Research Article

Perceptual mechanism underlying gaze guidance in chimpanzees and humans

Fumihiro Kano (1, 2)

Masaki Tomonaga (1)

1. Primate Research Institute, Kyoto University
2. Japan Society for Promotion of Science

fkanou@pri.kyoto-u.ac.jp

+81-80-6902-5013

Abstract

Previous studies comparing eye movements between humans and their closest relatives, chimpanzees, have revealed similarities and differences between the species in terms of where individuals fixate their gaze during free viewing of a naturalistic scene, including social stimuli (e.g. body and face). However, those results were somewhat confounded by the fact that gaze behavior is influenced by low-level stimulus properties (e.g., color and form), and by high-level processes such as social sensitivity and knowledge about the scene. Given the known perceptual and cognitive similarities between chimpanzees and humans, it is expected that such low-level effects do not play a critical role in explaining the high-level similarities and differences between the species. However, there is no quantitative evidence to support this assumption. To estimate the effect of local stimulus saliency on such eye-movement patterns, this study used a well-established bottom-up saliency model. In addition, to elucidate the cues that the viewers use to guide their gaze, we presented scenes in which we had manipulated various stimulus properties. As expected, the saliency model did not fully predict the fixation patterns actually observed in chimpanzees and humans. In addition, both species used multiple cues to fixate socially significant areas such as the face. There was no evidence suggesting any differences between chimpanzees and humans in their responses to low-level saliency. Therefore, this study found a substantial amount of similarity in the perceptual mechanisms underlying gaze guidance in chimpanzees and humans, and thereby offers a foundation for direct comparisons between them.

Key words: chimpanzees, eye-tracking, face, picture perception, saliency

Introduction

Eye-tracking methodology in human and nonhuman primates has been used for over 50 years (Fuchs, 1967; Yarbus, 1967). Eye-movement patterns of nonhuman primates show a significant degree of similarity with those of humans under similar experimental conditions (Kano and Tomonaga, 2009; Shepherd, Steckenfinger, Hasson, and Ghazanfar, 2010). Comparative studies of human and nonhuman primates have directly compared the species in order to clarify both similarities and differences in their eye-movement characteristics (Dahl, Wallraven, Bulthoff, and Logothetis, 2009; Gothard, Brooks, and Peterson, 2009; Guo, Robertson, Mahmoodi, Tadmor, and Young, 2003; Kano and Tomonaga, 2009; Keating and Keating, 1982; Nahm, Perret, Amaral, and Albright, 1997; Shepherd et al., 2010). Those similarities and differences have been an important source of information for the study of the evolution of visual behavior, social perception, and high-level cognition (Kano and Tomonaga, 2009; Shepherd et al., 2010). Although apes have been essential for this comparative approach, their eye-movement characteristics are largely unknown compared to those of the well-studied macaque species.

Recently, eye-tracking studies in chimpanzees (*Pan troglodytes*), the species most closely related to humans, have been reported (Hattori, Kano, and Tomonaga, 2010; Hirata, Fuwa, Sugama, Kusunoki, and Fujita, 2010; Kano and Tomonaga, 2009, 2010). Those studies have presented naturalistic images of scenes (including faces, bodies, etc.) to chimpanzees and humans under free-viewing conditions and compared their fixation patterns under similar experimental conditions. There are several advantages of comparing chimpanzees and humans for a free-viewing task. First, chimpanzees are the species most closely related to humans and are known to have similar perceptual mechanisms (Matsuno, Kawai, and Matsuzawa, 2004; Matsuzawa, 1985, 1990; Tomonaga and Matsuzawa, 1992). Second, the demands of a free-viewing task are small; the participants of both species need only look at the stimuli spontaneously and are not trained to solve particular problems using their eye movements. Third, for the same reason, a free-viewing task is relatively independent of the effect of reward or training. Therefore, we are able to efficiently and directly compare the species, find both similarities and differences between them, and discuss the extent to which chimpanzees and humans are similar and different in their perception and cognition.

In the previous study comparing chimpanzees and humans in a free-viewing task, it was found that the species were very similar in terms of where to fixate (i.e. scanpath similarity). For example, when presented with a scene including an entire body of a chimpanzee, a human, or another animal, both chimpanzees and humans concentrated fixations on the body, especially the face, rather than on the background. In addition, both species fixated on the face immediately after the image presentation (within the first few fixations). However, those responses differed quantitatively between the species; humans showed a higher proportion of face fixations than did chimpanzees. There seem to be several functional reasons for those similarities and differences between the species. First, faces are the most important source of social information (such as individuality and emotions) for both chimpanzees and humans (Chevalier-Skolnikoff, 1973; Parr, Dove, and Hopkins, 1998), and thus frequent inspection and immediate detection of facial characteristics may benefit them by enabling them to obtain such information efficiently. Second, humans have a specific form of facial communication; humans often engage in lengthy face-to-face communication, accompanied by intense eye contact (Argyle and Cook, 1976). Therefore, more frequent inspection of faces may benefit humans more specifically than chimpanzees in the context of their own form of facial communication.

There seem to be several factors that determine such similarities and differences. These include, for example, the perception of low-level visual properties (e.g. color, form), the perception of bodies and faces, and knowledge about the scenes (which the viewers had obtained through daily lives or experimental instructions). Previous studies using forced-choice discrimination paradigms have found that the perceptions of low-level visual properties involving color (Matsuno et al., 2004; Matsuzawa, 1985), form (Matsuzawa, 1990; Tomonaga and Matsuzawa, 1992) are largely similar between chimpanzees and humans. In addition, the mechanisms involving advanced social perceptions involving faces (Parr et al., 1998; Parr, Hecht, Barks, Preuss, and Votaw, 2009; Tomonaga, 2007, 2010; Tomonaga and Imura, 2009) and bodies (Tomonaga and Imura, 2008) are also similar between the species.

Because of these similarities between chimpanzees and humans, it is expected that the influence of low-level stimulus properties on their eye-movement patterns appears similarly in the two species and does not play a critical role in explaining for the overall similarities and differences between the species. However, there are no quantitative data to support this assumption. It is important to separate low-level from higher-level influences on eye-movement patterns in order to provide a foundation for direct comparison between the species. Therefore, this study aimed to elucidate the influence of low-level stimulus properties on the eye-movements of chimpanzees and humans.

