



Universiteit
Leiden

The Netherlands

Emotions through the eyes of our closest living relatives: exploring attentional and behavioral mechanisms

Berlo, E. van

Citation

Berlo, E. van. (2022, May 19). *Emotions through the eyes of our closest living relatives: exploring attentional and behavioral mechanisms*. Retrieved from <https://hdl.handle.net/1887/3304204>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3304204>

Note: To cite this publication please use the final published version (if applicable).

Chapter 4

Attentional selectivity for emotions: humans and bonobos compared

Abstract

Perceiving emotions in others is at the foundation of higher-order social cognition. Currently, we do not fully understand how evolution shaped the cognitive processes underlying emotion perception. Bonobos (*Pan paniscus*) are our closest relatives, and have more developed brain structures involved in emotion processing and exhibit stronger emotion regulation abilities compared to other apes. This makes bonobos an important animal model for understanding the evolutionary development of emotion perception. Here, we investigated how bonobos and humans attend to emotionally-laden scenes in a preferential looking task using eye-tracking. With Bayesian mixed modeling, we established that in both species attention is spontaneously sustained to emotional scenes of conspecifics rather than heterospecifics. Moreover, scenes displaying distress held attention longest compared to neutral scenes, consistent with studies finding an initial attentional bias towards potentially threatening signals. Additionally, bonobos and humans attended longer to sexual scenes compared to neutral scenes, in line with sex being highly rewarding in both species. Humans also attended longer to scenes involving grooming and embracing, as well as play. These findings suggest that emotional signals are relevant to bonobos and that eye-tracking can provide a unique window into apes' affective capacities.

Based on:

Van Berlo, E., Kim, Y., & Kret, M. E. (2021). Attentional selectivity for emotions: humans and bonobos compared. *Manuscript submitted for publication*.

Introduction

Emotional expressions are the conduit through which information about experiences, desires, and intentions are communicated to others. Perceiving emotions is therefore an adaptive process that is crucial to humans and other social animals (Ferretti & Papaleo, 2019; Kret et al., 2020; Nieuwburg et al., 2021). In humans, emotional information is so important that the brain prioritizes its processing even when attentional resources are limited. There is now some evidence that this emotion-biased attention is not only present in humans, but also in great apes (Kano et al., 2018; Kano & Tomonaga, 2010a; Kret et al., 2016; Pritsch et al., 2017; Van Berlo et al., 2020a). However, the manner in which great apes perceive others' expressions of emotions is not yet well understood. As emotions drive not only behavior, but also cognitive mechanisms such as memory, learning, attention, and decision-making (Dukes et al., 2021), examining how they are perceived and recognized by non-human animals can help us reconstruct the evolutionary history of (social) cognition in our species. Moreover, it will allow us to improve our understanding of affective states in animals. Through a comparative framework, in this paper, we investigate emotion-biased attention to emotionally salient scenes in humans as well as our closest relatives, bonobos (*Pan paniscus*).

The human brain is adept at selectively processing information about conspecifics (other members of the same species), and especially emotional expressions are an important source of information that can trigger selective attention (Treue, 2003). In humans, a robust body of evidence shows that emotionally salient information such as smiles or angry faces is preferentially remembered and attracts attention when attentional resources are limited (Petersen & Posner, 2012). Sensory systems are not only tuned to favor facial expressions but also whole body expressions of emotions (Kret et al., 2013a) as well as emotional scenes (Kret & Van Berlo, 2021). In general, the findings show that an attentional preference for emotionally salient information is closely tied to survival, punishment, and reward, thus likely rooted in evolutionarily old mechanisms (Öhman et al., 2001b), and likely shared with other species.

The importance of perceiving and recognizing emotional expressions is not uniquely human. In the last decade, most research efforts on the perception and recognition of emotional expressions have focused on the great apes; our closest extant relatives. Great apes express a large range of behaviors to communicate their desires and intentions to others, and primate brain circuits that are involved in the processing of social and emotional information are similar to that of humans (Hirata

et al., 2013; Pinsk et al., 2009; Tsao et al., 2008). Great apes are known to automatically mimic facial expressions of others (Davila-Ross et al., 2008; Laméris et al., 2020; Palagi et al., 2019b; Van Berlo et al., 2020b), which is often linked to emotion contagion, or the convergence of emotional experiences (Pérez-Manrique & Gomila, 2022) (also see (Adriaense et al., 2020) for a critical review). Furthermore, great apes console conspecifics in distress (Clay et al., 2018). Work on physiological determinants of emotion perception indicates that in chimpanzees (*Pan troglodytes*), seeing or hearing conspecifics fight creates changes in cortisol level, heart rate variability, skin temperature (Dezecache et al., 2017; Kano et al., 2016), and temperature in the inner ear (Parr & Hopkins, 2000). Finally, there is some evidence that great apes can discriminate between emotional faces of conspecifics (Buttelmann et al., 2009; Parr, 2001), and that memory is enhanced for emotional stimuli (Kano et al., 2008). This converging evidence, therefore, suggests that great apes share our sensitivity to emotional cues.

Some work has looked into the continuity of emotional expressions and their perception and recognition across different species. All mammals likely share homologous brain structures underlying emotional networks (Panksepp, 2011), and already over a century ago, Darwin theorized that expressions of emotions are universally shared among certain animals. Indeed, within the primate lineage, there is some overlap between human expressions of emotions and that of other primates (Kret et al., 2020). Moreover, one study showed that orangutans (*Pongo pygmaeus*) and human children looked longer at fearful human expressions, and the silent bared-teeth display of orangutans (Pritsch et al., 2017). These results suggest that emotional faces that carry a similar meaning in the two species (i.e., fear) are relevant enough to attend to. While this work is promising, it is clear that more research is needed to understand the phylogenetic continuity of emotional expressions and their perception across species. There is still a great deal to explore in terms of the mechanisms underlying emotion perception in great apes, and specifically, very little work has examined the attentional processes underlying emotion perception in these animals.

