Ernest B. Schoedsack. 1933. *King Kong*. USA.

Honda, Ishiro. 1954. *Godzilla* (Gojira). Japan.

Levinson, Barry. 1985. *Young Sherlock Holmes*. USA.

Lumière, Louis. 1895. *L'Arrivee d'un train a La Ciotat*. France.

Meyer, Nicholas. 1982. *Star Trek II: The Wrath of Khan*. USA.

Myrick, Daniel, and Eduardo Sánchez. 1999. *The Blair Witch Project*. USA.

Reitherman, Wolfgang. 1967. *The Jungle Book*. Argentina.

Sakaguchi, Hironobo, and Motonori Sakakibara. 2001. *Final Fantasy: The Spirits Within*. USA.

Zemeckis, Robert. 2004. *The Polar Express*. USA.