We used two approaches in order to separate low-level from high-level influences on eye movements. First, to simulate responses to local stimulus properties, we used the well-established bottom-up saliency model (Itti and Koch, 2000; Walther and Koch, 2006). This model estimates the local saliency of an image based on its low-level components -- such as color, intensity, and component orientations -- and predicts the locations of attention based on these local saliency values. The second approach used global manipulation of stimulus properties (e.g., stimulus properties such as color, configuration, frequency components, orientation, complexity, and location of scene features) and observed how participants changed their patterns of scanning in response to the manipulations.

In this study, we used similar stimulus sets to those used by Kano and Tomonaga (2009) and analyzed the participants' responses to social stimuli, especially to the face, as a main measure. In the previous study, chimpanzees and humans fixated the face more frequently than any other part of the scene. The frequent fixation to the face is most likely caused not only by the low-level saliency of the faces, but also by the participants' sensitivity to the social stimuli. This study aimed to investigate the extent to which such facial fixation patterns could be explained by the bottom-up saliency model and could be influenced by the global manipulation of stimulus properties in the scene.

Therefore, the topics we addressed in this study were as follows. (1) The degree of similarity in fixation distribution patterns between chimpanzees, humans, and those predicted by the bottom-up saliency model; we expected similar patterns of fixation distribution between chimpanzees and humans even when we controlled for the low-level saliency. (2) The extent to which the gaze of chimpanzees and humans is attracted by low-level saliency. (3) The extent to which the two species' facial fixation is influenced by the global manipulation of stimulus properties in the scene. For (2) and (3), again we expected a significant degree of similarity between the species given their perceptual similarities.

Method

Participants

Six chimpanzees (five females, one male; aged 9–31 years) and 16 humans (11 females, five males; aged 18–31 years; all Japanese) participated in this experiment. The chimpanzees were members of a social group of 14 individuals living in enriched outdoor compounds and attached indoor residences (Matsuzawa, Tomonaga, and Tanaka, 2006). They were highly experienced in observing pictorial representations appearing on a computer screen (Matsuzawa et al., 2006). No food or water deprivation occurred during the study period. Care and use of the chimpanzees adhered to the 2002 version of the Guidelines for the Care and Use of Laboratory Primates published by the Primate Research Institute, Kyoto University. The experimental protocol was approved by the Animal Welfare and Care Committee of the Institute and by the Animal Research Committee of Kyoto University. The human participants were graduate and undergraduate students, who participated in the experiment voluntarily. Informed consent was obtained from all human participants.

Apparatus

Both species used the same apparatus, in order to ensure the possibility of direct comparison between the species. Participants sat still and unrestrained in an experimental booth, with the eye-tracking apparatus and the experimenter separated by transparent acrylic panels (see S1). A table-mounted eye tracker measured their eye movements using infrared corneal reflection techniques (60 Hz; Tobii X120, Tobii Technology AB). This eye-tracker has wide-angle lenses (±40 degrees in the semicircle above the camera) and thus obviated the necessity to restrain the subjects. The eye-tracker and the 17-inch LCD monitor (1280×1024) were mounted on a movable platform, and the distance between the platform and the participants was adjusted to the point at which the gaze was most accurately recorded (60 ±10 cm). This flexible adjustment of the distance between the platform and the participants enabled us to record the gaze movements of chimpanzees without any head restraint. The participant's gaze was recorded as a relative coordinate with respect to the monitor size (i.e. not as the gaze angle). One degree of gaze angle corresponded to approximately 1 cm on the screen at a typical 60-cm viewing distance.

As a result of the training conducted during the study performed by Kano and Tomonaga (2009), the chimpanzees were already skilled at sitting still in front of an eye-tracker and looking upon request at a fixation point that appeared on the screen. Five-point calibration was conducted for humans; for chimpanzees, the calibration points were reduced to two in order to decrease the time required for each calibration process. However, for chimpanzees, the calibration was repeated until the maximum accuracy was obtained. The accuracy was checked by presenting to both species five fixation points on the screen. Using these calibration procedures, six participants of both species were tested for accuracy, and the errors were found to be small and comparable between the species (mean errors of 0.62 ±0.06 and 0.52 ±0.05 cm ± s.e.m. on the monitor for chimpanzees and humans, respectively). The drift (the calibration error due to changes occurring in the eye surface) was checked occasionally by presenting the fixation points to the participants again.

Stimuli

We prepared 20 color photographs of naturalistic scenes containing a human figure (Figure 2). We used only human figures (all Japanese; no chimpanzees or other animals) in this study because a previous study using an identical experimental procedure (Kano and Tomonaga 2009) found similar fixation patterns in both species for all animal figures. These 20 images served as the control condition. Eight experimental conditions were additionally prepared (for the details of manipulation procedure, see Table 1). In the monochrome, line drawing, and schematic drawing conditions, we eliminated color, low-spatial frequency component, and complex lines, respectively, from the original color scene and aimed to examine the influence of realistic appearance of a scene on the participants' response to the faces. In the blurred and silhouette conditions, we blurred and eliminated local features of face and body from the scene and aimed to examine the influence of those features on the responses. In the upside down and scramble conditions, we inverted and scrambled the scene, respectively, and aimed to examine the influence of orientation and arrangements of bodily parts on the response (i.e. we checked whether participants used only information that the head is above the body). In the headless condition, we eliminated the head from the body and aimed to examine whether participants used only bodily information to fixate the location where the head ought to be. Overall, these conditions aimed to observe whether participants used multiple cues to detect the location of faces in the scene. Each experimental condition was represented by five examples created by manipulating the control images. These five examples were pseudo-randomly selected from the entire set of 20 control images so that each control image was used at least once in the experimental conditions. In total, 60 stimuli were used (40 experimental and

20 control images). The images were converted into 1000 × 800 pixel images with surrounding gray frames (1280 × 1024 pixels in total; 37 × 30 degrees at a typical 60-cm viewing distance). We used Adobe Photoshop CS3 to process the images.