Two studies looking into implicit attention using a dot-probe paradigm found that bonobos attend faster to emotionally-laden scenes of others compared to neutral scenes (Kret et al., 2016), and especially of unfamiliar conspecifics (Van Berlo et al., 2020a). This effect has not been found in chimpanzees (Kret et al., 2018; Wilson & Tomonaga, 2018), but it is not yet clear whether methodological considerations (e.g., ecological validity of stimuli (Kret et al., 2018), or stimulus presentation duration

(Wilson & Tomonaga, 2018)) contributed to the null-results. Moreover, two recent studies showed that in apes, emotional cues such as the play face (Laméris et al., 2022) or snakes and food items (Hopper et al., 2021a) impact reaction time on an emotional Stroop task. Finally, eye-tracking studies with chimpanzees and orangutans revealed spontaneous gazing at negatively valenced emotional signals (Kano & Tomonaga, 2010a; Pritsch et al., 2017). Combined, these findings suggest that like in humans, apes' attention is tuned to emotionally salient information. However, different methodologies may tap into different attentional processes (with e.g., Stroop tasks measuring interference in attention, and dot-probes and eye-tracking likely measuring bottom-up or top-down attention), and few studies have directly compared how humans and great apes view emotional expressions or emotionally salient scenes. The aim of the current study is to further examine how apes, specifically bonobos, compare to humans in their allocation of attention to emotionally valent stimuli using eye-tracking. Compared to other apes, bonobos show marked differences in brain areas involved in social cognition, with a higher degree of connectivity and volume in the amygdala (regulating emotions, attention, memory, and social decision-making) and subgenual anterior cingulate cortex (regulating positive affect and arousal) (Issa et al., 2019; Stimpson et al., 2016). This makes them an interesting referential model to reconstruct the evolution of emotional capacities (Gruber & Clay, 2016). At the same time, bonobos are underrepresented in socio-cognitive studies due to their rarity and zoos and their endangered conservation status (Fruth et al., 2016). As such, bonobos' unique socio-emotional characteristics warrant a closer look at how this species perceives emotions.

To this end, we used a preferential looking paradigm with eye-tracking to investigate whether attention of bonobos (experiment 1) and humans (experiment 2) is preferentially sustained to emotionally-laden scenes of conspecifics or heterospecifics (i.e., the other species). Previous findings show that emotionally salient signals modulate the early stages of processing social signals (Hopper et al., 2021a; Kret et al., 2016; Laméris et al., 2022; Van Berlo et al., 2020a). Building on this, we expect that if emotions hold relevance to bonobos beyond an initial attentional bias, they will show a longer looking duration to emotional compared to neutral scenes, similar to humans. Moreover, we expect that bonobos and humans also attend longer to emotional scenes of heterospecifics, as there is some continuity between emotional expressions of great apes and humans (Kret et al., 2018).

Experiment 1: Examining biased attention to emotions in bonobos

Method

Participants

Our sample included four bonobos (Besede [12 yo], Kumbuka [18 yo], Monyama [7 yo], and Zuani [~16 yo]; all female) that are part of a social group of 12 individuals housed in the primate park Apenheul, Apeldoorn, The Netherlands. Except for Zuani, all bonobos took part in two prior touchscreen studies (Kret et al., 2016). At time of testing, none of the bonobos were pregnant nor were on contraceptives. During winter time (from November to the end of March), the park is closed for visitors, allowing us to conduct experimental research. All but one individual (Zuani) were born and raised in captivity. During non-testing hours, the bonobos had access to a 2812m² outdoor and 158m² indoor enclosure, and testing took place in the indoor enclosure.

The zoo kept the bonobos separated into two groups that varied in group composition on a weekly basis to mimic naturalistic fission-fusion dynamics. During testing periods, only one group of the bonobos was given access to the test apparatus. For ethical reasons, the group was never split further. This meant that when one individual was tested, its group members were present nearby. Water was available ad libitum, and food (a variety of vegetables, fruits, and branches and leaves) was provided four to five times a day, as well as nutritionally balanced mash.

Tests with the bonobos followed the EAZA Ex situ Program (EEP) guidelines, formulated by the European Association of Zoos and Aquaria (EAZA). Bonobos voluntarily participated in the experiment and were never restrained or forced to take part. Furthermore, only positive reinforcement (juice) was used during training and testing, and juice was also offered to the bonobos that did not take part in the experiment. Data were collected between February 2017 – March 2017, and December 2017 – April 2018.

Equipment

Our setup is comparable to those in other research facilities (see Hopper et al., 2020), and involves one PC running Tobii Studio (v.3.4.8), two computer screens (one for the experimenter, one for the participant, 1280x1024 pixels), a webcam to record the bonobos while they were tested, and a Tobii X2-60 eye tracker mounted on one of the screens. One computer screen, together with the eye

tracker and the webcam, was placed inside a wooden box inside the bonobos' enclosure (Figure 1).

The front of the wooden box was a 3 mm thick, scratch proof polycarbonate plate. At mouth's height, a drinking nozzle was attached to the panel. During the experiment, bonobos were rewarded with diluted juice (1 part syrup, 5 parts water) at short intervals (roughly every 5 seconds), and provided through the nozzle. To minimize distractions, other bonobos present in the enclosure were rewarded with the same juice by the caretaker, after they performed a body-part training that is used for veterinarian purposes. Bonobos were familiar with drinking from the nozzle because their enclosures were also fitted with these nozzles for drinking water. The computer and the other screen for the experimenter were located outside of the enclosure. This second screen displayed Tobii Studio Pro's Live Viewer, enabling the experimenter to track where bonobos were looking in real time.

