Animal Behavior and Cognition
ABC 2026, 13(1):1-21
DOI: https://doi.org/10.26451/abc.13.01.01.2026

Do Great Apes Know Each Other's Names? Probing Great Ape Comprehension of Social Vocal Labels

Laura S. Lewis[1,2,*], Fumihiro Kano[3,4], Jeroen M. G. Stevens[5,6], Jamie G. DuBois[2,7], Josep Call[2], and Christopher Krupenye[2,8]

[1] *Departments of Psychological & Brain Sciences and Anthropology, University of California, Santa Barbara, CA, U.S.A*
[2] *School of Psychology & Neuroscience, University of St Andrews, St Andrews, U.K.*
[3] *Kumamoto Sanctuary, Wildlife Research Center, Kyoto University, Kumamoto, Japan*
[4] *Center for the Advanced Study of Collective Behavior, University of Konstanz, Konstanz, Germany*
[5] *SALTO, Agro and Biotechnology, Odisee University of Applied Sciences, Sint Niklaas, Belgium*
[6] *Centre for Research and Conservation, Royal Zoological Society of Antwerp, Antwerp, Belgium*
[7] *Department of Psychology, University of Cambridge, Cambridge, U.K.*
[8] *Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD, U.S.A.*

*Corresponding author (Email: laurasimonelewis@gmail.com)

Citation – Lewis, L. S., Kano, F., Stevens, J. M. G., DuBois, J. G., Call, J., & Krupenye, C. (2026). Do great apes know each other's names? Probing great ape comprehension of social vocal labels. *Animal Behavior and Cognition, 13*(1), 1-21. https://doi.org/10.26451/abc.13.01.01.2026

Abstract – Humans use proper names as vocal labels to identify and communicate with and about social agents. The comprehension of spoken proper names requires the ability to interpret socially specific verbal signals, or social vocal labels, and use cross-modal perception to identify and discriminate between group members. Individuals that recognize and comprehend familiar proper names can use these labels to identify and discriminate between groupmates, gain third-party knowledge, and guide decision-making. Use of vocal labels for conspecifics is noticeably rare in the animal kingdom, and has only been found in species (dolphins, elephants, and marmosets) that are phylogenetically distant from humans. We therefore investigated the phylogenetic trajectory of this capacity by studying our closest living primate relatives, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*). We implemented a cross-modal non-invasive eye-tracking and playback study with multiple populations of apes (N = 24) living in zoos and sanctuaries, none of whom were specifically language-trained. We tested whether chimpanzees and bonobos spontaneously attend toward an image of a groupmate whose name has been called by a human caretaker. We found limited evidence that apes link the caretaker-given names of their groupmates to images on a screen, and therefore cannot make strong conclusions about apes' comprehension of these social vocal labels. Our playback and eye-tracking paradigm offers a novel tool for studying cross-modal perception and knowledge of vocal labels. Future work will be critical to identify the sociocognitive foundations underlying socially specific referential communication and the evolution of language.

Keywords – Language comprehension, Vocal labels, Social knowledge, Eye tracking, Primates, Cognitive evolution

The capacity to identify and evaluate individuals, especially based on their interactions with others, is advantageous for species that navigate large and complex social environments. This ability to assess others' actions and intentions based on their social interactions, or third-party knowledge, develops early in human ontogeny and was likely instrumental in human evolutionary history (Hamlin et al., 2007). Human fetuses 33 - 41 weeks of gestational age can discriminate between their mother's voices and others' voices, and between their mother's native language and a foreign language (Kisilevsky et al., 2009). By three months, human infants begin to differentiate novel actors based on third party interactions, and use these

interactions to generate third-party knowledge (Hamlin et al., 2007; Kinzler et al., 2007; Mascaro & Csibra, 2012; Thomsen et al., 2011). Third-party knowledge is foundational for humans' cooperative behavior, as it allows us to distinguish cooperative and norm-adherent individuals who benefit the group from cheaters and norm-violators who may not (Axelrod & Hamilton, 1981). By observing third-party interactions, individuals can gain third-party knowledge and utilize it to identify the most dominant individuals and cooperative allies and infer one's own social relationships with others. Third-party knowledge can also inform decisions around social interactions that maximize benefits, like strengthening social bonds with those who cooperate with others, and minimize risks, such as costly contest aggression with more dominant individuals. Critically, third-party knowledge can be obtained from many social contexts, including others' communicative interactions (e.g., Cheney & Seyfarth, 1980) and recruitment attempts (de Waal, 1982).