Procedure

Procedural differences for testing chimpanzees and humans were minimized to allow for direct comparisons between species. In each trial, an image was presented after participants focused on a fixation point that appeared at a random position on the screen. Participants were then allowed to view images freely. The participants of both species rarely kept gazing at the fixation point after the image presentation (i.e. spontaneous scanning was almost always observed). Stimuli were presented for 3 sec each. The presentation order of conditions and trials were randomized for each participant so that the same conditions were not presented more than three times in succession. 20 other stimuli depicting various interesting scenes (e.g. pictures of funny faces) were presented occasionally during the sessions in order to keep the participants interested. The entire session therefore consisted of 80 trials: 60 experimental stimuli and 20 others. The entire session was conducted on a single day for humans, whereas the session was divided among 15 days for the chimpanzees in order to decrease the time required for each daily experiment (each day used six examples for the chimpanzees). In a preliminary session, we confirmed that our human participants showed similar scanning patterns of bodies/faces when tested on separate days (comparing the results from this study with those from Kano and Tomonaga (2009)). Daily experiments lasted 10–15 min for the chimpanzees and 20 min for the humans. Human participants received book coupons as rewards after the session, and chimpanzees received a small piece of apple after each trial. The reward was given for chimpanzees in return for the initial fixation at the beginning of the trial, and thus was given independently of their viewing behavior during the image presentation. Overall, those procedural differences between the species were made in an effort to increase the motivation of both species to participate in the daily experiment, and to keep their interest during the presentation of each image (3 sec). Trials in which participants only glanced at the monitor (one or two fixations) were repeated after the whole session and were replaced by the new trials. As a result, we had no loss of trials for both species.

Data analysis

Fixation definition

A fixation was scored if the gaze remained stationary within a radius of 50 pixels for at least 75 ms (more than five measurement samples). Otherwise, the recorded sample was defined as part of a saccade. The records during the first 200 ms were eliminated from the analysis, thereby eliminating fixations that followed the offset of the initial fixation point.

Area of Interest (AOI)

Each stimulus was divided into areas of interest (AOI) for the purpose of quantitative comparison. Each scene was divided into background, face, torso, arms, and legs (Figure 3, bottom). Each AOI was drawn 20 pixels larger than the precise outline of the features to avoid errors in gaze estimations. The AOI of background, torso, legs, arms, and face were laid above in this order (i.e. face is the topmost). If two or more AOIs were duplicated, the samples were added to the upper AOI.

Chance level

The chance level was set on the assumption that participants would view images randomly. However, the participants generally showed a central bias in fixation distribution, while the model did not (evident by inspection of Figure 2). This needs to be controlled to compare participants with the model, because such central bias is known to be caused either by the participants' bias in scanning images or by the experimenter's bias in the arrangement of main objects in the scene (Henderson, Brockmole, Castelhano, and Mack, 2007; Tatler, 2007) (i.e., caused independently of the low-level stimulus properties). Therefore, in this study, we modified the definition of chance level by controlling for such particular bias shown by each participant. Specifically, we compared the particular scanpath, which was obtained from a participant (or the model) in a trial, with all the other scanpaths, which were obtained from the same participants (or the model) in all the other trials of the experiment. All data shown in this study were calculated as differences between the value obtained from the particular scanpath and the mean value obtained from the other control scanpaths (i.e., the chance level).

Saliency model

We used the well-established bottom-up model to estimate the low-level saliency of the images (Itti and Koch, 2000; Walther and Koch, 2006). This model processes the image with respect to several features -- such as color (red-green, blue-yellow), intensity, and orientation (0, 45, 90, 135 degrees) -- then extracts the local discontinuities in each feature, and finally combines them into a single 'saliency map' (Figure 1). The model then predicts a scanpath based on the saliency map, selecting salient locations in order of decreasing saliency. In this experiment, the saliency maps and the model scanpaths were generated by Saliency Toolbox 2.2 (http://www.saliencytoolbox.net) in Matlab with all-default parameters. We used the original resolution of images (1280×1024; including the surrounding gray frame) for the simulation in the model. Because this model does not predict the duration of each fixation, we arbitrarily set the scanpath length of the model as 9 fixations (about as long as chimpanzee scanpaths in 3-s viewing) to compare the model with the participants. There is no variance in the output when repeating the simulation.

To determine the saliency value at each fixated area, we employed the following procedures. First, saliency value was normalized within each map to a range of 0 (not salient) to 1 (highly salient). Second, to avoid errors in gaze estimation, the saliency map was divided into a 12×9 grid, and all saliency values (i.e. 1280×1024 samples, in total) were summed within each grid (i.e. each grid had approx. 100×100 samples). The fixated area was defined as the grid where the fixation was observed, and the saliency value of each grid was used for the saliency value at each fixated area.

Results

Figure 3a shows the distribution patterns of fixation over the scene in each species/model (the data were sampled from 20 control images). Comparing between species and between AOIs, we found a significant interaction ($F_{2.3,\,47} = 9.52$, $p < 0.001$, $\eta^2 = 0.32$)[1] because chimpanzees distributed their fixations over the scene more widely than did humans. Comparing between AOIs respectively for each species, we found significant main effects for both chimpanzees ($F_{1.8,\,9.3} = 23.80$, $p < 0.001$, $\eta^2 = 0.82$) and humans ($F_{2.4,\,37} = 358.86$, $p < 0.001$, $\eta^2 = 0.96$) because both species showed higher proportion of fixations on particular areas (the bodies, especially faces, rather than backgrounds) than would be expected by chance (represented as zero in the figures). This pattern of results emerged even when the model was subtracted from each species:

chimpanzees ($F_{1.8, 9.3}$ = 5.61, p = 0.003, η^2 = 0.52) and humans ($F_{2.4, 37}$ = 159.74, p < 0.001, η^2 = 0.91). This pattern emerged no later than the first two fixations, as shown in Figures 3b and 3c, and is consistent with the previous reports in humans (Crouzet, Kirchner, and Thorpe, 2010; Fletcher-Watson, Findley, Leekam, and Benson, 2008; Honey, Kirchner, VanRullen, 2008). The global similarities in distribution patterns of fixation among chimpanzees, humans, and the model suggest that the saliency model partially (but not fully) explained those patterns for the two species. Although chimpanzees were more similar to the model than were humans in that regard, it should be noted that this does not mean that the low-level visual saliency influenced chimpanzees more strongly than humans; this means that chimpanzees distributed their fixations over the scene more widely than did humans, but less widely than did the model.