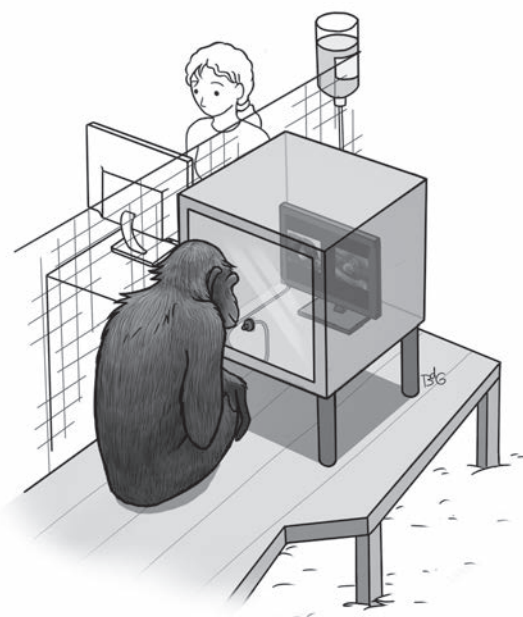


Figure 1. Drawing of the setup at primate park Apenheul. Illustration by Brenda de Groot.

Stimuli

Stimuli consisted of emotional and neutral scenes selected from previously validated sets (bonobos: Kret et al., 2016, humans: Kret & van Berlo, 2021; van Berlo et al., 2021). While it is common in psychological research to use isolated facial expressions of

emotions (see .e.g., Adolphs, 2002), we used a combination of expressions as well as emotional scenes. Emotional scenes can convey more contextual information, as they contain whole-body expressions that can communicate emotions as well as action intentions (De Gelder et al., 2010). Furthermore, previous studies have shown that emotional scenes modulate attention in a similar way as facial expressions (e.g., Kret et al., 2016; Kret & van Berlo, 2021; van Berlo et al., 2021), therefore indicating that they provide sufficient emotional information to the participant.

Emotional scenes involved individuals engaged in socially relevant behavior and/or having an emotionally relevant facial or bodily expression. Though it can be argued that we do not exactly know what bonobo emotions are, we do know the social relevance of certain facial expressions (such as the fear-grin, the relaxed open-mouth play face and yawning) and socio-emotional behaviors (sex, grooming) (De Waal, 1988). The fear-grin is often expressed during stressful situations and agonistic interactions, while the relaxed open-mouth face (or play face) is expressed during playful interactions (De Waal, 1988). Yawning is a widespread behavior in vertebrates and it is highly contagious (Demuru & Palagi, 2012; Massen et al., 2015; Palagi et al., 2014; Van Berlo et al., 2020b). Its contagiousness is linked to social closeness, and yawning could therefore serve a social function (Casetta et al., 2021; Norscia et al., 2020). Furthermore, yawns capture immediate attention in bonobos (Kret et al., 2016). Other socio-emotional behaviors that are relevant to bonobo society are sex and grooming. Bonobos use sex to prevent or resolve conflicts and reduce stress levels (De Waal, 1988). Grooming is an important social behavior used to form and strengthen social bonds between individuals (Dunbar, 1991). As such, emotional scenes in our task consisted of one or more bonobos playing, having sex or displaying an erection (male) or a large swelling (female), grooming, displaying distress, and yawning. Neutral scenes consisted of one or more bonobos lying down, sitting or walking with a neutral facial expression (see Tables S1 and S2 in supplements).

To make direct comparisons between bonobos and humans possible, we selected emotional scenes of humans that were equivalent to or an approximation of the emotional bonobo scenes. The stimuli consisted of humans playing, having sex (specifically: engaged in a romantic embrace), embracing ("grooming"), displaying distress (crying), and yawning. As there is no clear human equivalent for grooming in humans, we opted to use embracing as it is a reflection of social closeness and involves physical contact, just like grooming (Forsell & Åström, 2012). Neutral scenes of humans depicted one or more

individuals lying down on grass, sitting, walking, or cycling with a neutral facial expression (see Table S3 for more information on the composition of the scenes). In total there were 10 unique stimuli per emotional category (5) and per species (2), as well as 100 unique neutral stimuli, as each emotional stimulus was matched with a neutral stimulus. Stimuli consisted of a subset of the validated sets by Kret et al. (2016) and van Berlo & Kret (2021). They were colored pictures with a dimension of 500x430 pixels, matched on luminance level and number of individuals depicted as much as possible.

Calibration

Before commencing testing, we conducted a manual two-point calibration using the infant calibration procedure in Tobii Studio. We used a relatively small number of reference points because apes tended to look only very briefly at the points. However, two-point calibrations are often used in great ape research as they are reasonably sufficient for the research questions asked, and also attainable given the constraints of working with animals (Hopper et al., 2021b). A small video displaying penguins (270x155 px) was used for the reference points. Calibrations were repeated until a sufficient calibration was obtained (i.e., Tobii Studio indicated no large calibration errors). For each individual, we continued using their first successful calibration throughout the entire experiment. To make sure that the calibration remained sufficient over time, we showed bonobos a 9-point grid before the start of each test session and visually inspected the accuracy of the calibration (see supplements for more information regarding calibration).

Procedure

Before commencing the experiment, bonobos were familiarized with the setup by showing each individual at least two sets of 10 trials with stimuli of animals and objects. Due to time constraints, once four individuals were able to drink from the setup during most of the practice sessions, we moved on to the experiment. Bonobos then participated in an experiment in which they could freely view socio-emotional and neutral scenes (presented at the same time) of unfamiliar conspecifics and of unfamiliar humans (Figure 2). Because the bonobos were not physically separated from other group members, the progression from trial to trial was manually controlled by the experimenter. This was done to ensure that data would only be collected when bonobos were attending the screen, and not when there were disturbances such as individuals moving away from the setup or individuals being distracted by others.

Each test session consisted of 10 trials and started with a 9-point grid to check for calibration accuracy, shown until the experimenter manually continued the experiment. The presentation of the grid was followed with a black screen displayed for 4 seconds, and subsequently followed by a fixation video (a sped-up nature movie) positioned in the middle of the screen. Only when the participant's fixation was on the video for more than one second, the experimenter moved on to the next trial sequence. Bonobos were then presented with two stimuli on the left and right side of the screen; one emotional and one neutral image (location was counterbalanced). Stimuli were presented for 3 seconds, in accordance with previous eye tracking tasks with great apes (e.g., Kano et al., 2015, and see Hopper et al., 2020, for a review). After 3 seconds, the experiment continued with a black screen shown for 4 seconds, and this concluded a trial. After 10 trials, the task ended automatically. Bonobos first completed all the trials with bonobos before moving on to the human stimulus set.