One fundamental prerequisite to gaining third-party knowledge is an understanding of the communicative and referential signals involved in a social interaction. Humans use proper names, or nouns that consistently refer to specific individuals, as vocal labels to initiate and coordinate social interactions (Kripke, 1980; Sorrentino, 2001). Labeling and naming other social agents is a defining aspect of human language and is an important building block for our complex communication systems and social interactions (Hurford & Hurford, 2007; Seyfarth et al., 2005). Children two years of age can already comprehend that proper names refer to unique individuals, and are sensitive to a speaker's knowledge when learning proper names (Kripke, 1980; Sorrentino, 2001). A key question is how humans have evolved the cognitive capacities that allow for comprehension of "social vocal labels," which we define here as a signal that refers to a specific social agent, such as proper names. One unique way to probe the evolutionary pressures that may have shaped the use of social vocal labels in humans is to investigate these capacities in our closest living phylogenetic relatives, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*). Exploring whether these species can comprehend social vocal labels can provide insight into whether these capacities are shared with our great ape cousins and were likely present millions of years ago in our common evolutionary ancestors, or whether humans may have uniquely evolved the ability to use and comprehend these types of labels. While bonobos and chimpanzees do exhibit complex multimodal communication, including combinations of referential gestures and vocalizations (Doherty et al., 2023; Genty & Zuberbühler, 2014; Graham et al., 2018, 2022), it is currently unknown whether these species can comprehend or use social vocal labels.

However, there is some evidence that chimpanzees and bonobos can comprehend "functionally referential signals," or signals that are both conceptually and perceptually specific (Cäsar & Zuberbühler, 2012; Macedonia & Evans, 2010). For example, bonobo Kanzi, who was raised in a language-enriched environment and heavily trained to use lexigrams, was reported to spontaneously acquire new symbols with which he had no training and use them in communicative contexts (Rabinowitz, 2016; Savage-Rumbaugh et al., 1986, 1993). It has also been said that Kanzi could recognize over 100 spoken words in multiple communicative contexts, including spoken proper names (Rabinowitz, 2016; Savage-Rumbaugh et al., 1986, 1993). Lastly, a chimpanzee demonstrated a negatively shifted event-related potential (ERP) in response to the sound of her own caretaker-given name being played from a speaker (Hirata et al., 2011; Ueno et al., 2009). However, it is currently unknown whether, in the absence of extensive training, great apes can spontaneously acquire and recognize spoken proper names as referring to other specific, unique individuals.

Many other primate and bird species use functionally referential signals in both predation and feeding contexts (Cäsar & Zuberbühler, 2012; Macedonia & Evans, 2010). For example, vervet monkeys (*Chlorocebus pygerythrus*) produce and comprehend distinct alarm calls that refer to different predators (Cheney & Seyfarth, 1980; Seyfarth et al., 1980). However, only a few nonhuman animal species use referential signals in social contexts (Faragó et al., 2010; Gouzoules et al., 1984; Rendall et al., 1999). Faragó and colleagues (2010) found that in domestic dogs (*Canis familiaris*), growls expressed during play, food guarding, and threat response all differ acoustically and produce different, but context-appropriate, responses in the listener. In rhesus macaques (*Macaca mulatta*), scream vocalizations during agonistic encounters vary according to the particular class of opponent and level of physical aggression experienced by the caller (Gouzoules et al., 1984). Baboons (*Papio ursinus*) also produce two harmonically distinct

types of grunts, one to initiate movement and the other when approaching infants, that elicit different and appropriate responses in the listeners (Rendall et al., 1999).

However, only three nonhuman animal species have been found to use social vocal labels. These studies typically employ playback methods that measure turning toward and looking at speakers that produce distinct vocalizations from specific individuals. Bottlenose dolphins (*Tursiops truncatus*) link vocal and visual social information through the use of "signature whistles," which are distinct, individual vocalizations that are used to broadcast the identity of the caller (Janik & Sayigh, 2013; Quick & Janik, 2012). Common marmosets (*Callithrix jacchus*) produce "phee" calls as vocal labels for conspecifics, comprehend when these labels are directed at them, and respond correctly to the distinct phee calls that are directed at them (Oren et al., 2024; however, see Jaakkola, 2025). Female African elephants (*Loxodonta africana*) exhibit calls that are vocally distinct at the individual and group level, and social group membership is a better predictor of call similarity as compared to genetic relatedness, although it is still unknown whether other elephants comprehend these calls as specific vocal labels (Pardo et al., 2024).