Indeed, chimpanzees and humans did not significantly differ in their responses to low-level visual saliency. There was no significant effect of species in the saliency values at fixation (Figure 4); neither the main effect of species ($F_{1, 20}$ = 0.014, p = 0.90, η^2 = 0.001) nor the interaction between species and fixation order ($F_{5, 100}$ = 0.46, p = 0.80, η^2 = 0.023) was significant. Overall, however, both species fixated on salient regions in the scene more than would be expected by chance: the mean saliency values for the first 6 fixations were significantly higher than zero in both chimpanzees (t(5) = 9.83, p < 0.001) and humans (t(15) = 19.27, p < 0.001). This pattern emerged more strongly for the earlier than for the later fixations: saliency value decreased as a function of increasing fixation order ($F_{5, 100}$ = 3.20, p = 0.010, η^2 = 0.13). These results suggest that the saliency model predicted the distribution patterns of fixation in both chimpanzees and humans better than chance, especially for the early fixations. However, it should be noted that this result does not necessarily mean that the low-level saliency alone influenced the species' distribution patterns of fixation, because such frequently fixated areas (e.g., bodies and faces) were in general more visually salient (because of the complexity of lines, for example) as well as more informative than the other areas of the scene (Figure 3; refer to (Henderson et al., 2007) for a similar discussion).

We then examined the effect of image manipulations on the fixation patterns of chimpanzees and humans (Figure 5). Figure 5b shows the proportion of fixations on the faces as a function of image manipulations. There was no interaction of species with condition ($F_{3.5,\ 70}$ = 1.13, p = 0.34, η^2 = 0.05). The main effect of species was significant: humans showed a higher proportion of fixations on faces than did chimpanzees ($F_{1,\ 20}$ = 5.51, p = 0.029, η^2 = 0.21), which is consistent with the aforementioned result. The main effect of condition ($F_{3.5,\ 70}$ = 5.33, $p < 0.001$, η^2 = 0.21) was significant: participants showed a lower proportion of face fixations in the headless than the other conditions (as was revealed by the pair-wise comparisons with Bonferroni's correction).

However, even in the headless condition, both species showed a higher proportion of fixations on the face original locations than would be expected by the model (as was revealed by the post-hoc *t*-tests). This means that even when a head was actually absent from the scene, both species concentrated fixations on the area where the face would have been (i.e. above the body).

Figure 5c shows the mean saliency values at the first 6 fixations as a function of image manipulations. The main effect of condition was significant ($F_{8,\ 160}$ = 46.93, $p < 0.001$, η^2 = 0.70), probably modulated by saliency (or informativeness) in local features of the scene, which was an outcome of image manipulations. Importantly, there was no effect of species despite these image manipulations, either the main effect of species ($F_{1,\ 20}$ = 0.017, p = 0.89, η^2 = 0.001) or the interaction between species and condition ($F_{8,\ 160}$ = 1.18, p = 0.31, η^2 = 0.05).

Discussion

Chimpanzees and humans distributed fixations over the scene non-randomly, and showed higher fixation proportions on particular areas of the scene, especially faces, than would be expected by the saliency model. However, humans showed an even higher proportion of fixation on the bodies and faces than did chimpanzees. These results emerged even at the first two fixations, at the earliest moments of scene inspection, suggesting that those fixation patterns reflect automatic rather than voluntary control of gaze. Saliency values of chimpanzees and humans in the fixated region were higher than would be expected by chance, suggesting that low-level saliency partially (but not fully) predicted the species' distribution patterns of fixation. However, chimpanzees and humans did not significantly differ in their responses to low-level saliency. None of global manipulations of stimulus properties in the scene (color, configuration, frequency components, orientation, complexity, and local features) critically altered both species' strong tendency toward fixating faces, suggesting that both species used multiple cues to fixate faces. In addition, although those manipulations changed the extent to which low-level saliency influenced both species, chimpanzees and humans did not differ in the degree of change in the response.

Therefore, chimpanzees and humans seem to be qualitatively similar in the sense that both species have an enhanced perceptual mechanism to guide their fixation location, one which is more complex than would be presumed on the basis of the saliency model (i.e. color, intensity, and orientations), and have multiple strategies to perceive the location of faces. Quantitatively, these two species did not differ significantly in their responses to low-level saliency, suggesting that they have similar perceptual mechanisms to guide the fixation locations.

Einhäuser et al. (Einhuäuser, Kruse, Hoffmann, and König, 2006) used the standard saliency model to predict the fixation location of monkeys (rhesus macaques) and humans when presented with the still images of naturalistic scene (without social contents). They found that monkeys and humans did not differ significantly in their responses to low-level saliency when viewing those images, which is consistent with the present study comparing chimpanzees and humans. However, when the luminance-contrast (or the saliency) was manipulated locally in the image, the monkeys responded to those manipulated areas more strongly than did the humans. In the similar analysis to that of Einhäuser et al. (2006) and this study, Berg et al. (Berg, Boehnke, Marino, and Munoz, and Itti, 2009) found that, when presented with dynamic scenes including various social, non-social, and narrative contents, humans responded to the low-level visual saliency more strongly than did monkeys (perhaps because monkeys tended to move their eyes independently of the stimuli (e.g. inattentiveness to the stimuli) or show a large degree of individual differences in their fixation patterns), which is somewhat inconsistent with Einhäuser et al. (2006) and this study. Therefore, multiple factors seem to be involved in the species difference in the responses to the low-level saliency. To clarify those factors, it is necessary to directly compare between the three species for their fixation patterns when presented with various contents of still and dynamic scenes.

Cerf et al. (Cerf, Harel, Einhäuser, and Koch, 2008) have shown that the addition of a “face channel” into the standard saliency model better predicts the fixation patterns of human participants viewing a naturalistic scene that includes faces. They used an established face detector algorithm for that purpose, which predicts the location of faces based on local facial features (e.g. local discontinuities in intensity around eye and nose regions). The distribution patterns of fixation observed in this study suggest that chimpanzees and humans have such a face perception channel in addition to the low-level channels. However, the mechanism underlying such a face channel seems more complex in chimpanzees and humans than would be assumed by the face detector algorithm. This is because chimpanzees and humans concentrated fixations on the faces even when local features of faces were significantly reduced (schematic and blurred) or completely silhouetted out of the scene. They did so even when the faces were removed completely (headless), suggesting that chimpanzees and humans can use the bodily configuration alone to fixate where faces ought to be. On the other hand, chimpanzees and humans also seem to be able to use local cues to fixate faces, because they concentrated fixations on the face parts even when bodily configuration was disrupted (scrambled). Therefore, chimpanzees and humans seem to have an enhanced perceptual mechanism to guide their fixations to a face, a mechanism that is more complex than would be assumed by the standard saliency model or the saliency model combined with face detection.