On average, the bonobos were tested on 33.5 sessions ($SD = 3.12$), and 355 trials (trials with bonobo stimuli: $M = 191.5$, $SD = 23.84$; trials with human stimuli: $M = 163.5$, $SD = 46.57$). Furthermore, trials were repeated in order to compensate for data loss (e.g., due to disruptions by other bonobos). On average, all unique stimulus combinations were repeated 3.59 times ($SD = 1.49$).

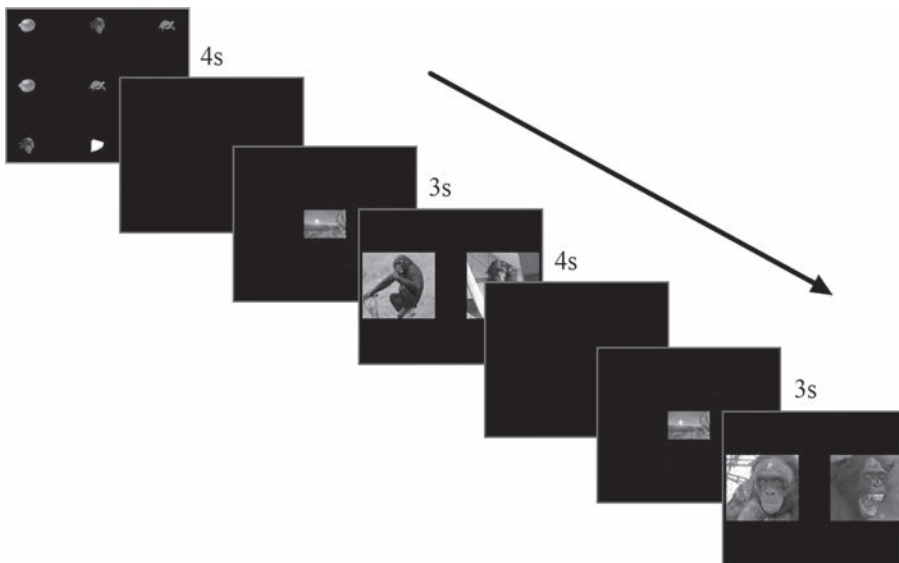


Figure 2. Trial sequence for participants. The test started with a 9-point grid, and each trial started with a black screen (4s), followed by a fixation video. Finally, two stimuli were shown on both sides of the screen (3s).

Data preparation

Because we used only one calibration per bonobo throughout the entire experiment rather than re-calibrating the bonobos for each experimental session, before analyzing the data, we checked whether the raw fixation data per bonobo and per session reasonably matched with the areas of the stimuli on the screen. We plotted all the gaze data for each individual onto a mapping of our screen and the location of the stimuli on the screen. We found that for two apes, in some sessions there were consistent shifts in gaze data to the left or to the right relative to the position of the stimuli on the screen.

Using K-means clustering in a custom script in Python, we established the difference between the gaze data collected by the eye tracker and the true centroids of the stimuli displayed on the left and right side of the screen. Based on these findings, we corrected 37/54 sessions for Monyama (average offset of +134 pixels), and 39/46 sessions for Zuani (average offset of -141 pixels) (see supplements for more information on how we corrected these sessions).

Next, two regions of interest (ROIs) were defined in Tobii Studio. We drew a 500x512 square around each of the stimuli (sized 500x430, thus the ROI was slightly larger in length than the stimuli to compensate y-axis inaccuracies in the gaze data; Figure S3). Through Tobii Studio's *Statistics* option, we extracted data on *Total Fixation Duration* per ROI using the Tobii Fixation Filter. Finally, after processing the *Total Fixation Duration* gaze data, we noticed that there were 19 trials where the total fixation duration was higher than 3 seconds ($M = 4.47s$, $SD = 1.09$), possibly due to Tobii registering a fixation that extended beyond the duration of the stimulus presentation. These isolated cases were removed from further analyses.

Statistical analyses

We used Bayesian mixed modeling in order to assess support for our hypotheses. We were interested in the total looking duration to emotional stimuli across trials. Our dependent variable was therefore the proportional looking duration to emotional stimuli (based on Tobii Studio's *Total Fixation Duration*. From here on: PLD_{emotion}), calculated by dividing the looking duration to the target by the sum of the looking duration to the target and distractor. The target was the emotional stimulus, and the distractor a neutral stimulus of the same species. A PLD_{emotion} higher than 0.5 indicates a longer looking duration to the target.

Within a three-second trial, bonobos on average looked 1.93 ($SD = 0.78$) seconds to the target and distractor combined (raw, unweighted values) when bonobo stimuli were displayed, and 2.04 seconds ($SD = 0.72$) when human stimuli were displayed. Thus, as some trials were more reliable than other trials, and to account for variation in overall attention to the target and distractor during the trial, we also calculated a weight for each trial. We calculated the weight by dividing the sum of the looking duration to the target and distractor by the average looking duration to the target and distractor per participant. The weight gives more importance to trials in which the participant paid more attention to the stimuli, and less importance to trials where participants were relatively inattentive. Weights were added to our models for all measures of interest ($M = 1$, $SD = 0.38$, range [0.01 – 1.69]).

We used zero-one-inflated beta (ZOIB) regression to account for 0's, 1's, and the data between the range [0, 1]. For our measure of interest, proportional looking duration to emotional stimuli ($PLD_{emotion}$) across trials, we ran three separate models. In the first model, we examined whether the $PLD_{emotion}$ was higher than 0.5, i.e., whether participants look more than 50% of the time to emotional stimuli. In the second model, we assessed whether the $PLD_{emotion}$ differed between *Species* displayed on the stimuli (i.e., human or bonobo). In the third model, we zoomed in on the specific emotion categories and examined whether there was an interactive effect between *Species* and *Emotion Category* on the $PLD_{emotion}$.