Given the phylogenetic distance between these various species, it is difficult to determine whether the capacity to use and comprehend socially specific referential signals is homologous or has evolved independently in these differing lineages. Since bonobo Kanzi appeared to recognize spoken proper names, one question is whether his rich symbolic training and linguistic experience are necessary for this capacity to emerge, and whether it is widespread and occurs spontaneously in other non-language-trained apes. Critically, this leaves open the question of whether untrained nonhuman animals can produce and exploit vocal labels to predict and eavesdrop on third-party interactions. In these previous studies, the focus was on whether nonhuman animals could both generate and understand name-like vocal labels, potentially as referential symbols. In the present study, our focus was solely on comprehension of such signals, in the sense of associating a vocal label produced by a third party with another individual. Such an association could arise if the label is understood as a symbol, or simply if animals learn that it typically precedes some interaction with the target individual, either of which could help animals to effectively eavesdrop on third-party interactions.

One method to investigate how nonhuman animals may integrate visual and auditory information is by exploring cross-modal perception. Cross-modal perception is defined as the ability to integrate information between different sensory modalities (Davenport et al., 1973). Auditory-visual recognition describes the cross-modal capacity to integrate auditory and visual information, and has been found in preverbal infants as well as a number of nonhuman primates. The canonical test for the ability to cross-modally integrate visual and auditory information is to present an individual with two images or videos on a screen, and then to simultaneously present auditory information that is consistent with one *but not both* of the visual stimuli through speakers located near the screen. These stimuli often include visual facial cues of conspecifics, and corresponding auditory speech or vocalizations. Participants are thought to accurately integrate visual and auditory information if they look at the image or video that is consistent with the auditory information. This paradigm has been utilized to successfully demonstrate that infants before 6 months of age exhibit some forms of auditory-visual cross-modal perception (Kuhl & Meltzoff, 1982, 1996; Lewkowicz & Turkewitz, 1980). This method has also revealed that other animal species such as horses and dogs can cross-modally represent other individuals (Adachi et al., 2007; Adachi & Fujita, 2007; Nakamura et al., 2018; Proops et al., 2009) In addition, lemurs, monkeys, and apes have also demonstrated some forms of cross-modal perception through the use of other methodologies such as match-to-sample and olfactory tasks (Adachi & Fujita, 2007; Carvajal & Krupenye, 2025; Davenport et al., 1973; Evans et al., 2005; Ghazanfar & Logothetis, 2003; Hashiya & Kojima, 2001; Kulahci et al., 2014; Tomasello & Call, 1997). Finally, Sliwa and colleagues (2011) used an eye-tracking and preferential-looking task with rhesus macaques and found that they spontaneously matched the faces of familiar humans and conspecifics with their respective voices. The methods and results of this study served as a key influence for the implementation of our own task with chimpanzees and bonobos.

Thus, in the present study, we investigated whether non-language-trained chimpanzees and bonobos could cross-modally integrate auditory and visual perception to comprehend a key feature of human communication: social vocal labels. We designed a task similar to previous methods that

investigated cross-modal perception in human infants and other species: we presented apes with two side-by-side same-sex images of current groupmates, while noninvasively recording their gaze with an eye-tracker. After presenting the images for 1 s to capture baseline patterns of attention, we then played a recording of a familiar caretaker calling one of the groupmates' names (given to them by caretaker) while continuing to present the two images on the screen. We used eye tracking to map apes' gaze across the images as they heard the name of one groupmate being called (similar to previous research, which typically recorded looking times by hand, frame-by-frame). Based on previous findings with human infants and other nonhuman animals, we predicted that if apes can recognize each other's spoken proper names, they should spontaneously match the face of a groupmate with their spoken caretaker-given name, as indicated by increased attention towards the image of the groupmate whose name was called as compared to the image of the groupmate whose name was not called.

Methods

Ethics Statement

Experimental protocols adhered to the School of Psychology and Neuroscience Animal Ethics Committee at the University of St Andrews and to approval by each participating animal care institution. Edinburgh and Kumamoto Sanctuary participants were tested in the testing rooms prepared for each species, whereas the Planckendael participants were tested in their large indoor enclosure. Apes' daily participation in this study was completely voluntary. They received regular feedings and daily enrichment and had ad libitum access to water. Animal husbandry and research protocols complied with international standards (the Weatherall report, The use of nonhuman primates in research) and institutional guidelines (Kumamoto Sanctuary: Wildlife Research Center Guide for the animal research ethics; Edinburgh and Planckendael Zoos: EAZA Minimum standards for the accommodation and care of animals in zoos and aquaria; WAZA Ethical guidelines for the conduct of research on animals by zoos and aquariums; Guidelines for the treatment of animals in behavioral research and teaching (ASAB/ABS)).