Notwithstanding those similarities between the species, chimpanzees and humans differ quantitatively in the distribution patterns of fixations. Humans showed a higher proportion of fixations on bodies and faces than did chimpanzees. As clarified above, it is unlikely that this species difference resulted from their differential responses to the low-level visual properties (or in their differential tendencies for central bias). It is also unlikely that this species difference resulted from the use of human, as opposed to chimpanzee figures as stimuli, because a previous study (Kano and Tomonaga, 2009) obtained the same patterns of results when using chimpanzees and other mammals as the stimulus models. Therefore, we interpreted the results in the following two ways.

First, although the results suggested that both species have similar mechanism to guide their gaze to the social stimuli (body and face), those mechanisms may operate differently in each species. For example, humans may put more emphasis on the body/face channels to create the saliency map, and so humans may perceive bodies and faces as more salient than chimpanzees do. Second, humans, compared to chimpanzees, may have a stronger tendency to process scenes in a top-down rather than a bottom-up manner, and thus would be expected to show a higher proportion of fixations on the semantically informative areas such as bodies and faces. Further studies are necessary to test these two possibilities.

In summary, this study presented, to chimpanzees and humans, naturalistic (unmanipulated) scenes including body, face, and their manipulated representations. We then compared among the two species and the saliency model for the fixation patterns on the images. We found that the saliency model did not fully predict the fixation patterns actually observed in chimpanzees and humans. In addition, both species used multiple cues to fixate the face. There was no evidence suggesting any differences between chimpanzees and humans in the perception of low-level saliency (e.g. color, intensity, or orientations). Therefore, we showed a substantial amount of similarities in the perceptual mechanism underlying gaze guidance between chimpanzees and humans, and thereby offer a foundation for the direct comparison between the species. Further studies are necessary to elucidate the high-level similarities and differences between the species (e.g. social sensitivity, knowledge-based attention).

Acknowledgements

This research was financially supported by the Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology

(MEXT) of Japan Grants-in-Aid for Scientific Research (nos. 16002001, 19300091, 20002001, 212299), and the JSPS/MEXT global COE programs (D07 and A06). We thank Drs. T. Matsuzawa, I. Adachi, S. Hirata, and Y. Hattori for their help and invaluable comments. We also thank Dr. J. Call for the comments on an earlier draft of the paper. We also thank the Center for Human Evolution Modeling Research at the Primate Research Institute for the daily care of the chimpanzees.

The experiments comply with the current laws of the country in which they were performed.

The authors declare that they have no conflict of interest.

References

Argyle, M., Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *J Vis, 9*(5), 1-15.

Cerf, M., Harel, J., Einhuäuser, W., Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Adv Neural Inform Process Syst, 20*, 241-248.

Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 11-89). New York: Academic Press.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. Fast saccades toward faces: Face detection in just 100 ms. *J Vis, 10*(4), 1-17.

Einhäuser, W., Kruse, W., Hoffmann, K. P., & König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Res, 46*(8-9), 1194-1209.

Dahl, C. D., Wallraven, C., Bulthoff, H. H., Logothetis, N. K. (2009). Humans and macaques employ similar face-processing strategies. *Curr Biol, 19*(6), 509-513.

Fuchs, A. F. (1967). Saccadic and smooth pursuit eye movements in the monkey. *J Physiol, 191*(3), 609-631.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*, 571-583.

Gothard, K. M., Brooks, K. N., Peterson, M. A. (2009). Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim Cogn, 12*(1), 155-167.

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., Young, M. P. (2003). How do monkeys view faces?—a study of eye movements. *Exp Brain Res, 150*(3), 363-374.

Hattori, Y., Kano, F., Tomonaga, M. (2010). Differential sensitivity to conspecific and allospecific cues in chimpanzees and humans: A comparative eye-tracking study. *Biology Lett*. *6*(5), 610-613

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Neitherlands: Elsevier.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *J Vis, 8*(12), 1-13.

Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., Fujita, S. (2010). Facial perception of conspecifics: chimpanzees (*Pan troglodytes*) preferentially attend to proper orientation and open eyes. *Anim Cogn, 13*(5), 679-688.

Itti, L., Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res, 40*(10-12), 1489-1506.

Kano, F., Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative eye-tracking study. *Proc Roy Soc B, 276*(1664), 1949-1955.

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

Keating, C. F., Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception, 11*(2), 211-219.

Matsuno, T., Kawai, N., Matsuzawa, T. (2004). Color classification by chimpanzees (*Pan troglodytes*) in a matching-to-sample task. *Behav Brain Res, 148*(1-2), 157-165.

Matsuzawa, T. (1985). Colour naming and classification in a chimpanzee (*Pan troglodytes*). *J Hum Evol, 14*(3), 283-291.

Matsuzawa, T. (1990). Form perception and visual acuity in a chimpanzee. *Folia Primatol, 55*(1), 24-32.

Matsuzawa, T., Tomonaga, M., Tanaka, M. (2006). *Cognitive development in chimpanzees*. Tokyo: Springer.

Nahm, F. K. D., Perret, A., Amaral, D. G., Albright, T. D. (1997). How do monkeys look at faces? *J Cogn Neurosci, 9*(5), 611-623.

Parr, L. A., Dove, T., Hopkins, W. D. (1998). Why faces may be special: evidence of the inversion effect in chimpanzees. *J Cogn Neurosci, 10*(5), 615-622.

Parr, L. A., Hecht, E., Barks, S. K., Preuss, T. M., Votaw, J. R. (2009). Face processing in the chimpanzee brain. *Curr Biol, 19*(1), 50-53.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol, 20*(7), 649-656.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *J Vis, 7*(14).

Tomonaga, M. (2007). Visual search for orientation of faces by a chimpanzee (*Pan troglodytes*): face-specific upright superiority and the role of facial configural properties. *Primates, 48*(1), 1-12.