In all of our models, we used weakly informative priors, specifically a student-t (default) prior ($df = 3$, $M = 0$, $SD = 2.5$) for the standard deviation coefficient, and a normal distribution ($M = 0$, $SD = 1$) for all other coefficients. *Species* and *Emotion Category* were treatment (dummy) coded. For each model, we report the median estimate coefficient, together with the 89% credible interval (either the Highest-Density Credible Interval [HDI; a “summary credible interval” for the posterior distribution] or the 89% Highest-Posterior Density [HPD; the shortest possible credible interval]). For comparisons between conditions, we report the odds ratio (OR). We also report the probability of direction (pd), which indicates the certainty that an effect goes in a specific direction (Makowski et al., 2019c, 2019a).

To establish model convergence, we followed the guidelines set out in the WAMBS checklist by de Paoli & van de Schoot (2017). We assessed trace and autocorrelation plots, the Gelman-Rubin diagnostic values (convergence indicated by a value close to 1), and density histograms for the posterior distributions. We conducted all of our analyses using RStudio (v. 1.4.1106, R Core Team, 2020) and the package brms (Bürkner, 2017, 2018).

Results

In model 1, we did not find evidence that the looking duration to emotional stimuli (PLD_{emotion}) of bonobos is higher than 50% ($Mdn = 0.50$, 89% CI [0.46 – 0.53], $pd_+ = 58\%$, Table 1), meaning that bonobos did not reliably look longer at emotional stimuli of other bonobos and humans compared to neutral stimuli.

In our second model, we examined the effect of *Species* on the stimulus (bonobo or human) on PLD_{emotion} . We found that for both species, the PLD_{emotion} did not reliably deviate from 50% (bonobo stimuli: $Mdn = 0.52$, 89% CI [0.48 – 0.55], $pd_+ = 82\%$; human stimuli: $Mdn = 0.47$, 89% CI [0.44 – 0.51], $pd_+ = 90\%$, Table 1). However, we found robust evidence for a difference between the PLD_{emotion} for stimuli depicting humans and those depicting bonobos ($OR = 1.17$, 89% HPD [1.09 – 1.26]); bonobos looked relatively longer to emotional stimuli of other bonobos than to emotional stimuli of humans.

In our third model where we zoomed in on the specific emotion categories, we found robust evidence for a longer PLD_{emotion} of stimuli depicting distressed bonobos ($Mdn = 0.54$, 89% CI [0.51 – 0.58], $pd_+ = 96\%$, Table 1 and Figure 3a). For the sex category, the effect was in the expected direction (as indicated by the probability of direction; $pd_+ = 93\%$), but not very strong ($Mdn = 0.54$, 89% CI [0.50 – 0.57]). Finally, we found robust evidence for a lower PLD_{emotion} of stimuli depicting humans having sex ($Mdn = 0.40$, 89% CI [0.36 – 0.43], $pd_+ = 100\%$, Table 1, Table S4, and Figure 3c).

Table 1. Overview of results per factor's level of interest for the three models. Robust effects are in bold.

Model	Species on stimulus	Emotion Category	Median	89% CI	pd
1 (Intercept)	Bonobo and human	All	0.50	0.46 – 0.53	0.58
2 (Species)	Bonobo	All	0.52	0.48 – 0.55	0.82
3 (Species*Emotion Category)	Human	All	0.47	0.44 – 0.51	0.90
	Bonobo	Distress	0.54	0.51 – 0.58	0.96
		Yawn	0.50	0.46 – 0.54	0.51
		Groom	0.49	0.45 – 0.53	0.61
		Sex	0.54	0.50 – 0.57	0.93
		Play	0.49	0.46 – 0.53	0.61
	Human	Distress	0.53	0.48 – 0.56	0.85
		Yawn	0.47	0.43 – 0.51	0.91
		Groom	0.49	0.45 – 0.53	0.73
		Sex	0.40	0.36 – 0.43	1.00
Yawn		0.50	0.46 – 0.54	0.55	

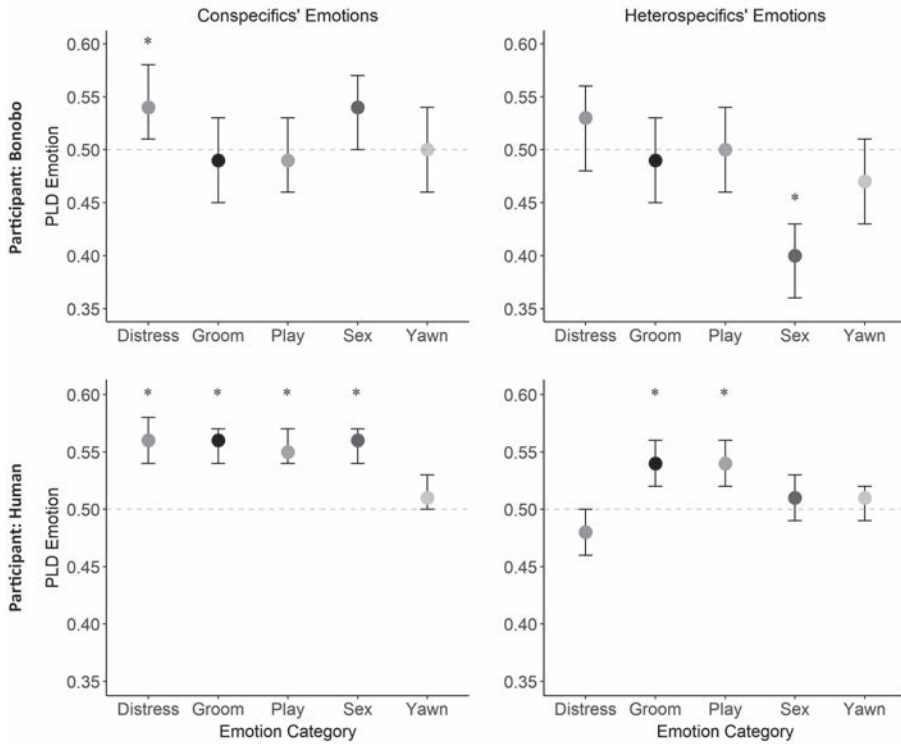


Figure 3. Graphs displaying the proportional looking duration to emotional stimuli (PLD Emotion) of conspecifics and heterospecifics by bonobos and humans. Error bars reflect the 89% credible interval, dots represent the median. Asterisks indicate robust effects.