Participants

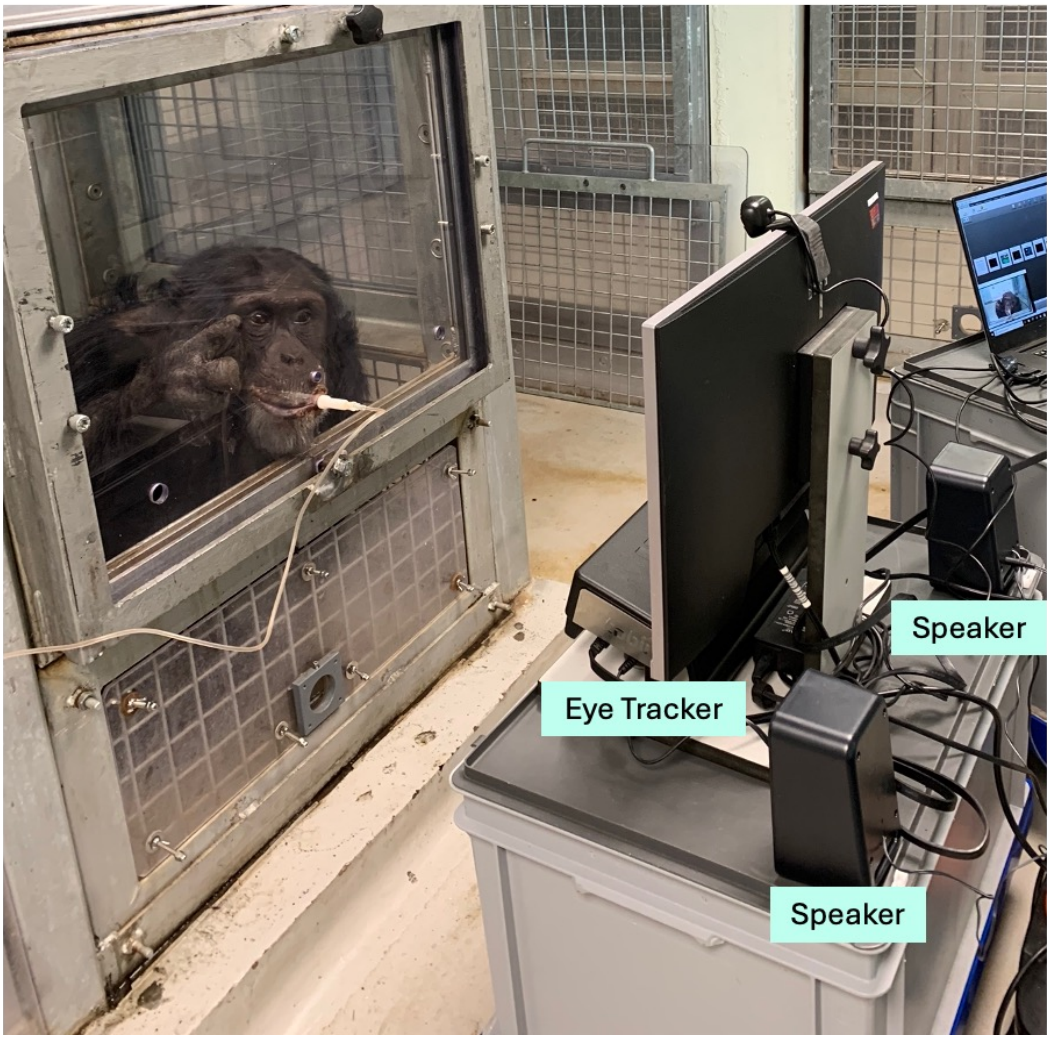
We tested 24 ape participants derived from two groups per species, living in three locations: Edinburgh Zoo, Scotland (7 chimpanzees: 2 females, 5 males), Kumamoto Sanctuary, Japan (6 chimpanzees: 5 females, 1 male; 6 bonobos: 4 females, 2 males), and Planckendael Zoo, Belgium (5 bonobos: 2 females, 3 males). Apes participated in this study from March 2019 - August 2019 and ranged in age from 2 to 46 years (bonobo mean = 21.19 ± 13.95 years (SD); chimpanzee mean = 21.43 ± 9.72 years (SD) (see Tables S1 and S2 for more details).