Tomonaga, M. (2010). Chimpanzee eyes have it? Social cognition on the basis of gaze and attention from the comparative-cognitive-developmental perspective. In E. Lornsdorf, S. Ross & T. Matsuzawa (Eds.), *The mind of the chimpanzee:*

Ecological and empirical perspectives. Chicago: University of Chicago Press.

Tomonaga, M., Imura, T. (2008). Chimp in the shadow: Efficient detection of chimpanzee body by chimpanzees? *Primate Res, 24*(S), 14-15 (Japanese abstract only).

Tomonaga, M., Imura, T. (2009). Faces capture the visuospatial attention of chimpanzees(*Pan troglodytes*): evidence from a cueing experiment. *Front Zool, 6*(1), 14.

Tomonaga, M., Matsuzawa, T. (1992). Perception of complex geometric-figures in chimpanzees (Pan troglodytes) and humans (Homo sapiens): Analysis of visual similarity on the basis of choice reaction time. *J Comp Psychol, 106*(1), 43-52.

Walther, D., Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395-1407.

Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum Press.

Content Note

1. In the ANOVAs, in cases in which the assumption of homogeneity of variance was violated, the Greenhouse-Geisser correction was applied, and corrected *P* values were calculated.

Tables

Table 1. *Procedures Used for Image Manipulation*

condition	n	procedure
control	20	
monochrome	5	The color was removed from the original photographs.
line drawing	5	Only edges were extracted from the monochrome photograph (with a Photoshop function), and binary image processing techniques simplified the image (emphasizing the fat lines and eliminating the thin lines and small dots).
schematic drawing	5	The edges were roughly traced with simple black circles and lines.
blurred	5	The edges were blurred to the extent that the facial features were not recognizable (a Gaussian blur 20 pixels in diameter).
silhouette	5	The figure was colored in black, and binary image processing techniques transformed the background into black and white patches.
upside down	5	The original photographs were turned upside down.
scrambled	5	The original scenes were superimposed into a 6 × 5 matrix, and each block of the matrix was randomly rearranged. A matrix was defined so that a block includes the whole face (i.e. the face was intact).
headless	5	The head was eliminated so that the background was visible through the regions in which the head was previously located. To this end, the headless figure was cropped in the first image and superimposed on the second image that contains only background.

Figure legends

Figure 1

Scanpaths of a chimpanzee and a human, each superimposed on the naturalistic scene (a) and fine art painting (b; Paul Klee, 1923, "*Puppet Theater*"; see Supporting material for the quantitative data). Fixations and saccades are indicated by dots and lines, respectively. The stimuli were presented for 3 sec. each. Also shown are a raw saliency map and the scanpath predicted by the model. Bright areas indicate areas of high saliency. By design, the model made 9 fixations on the images in the order of decreasing saliency.

Figure 2

The locations of all fixations made by a chimpanzee, a human, and the model. While the model showed a relatively even distribution of fixations over the scene, the chimpanzee and the human showed a central bias in the distribution. Therefore, the chance level (random gaze pattern) was adjusted to control for this observed bias (see text).

Figure 3

(a) Proportion of fixations on each area of interest (AOI; see the diagram for an example) in each image by chimpanzees (n = 6) and humans (n = 16). (b) Proportion of images (n = 20) in which a fixation was observed in each AOI at each fixation order. The first 6 fixations are presented here. (c) The sum proportion of images at the first two fixations, showing that the results from (a) are evident no later than the first two fixations. The data are from the control condition. All data are shown as the difference from the chance level. T-tests compared between chimpanzees and humans, and between each species and the model (one-sample). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Error bars indicate s.e.m.

Figure 4

The saliency values at the first 6 fixations. The saliency value was standardized, and ranges from 0 (not salient) to 1 (highly salient). The data are taken from the control condition. n.s. not significant. Error bars indicate s.e.m.

Figure 5

(a) Examples of stimuli presented in each experimental condition. Note that the original stimuli were in color. (b) Proportion of fixations on the face in each image by chimpanzees (n = 6), humans (n = 16), and the model. (c) The mean saliency values at the first 6 fixations for chimpanzees and humans. All data are shown as the difference from the chance level. T-tests compared between chimpanzees and humans, and between each species and the model (one-sample). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s. not significant (the P values for Figure 5c are 0.75, 0.20, 0.29, 0.78, 0.74, 0.19, 0.26, 0.86, 0.19, for each condition, from left to right). Error bars indicate s.e.m.