Conclusion

Overall, we found that bonobos attended longer to emotional scenes of conspecifics (i.e., other bonobos) than to emotional scenes of heterospecifics (i.e., humans). When viewing emotional scenes of conspecifics, bonobos preferred to look at distressed others and sexual scenes compared to neutral scenes.

Experiment 2: Examining biased attention to emotions in humans

Method

Participants

Participants were visitors of primate park Apenheul. In total, 100 adults participated (Age category 18-30: $N = 57$, 31-50: $N = 33$, 51-80: $N = 9$; 58 women, 41 men). We tested participants in the visitor's area of the bonobo enclosure, where we set up a long table with cubicles in which we could test participants. We actively recruited participants by approaching them when they walked past the indoor bonobo enclosures and our setup. Participants were told that the bonobos participated in several experiments, and that we were now collecting human data using the same tasks. Data were collected between April and May 2017.

Stimuli

The same stimulus material was used as in Experiment 1. Like the bonobos, human participants saw both bonobo and human stimuli (see supplements Tables S2 and S3 for more information on the stimuli).

Equipment

Humans were tested near the indoor enclosures of the bonobos. We had a special corner dedicated to comparative research, consisting of two cubicles. One cubicle was specifically for this study. We tested participants using a 19" laptop (1920x1200 pixels) and a Tobii X2-60 eye tracker with Tobii Studio.

Calibration

Human participants were calibrated using the 5-point automated calibration procedure in Tobii Studio. Calibrations were accepted when the error displayed after finishing the calibration was minimal (less than a degree).

Procedure

Human participants were actively recruited by research assistants in the park. Visitors were approached and asked if they were interested in participating in a short, 10 minute task that was also completed by the bonobos. If visitors were interested, they were given a consent form to sign, thereby giving the experimenter permission to use their data for further analyses and publication. Participants then sat down

behind the laptop and the experimenter started the 5-point calibration procedure. After finishing the calibration, participants filled in their age and sex in the task, and then the experimenter started the task. After finishing the task, participants were given the opportunity to ask more questions about the study and were given a debrief form containing the explanation and goal of the study.

To make direct comparisons between bonobos and humans possible, the difference in the task completed by both species was kept to a minimum. Whereas bonobos first completed all trials with bonobo stimuli and then trials with human stimuli, human participants first completed 10 trials of either bonobo or human stimuli, followed by 10 trials of the opposite species, and then followed by yet another 10 trial of the species that they started out with. Human participants thus completed 30 trials in one session. We created 10 versions of the task to control for order effects. In version 1, participants completed 10 trials with human scenes, followed by 10 trials containing bonobo scenes, and then another 10 trials with human scenes. In version 2 of the task, participants started with 10 trials with bonobo scenes, followed by 10 trials with human scenes, and again 10 trials with bonobo scenes. We continued alternating this sequence for the remaining 8 versions. We tested 10 participants per version of the task. This meant that every participant saw stimuli only once, but since we had 100 unique stimuli (50 combinations) and 10 versions of the task, every stimulus combination was repeated three times overall, resulting in 30 datapoints per stimulus combination.

In human participants, the trial sequence was fully automated. Because the bonobos could not be instructed, humans received minimal instructions as well, namely that they should pay attention to the screen and not move their head too much. Similar to the bonobo version of the task, humans started out with a 9-point grid that was shown for 3 seconds. The grid was followed by a black screen for 4 seconds, and then the fixation video for 3 seconds. Next, two stimuli of an emotional and a neutral bonobo or human were shown for 3 seconds, followed by a black screen shown for 4 seconds (Figure 2). Participants could take a short break between every set of 10 trials where they were allowed to move their head, but were requested to remain seated. When ready, participants could continue to the next 10 trials by pressing the space bar, followed by the 4 seconds black screen indicating the start of a new trial. At the end of the last set of 10 trials, participants saw a screen on which they were thanked for their participation.

Data preparation

After data collection finished, we realized that in version 3, 6, and 9 of the task, we accidentally showed one stimulus twice. These repetitions were removed from further analyses (31 datapoints). Furthermore, for five participants, there was a technical malfunction with the eye tracker resulting in 60% or more data loss. Thus, the data of these participants were excluded from further analyses. Similar to Experiment 1, we created ROIs in Tobii Studio, and extracted data on *Total Fixation Duration* per ROI using the Tobii Fixation Filter (see Experiment 1).

Statistical analyses

The analysis procedure for humans was similar to that of the bonobos. We were interested in the total looking duration to emotional stimuli across trials, thus calculated the proportional looking duration to emotional stimuli (PLD_{emotion}). Within the three-second trial window, human participants looked on average 2.66 s ($SD = 0.38$) to the target and distractor combined (raw, unweighted values) when human stimuli were displayed, and 2.64 s ($SD = 0.43$) when bonobo stimuli were displayed. Similar to what we did for the bonobos, we calculated the weight of a trial depending on how long a participant looked at the stimuli relative to their average looking duration to the stimuli ($M = 1$, $SD = 0.15$, range [0.005 – 1.51]).