Apparatus

We used previously established eye-tracking procedures and comparable setups across facilities (Hopper et al., 2021; Kano & Tomonaga, 2009; Krupenye et al., 2016; Lewis et al., 2021; Lewis & Krupenye, 2022). Images were presented to apes through a transparent polycarbonate or acrylic panel on a 23" LCD monitor situated directly outside of their enclosures at a distance of approximately 60 cm from the subject's face. Subjects' eye movements were non-invasively recorded via an infrared eye-tracker (X120 in Edinburgh and Planckendael, X300 in Kumamoto, Tobii Technology AB, Stockholm, Sweden), positioned directly below the monitor, which mapped their gaze onto the stimulus images. Stimulus presentation and data collection were controlled using Tobii Studio. Apes were continuously provided a small amount of diluted fruit juice (provided irrespective of viewing patterns) which was delivered through a plastic nozzle positioned on the transparent panel, directly in front of the eye-tracker. Providing juice in this way encourages apes to voluntarily position themselves at the eye-tracking setup, minimizes head movements, and optimizes corneal reflection measurements (see Figure 1).

Figure 1

Experimental Eye-Tracking Set Up at Edinburgh Zoo, Scotland



Note. The experimental eye tracking set up at Edinburgh Zoo, including (from left) a chimpanzee participant watching images presented on the monitor screen, below which is the Tobii X120 eye tracker. Two speakers are placed directly behind the monitor, and the experimenter laptop is placed to the right of the monitor.

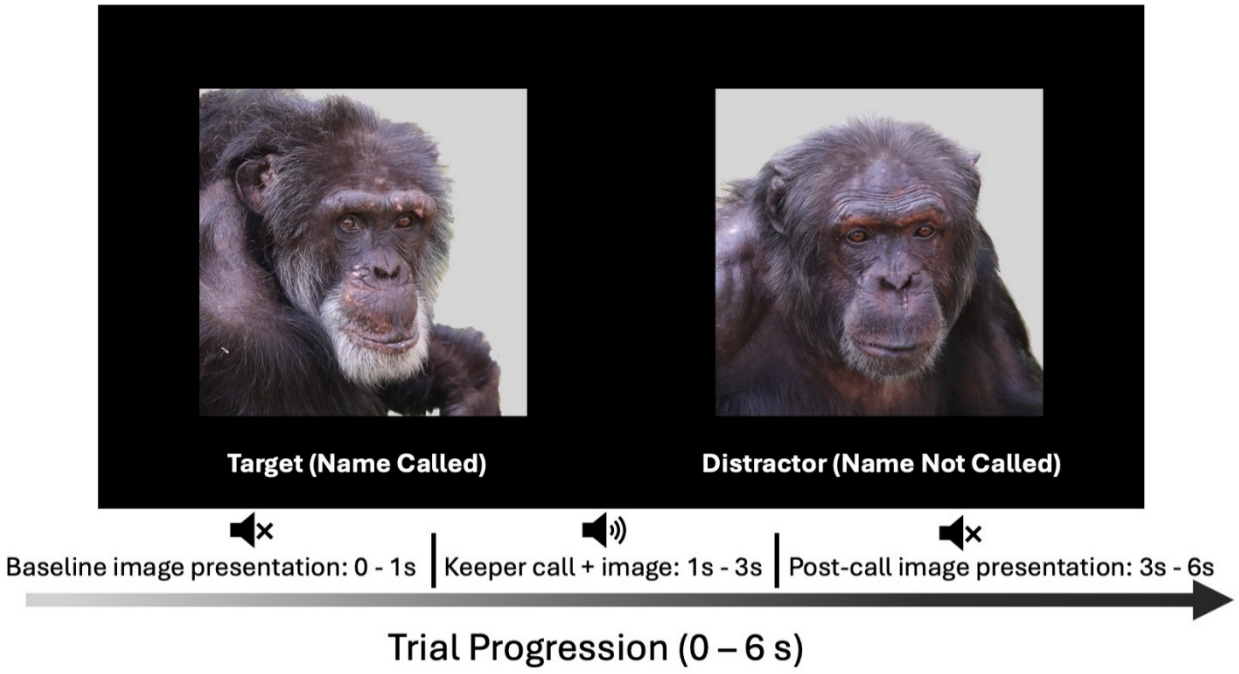
Before testing, we conducted a two-point calibration for each ape subject by presenting a small video clip (and often a small piece of real fruit) on each reference point. This two-point calibration procedure is routinely used in eye-tracking studies with great apes to provide high quality data and minimize the potential loss of subjects who would not reliably attend to a larger number of calibration points (Hopper et al., 2021; Kano & Call, 2014; Kawaguchi et al., 2019). After each calibration was obtained, we manually checked the accuracy of the calibration using nine points on the screen and repeated the calibration process if the original calibration was suboptimal. Each subject's unique calibration was used throughout the entire testing period. Prior to the start of every test session, the calibration accuracy was visually checked with at least one of the nine points. Calibration errors are typically less than a degree with this procedure, and any error of this size will not impact the ability to determine preferential looking to images (Kano et al., 2011).