Figure

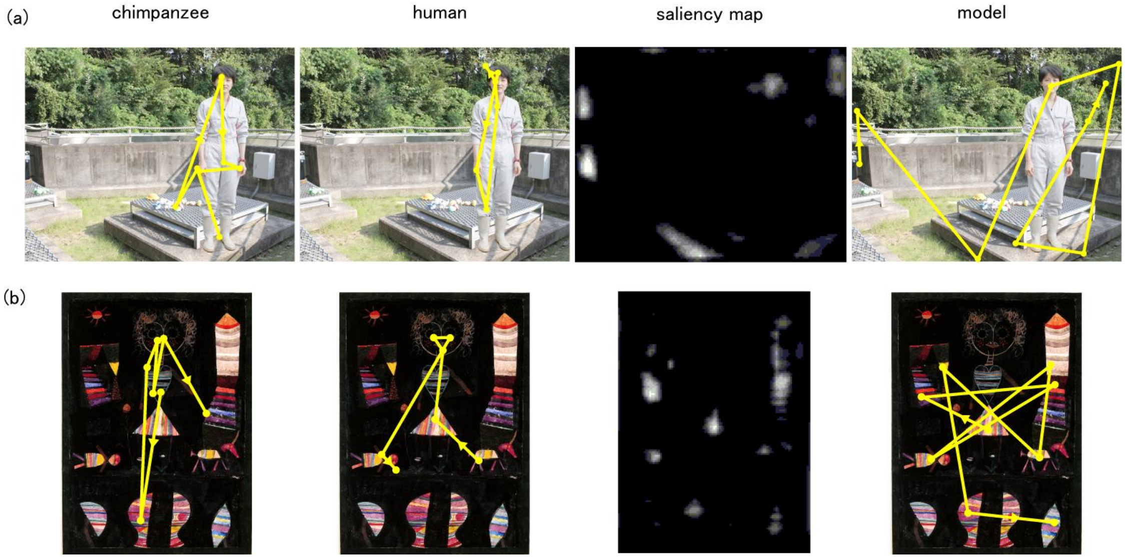


Figure 1

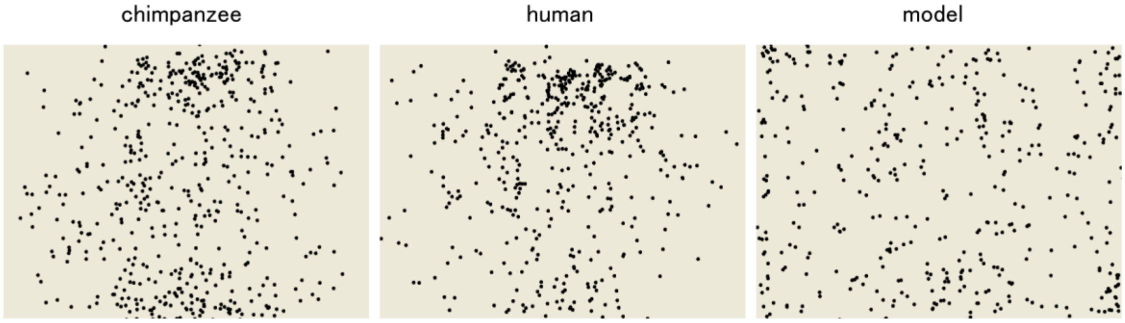
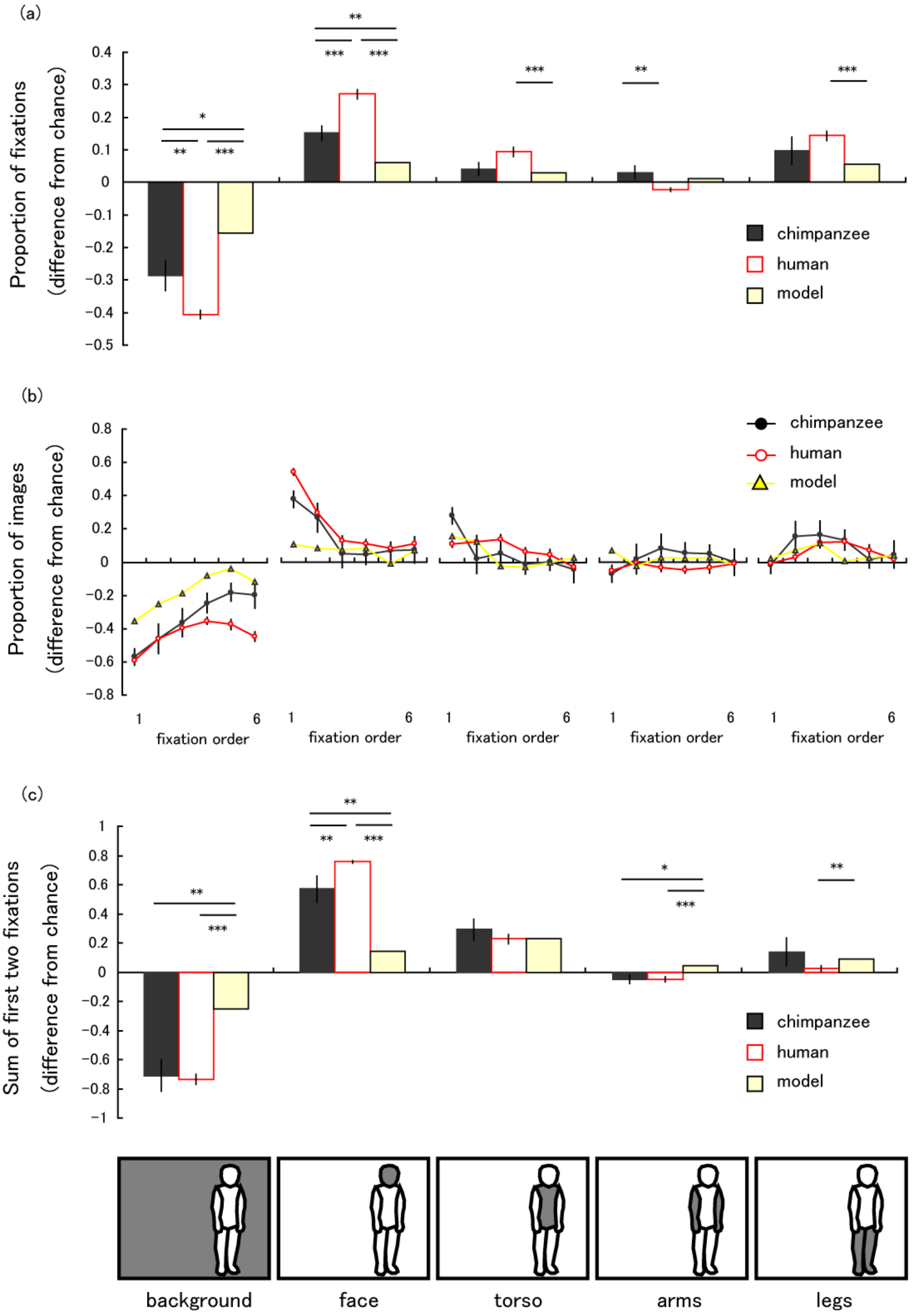