For the PLD_{emotion} across trials, we ran Bayesian zero-one-inflated beta regression models, similar to Experiment 1. Model 1 involved only the intercept, model 2 examined effects of *Species* displayed on the stimulus, and model 3 assessed an interaction effect of *Species* and *Emotional Category*. All models included a random intercept for *ID* (participant), and used weakly informative priors. Each model was checked using the WAMBS checklist (Depaoli & van de Schoot, 2017). We conducted all of our analyses using RStudio (v. 1.4.1106, R Core Team, 2020) and the package brms (Bürkner, 2017, 2018).

Results

In model 1, we found robust evidence for a longer PLD_{emotion} in human participants ($Mdn = 0.53$, 89%CI [0.52 – 0.54], $pd_{+} = 100\%$, Table 2), meaning that humans looked relatively longer to emotional stimuli than to neutral stimuli. In the second model with *Species* included as a factor, we found robust evidence for longer PLD_{emotion} of stimuli depicting humans ($Mdn = 0.55$, 89% CI [0.54 – 0.56], pd_{+}

= 100%, as well as those depicting bonobos ($Mdn = 0.52$, 89%CI [0.51 – 0.52], $pd_+ = 100\%$, Table 2). Additionally, we found robust evidence for a difference between $PLD_{emotion}$ to stimuli depicting bonobos or humans ($OR = 0.88$, 89% HDI [0.84 – 0.91]). Thus, the looking duration to emotional stimuli was higher for human emotions than for bonobo emotions.

When examining the specific emotion categories per species in model 3, we found robust evidence that humans looked longer at other humans in distress ($Mdn = 0.56$, 89% CI [0.54 – 0.58], $pd_+ = 100\%$), having sex ($Mdn = 0.56$, 89% CI [0.54 – 0.57], $pd_+ = 100\%$, or humans playing ($Mdn = 0.55$, 89% CI [0.54 – 0.57], $pd_+ = 100\%$), grooming/embracing ($Mdn = 0.56$, 89% CI [0.54 – 0.57], $pd_+ = 100\%$, Table 2 and Figure 3b). For the yawning category, the effect was in the expected direction ($pd_+ = 90\%$), but it was weak ($Mdn = 0.51$, 89% CI [0.50 – 0.53]). For the bonobo category, we found robust evidence for $PLD_{emotion}$ of stimuli of grooming ($Mdn = 0.54$, 89%CI [0.52 – 0.56], $pd_+ = 100\%$) and playing bonobos ($Mdn = 0.52$, 89% CI [0.52 – 0.56], $pd_+ = 100\%$). We also find a weak effect of humans looking towards neutral scenes that were matched with distressed bonobos ($Mdn = 0.48$, 89% CI [0.46 – 0.50], $pd_- = 96\%$, Table 2, Table S5, and Figure 3d).

Table 2. Overview of results per factor level of interest for the three models. Robust effects are in bold.

Model	Species on stimulus	Emotion Category	Median	CI 89%	pd	
1 (Intercept)	Bonobo and human	All	0.53	[0.52 – 0.54]	1.00	
2 (Species)	Bonobo	All	0.52	[0.51 – 0.52]	1.00	
3 (Species*Emotion Category)	Human	All	0.55	[0.54 – 0.56]	1.00	
		Distress	0.48	[0.46 – 0.50]	0.96	
		Yawn	0.51	[0.49 – 0.52]	0.78	
	Bonobo	Groom	0.54	[0.52 – 0.56]	1.00	
		Sex	0.51	[0.49 – 0.53]	0.87	
		Play	0.54	[0.52 – 0.56]	1.00	
		Human	Distress	0.56	[0.54 – 0.58]	1.00
			Yawn	0.51	[0.50 – 0.53]	0.90
			Groom	0.56	[0.54 – 0.57]	1.00
Sex	0.56		[0.54 – 0.57]	1.00		
		Play	0.55	[0.54 – 0.57]	1.00	

Conclusion

In general, humans attended longer to emotional scenes compared to neutral scenes. This general emotion bias was also present for scenes of bonobos, although it was less pronounced. Humans tended to look longer at all types of emotional scenes involving humans, although evidence for a bias towards yawning was not robust.

Discussion

Emotions and their perception in non-human animals are intriguing, yet elusive (Anderson & Adolphs, 2014). To progress our understanding of when and how the brain evolved to efficiently process emotionally salient cues, we set out to study attention for emotions in our closest relatives, bonobos, and in humans. We found that both species preferentially attended to conspecific over heterospecific emotional scenes. Moreover, attention appeared to be strongly tuned to conspecifics in distress. Furthermore, bonobos showed an (albeit weak) attentional bias towards sex stimuli, while humans tended to look longer at emotional scenes across all categories. Below, we first discuss the findings in experiment 1, followed by a comparison between the results of humans (experiment 2) and bonobos.

In the first experiment, we partially confirmed our expectation that bonobos preferentially look at emotional scenes over neutral scenes of other bonobos and humans. Seeing distressed others can be a very salient cue, for instance, because detecting potential social or environmental threats can be crucial to survival (Öhman et al., 2001a). Similarly, in bonobos, socio-sexual interactions play a major role in preserving stability in the group (for instance to ameliorate tension) (Genty et al., 2015), and sexual stimuli may therefore receive enhanced attention. Bonobos showed no pronounced attention bias towards playful, grooming, or yawning scenes. These results are somewhat surprising, as a previous study found an implicit attentional bias towards scenes depicting yawning and grooming (in addition to sexual scenes) (Kret et al., 2016), and one study found that playful scenes interfered with bonobos' attention in an emotional Stroop task (Laméris et al., 2022). However, this could be explained by the results capturing different attentional processes, with reaction time paradigms possibly tapping into bottom-up attention, and eye-tracking paradigms having the potential to also measure top-down attention

(Belopolsky et al., 2011). Our data are not fine-grained enough to disentangle the two processes, as eye-tracking is not yet optimized for apes. However, given that bonobos do appear to have an immediate bias towards playful and yawning scenes, but not attend to them longer when given the opportunity to do so, these categories likely elicit a bottom-up attentional process. Future studies could focus on distinguishing between bottom-up and top-down attentional processes, especially now that new eye-tracker models allow for greater sampling rates and are more forgiving in terms head movements (which is important when working with animals).

We expected a similar (but less pronounced) attentional bias pattern when the bonobos viewed emotional scenes involving humans. Although we did not find robust evidence that bonobos looked longer at any of the human emotions compared to neutral scenes, the looking duration pattern was similar to viewing bonobo scenes. Specifically, bonobos seemed to look slightly longer at humans expressing distress. These findings may be explained by human expressions of distress sharing similar morphological action tendencies as bonobo expressions of distress. For instance, a general feature of fearful expressions is the tendency to make oneself small, indicating weakness or submissiveness, and this occurs in humans and many other primates (Kret et al., 2020). Moreover, the scream face of apes shares a lot of morphological similarities with its human equivalent (Parr et al., 2007). Furthermore, the finding that bonobos looked longer to the neutral scenes that accompanied a human sex scene is curious. In the neutral scenes, people wore more clothing, which may provide a salient cue (e.g., due to more variation in patterns) than seeing people without clothes in the sex scenes (Van Renswoude et al., 2019). Finally, as scenes showing bonobos engaged in play or a grooming bout did not hold attention longer than neutral scenes, the human variant of these scenes is likely also not very salient to bonobos.

In experiment 2 with human participants, we found an overall preference for viewing emotional scenes over neutral scenes, and with human emotional scenes receiving slightly more attention than bonobo scenes. Humans showed the most pronounced effect in the distress category, with longer looking durations towards distressed conspecifics compared to the neutral scenes. Moreover, humans also preferentially looked at individuals that were embracing each other, playing, or having sex. An implicit attentional bias for threatening signals has been studied a great deal in humans. Most studies indicate that in highly anxious individuals, attention to negative or threatening stimuli is strongly prioritized (Armstrong & Olatunji, 2012). Results on non-anxious individuals are mixed, showing that an implicit bias towards

positive emotional expressions also occurs (Becker et al., 2011). In a previous study, we found an implicit attentional bias towards stressful scenes in a heterogeneous human population, as well as to scenes involving sex and yawning (Kret & Van Berlo, 2021). Here, we add to the existing literature by showing that emotional scenes also spontaneously hold attention for longer durations in a task without a clear goal to the participants, and even when a competing social, but emotionally-neutral stimulus is present.

Interestingly, the attentional pattern of humans for human emotional scenes differed from that for bonobo emotional scenes. Humans looked longer at bonobos engaged in grooming and play compared to neutral scenes, but not at sex or yawning scenes, even though we found an effect for these two categories within the human scenes. Furthermore, we found weak evidence that humans looked longer at neutral scenes rather than bonobo distress scenes; the opposite from what we found for distress scenes of humans. In a previous study, adults rated distress scenes of bonobos as negative and highly arousing (similar to ratings of distress scenes of chimpanzees (Kret et al., 2018)), possibly due to canine visibility (Kret & Van Berlo, 2021). In our study, participants may have looked away from the distress scenes because they are intense in terms of emotional arousal, but this remains speculative. To date, very little work has examined how humans view (other) emotional expressions of primates (see e.g., Kret & Van Berlo (2021); Maréchal et al. (2017)). As such, future work on attentional biases could benefit from including questionnaires that measure participants' interpretation of and feelings towards the stimuli.

Compared to our bonobo sample, humans appear to preferentially sustain attention to emotional scenes across all categories. A possible explanation for this difference is that humans have evolved exceptionally distinctive and exaggerated communicative faces in order to communicate more effectively (Kret et al., 2020), and therefore also have a sensitivity to a wider range of expressions. Nevertheless, alternative explanations, particularly relating to our methodology, must be considered.

We report several limitations to our study. First, we used static images of emotional expressions instead of dynamic scenes. Studies have suggested that the dynamic facial expressions of emotion provide richer information than static expressions, causing stronger activation in brain regions associated with emotion recognition (Arsalidou et al., 2011). Second, we made use of more complex social and emotional scenes rather than isolated facial expressions, potentially providing more contextual information. However, it is possible that by providing this context, we increase the

complexity of the stimuli, therefore making the interpretation of the stimuli more ambiguous (Tottenham et al., 2013). A combination of these two interpretations may explain our bonobo results in that our stimuli may underrepresent the interest bonobos have in emotionally-salient information. Nevertheless, it is important to note that humans do show an emotional bias across all the categories of emotions even with a similarly prepared stimulus set. Moreover, in a follow-up experiment where we zoomed in on facial expressions rather than scenes (as well as investigating effects of expression channels such as face vs. body), an emotional bias was not observed (in prep, (Kim et al., 2021)). At the moment, it is difficult to know how bonobos interpret emotional images and whether emotional scenes are better at providing more salience than isolated faces. Future research could use dynamic emotional cues using videos or a combination of images with sound, as this has previously proved to be successful in uncovering an emotion bias in for instance chimpanzees (Kano & Tomonaga, 2010a).

Another limitation of our study is the small sample size. Moreover, we were only able to test female bonobos. The reason for this is that bonobos are rarely found in zoos and face a high risk of extinction (Fruth et al., 2016), and even fewer are accessible for research purposes. As such, we cannot extrapolate our findings to the entire species. Nevertheless, our results converge with a small, but growing body of experimental studies indicating that bonobos and other apes are sensitive to the emotional cues of others (Kano et al., 2016; Kret et al., 2016; Laméris et al., 2022; Pritsch et al., 2017; Van Berlo et al., 2020a), and showing that bonobos have remarkably well-developed brain structures that are important for emotion processing (Issa et al., 2019; Stimpson et al., 2016).

Perceiving emotions in others is at the foundation of more complex socio-cognitive abilities such as cooperation and empathy (Levine et al., 2018). Our findings show that bonobos, like humans, voluntarily look longer at emotionally salient signals such as distress and sex. Our findings converge with previous studies, suggesting that the groundwork for higher social cognition is likely shared with our closest living relatives.