Figure 2



609

610 Figure 3

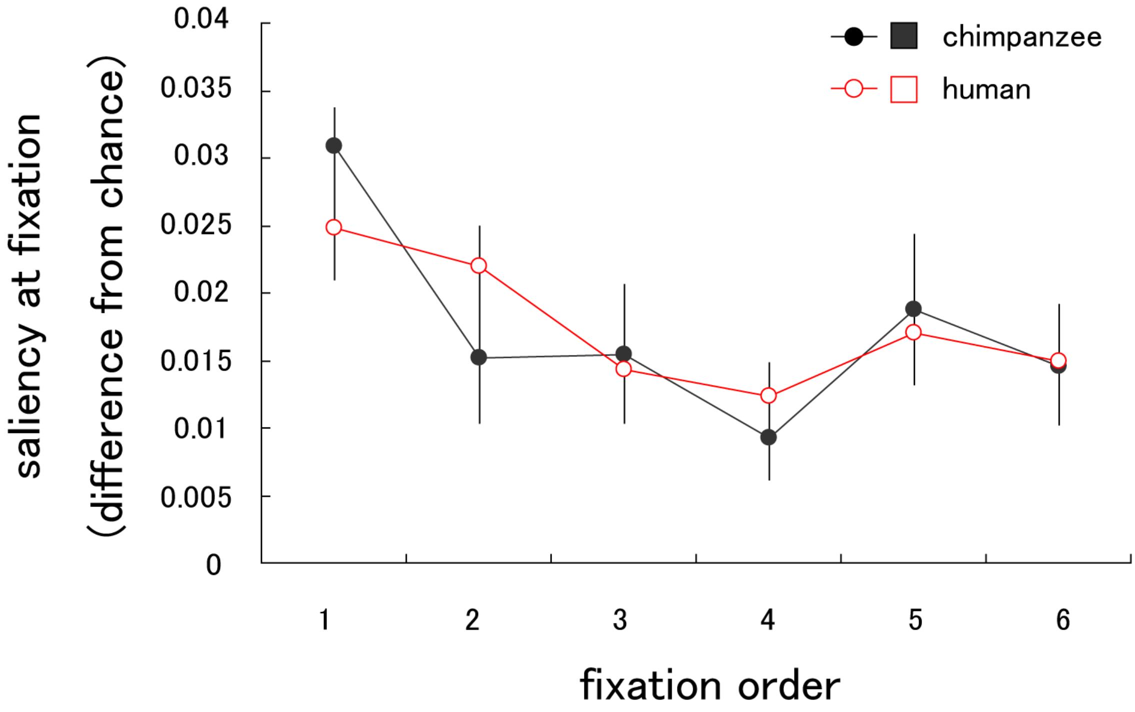

0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
saliency at fixation
(difference from chance)
1
2
3
4
5
6
fixation order
chimpanzee
human

Figure 4

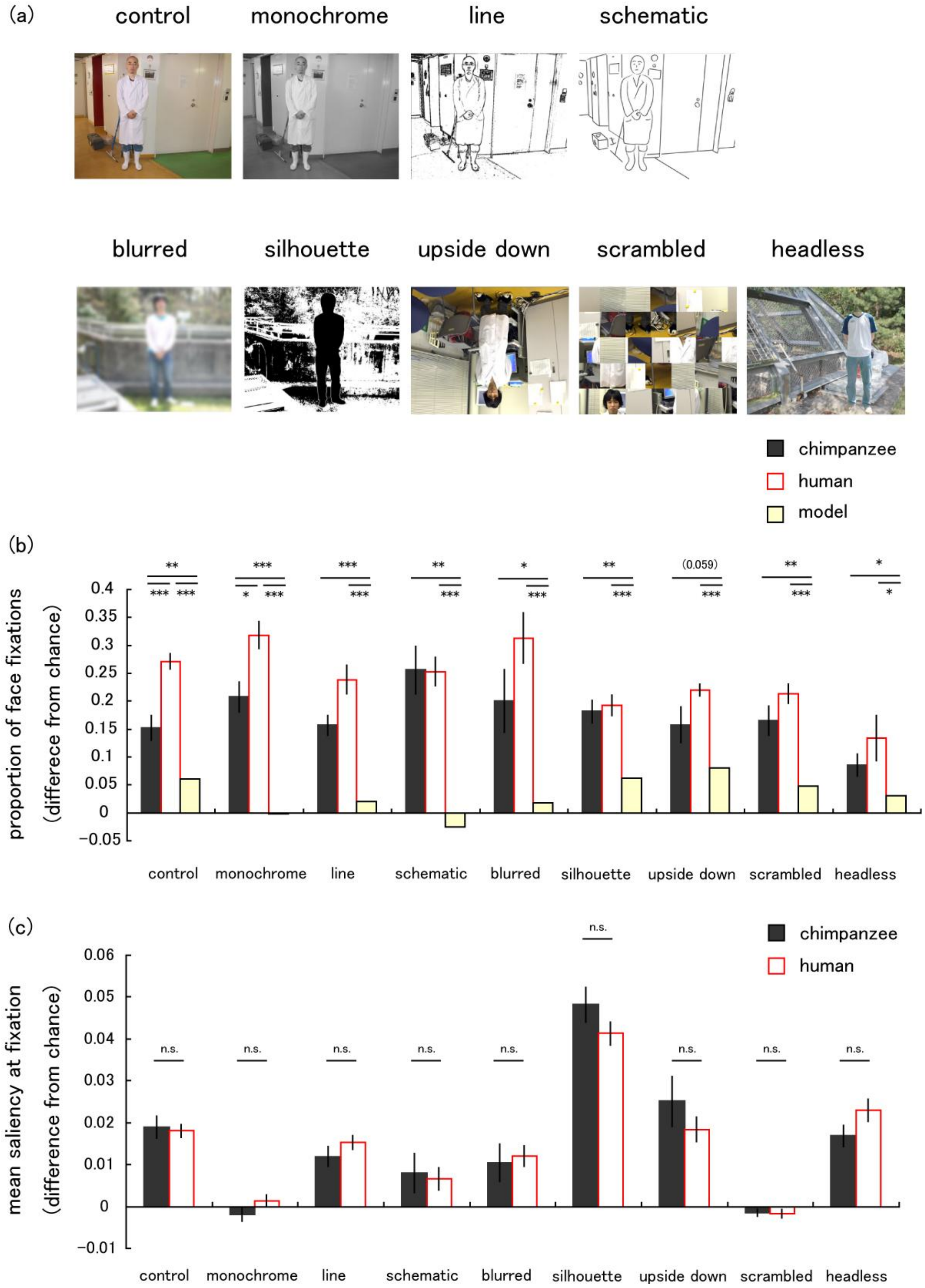

(a)
control
monochrome
line
schematic
blurred
silhouette
upside down
scrambled
headless
chimpanzee
human
model
(b)
proportion of face fixations
(difference from chance)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−0.05
**
*** ***
***
* ***
***
***
**
***
*
***
**
***
(0.059)
***
**
***
*
*
control
monochrome
line
schematic
blurred
silhouette
upside down
scrambled
headless
(c)
mean saliency at fixation
(difference from chance)
0.06
0.05
0.04
0.03
0.02
0.01
0
−0.01
n.s.
chimpanzee
human
control
monochrome
line
schematic
blurred
silhouette
upside down
scrambled
headless

Figure 5