Stimuli

The visual stimuli in this study consisted of static 600 x 600 pixel close-up color photographs of forward-facing conspecific faces with neutral facial expressions (hereafter referred to as "avatars") and a gray background (Figure 2). Each trial featured two images of two different groupmates of the same sex, presented on the center left and center right regions of a black 1920 x 1080 pixel screen (locations were counterbalanced across trials). The distance between the right-most edge of the left avatar and the leftmost edge of the right avatar was 247 pixels, and the brightness, contrast, and blurriness of photographs were kept as consistent as possible across stimuli. While the images were presented on the screen, we played a recording of a familiar caretaker calling the name of one of the groupmates presented in the images on the screen (more detailed procedural information provided below). Hereafter, we refer to the groupmate whose name was called as the "target avatar", and the groupmate whose name was not called as the "distractor avatar".

Figure 2

Example Trial Image and Trial Sequence Progression



Chimpanzees at Edinburgh Zoo and bonobos at Planckendael Zoo experienced two stimulus sets, one of female avatars and the other male avatars, with each set comprised of three images of adult female groupmates and three images of adult male groupmates, respectively (a single image of each groupmate was used). As the Kumamoto bonobo population only had two male groupmates, the male stimuli set included only pictures of these two males (the female stimulus set included images of three female groupmates). The Kumamoto chimpanzee population only has one male groupmate, and thus these individuals only saw a stimulus set of three images of female groupmates.

Within a stimulus set, each groupmate avatar was paired with every other same-sex avatar in the set (three groupmates = three pairs: AB, AC, BC). To control for both the location and identity of the target avatar, each stimulus pair was shown four times: once each with (1) individual 1 as the target avatar (whose name was called) on the left, (2) individual 1 as the target avatar on the right, (3) individual 2 as the target avatar on the left, and (4) individual 2 as the target avatar on the right. Thus, in total, the Edinburgh

chimpanzees and Planckendael bonobos each saw 24 trials ([3 female pairs + 3 male pairs] x 4 presentations). The Kumamoto bonobos saw 16 trials ([3 female pairs + 1 male pair] x 4 presentations), and the Kumamoto chimpanzees saw 12 trials total (3 female pairs x 4 presentations). Within each group, the majority of individuals received identical stimuli. If, however, a participant was included in the standard stimulus set for their group, for their stimulus set, their own image was replaced with that of a different member of their group. In cases where no additional individuals of the same sex existed within the group, these individuals received fewer trials than others in their group.

Procedure

At all three facilities, apes voluntarily entered the testing room. At Kumamoto Sanctuary and Planckendael Zoo, apes were sometimes temporarily separated from groupmates for testing. At Edinburgh Zoo, apes were not separated but tests were only administered when other groupmates were at least 1 m away from the subject such that interference was unlikely. All of the apes included in this study had already participated in other eye-tracking studies, and thus did not require habituation as they were already familiar with the experimental set-up.

The test trials were administered in clusters of two. The side (left, right) on which the target avatar was presented was counterbalanced within and across clusters. Chimpanzees at Edinburgh Zoo and bonobos at Planckendael Zoo saw clusters that had one male trial and one female trial, and half of the clusters presented a female trial first while the other half presented a male trial first. The Kumamoto bonobos saw four clusters that had one male trial and one female trial, and four clusters that had two female trials (because of a lower number of male trials). For the Kumamoto bonobos, the presentation alternated between clusters with two female trials and clusters with one male and one female trial. The Kumamoto chimpanzees only saw trials with female avatars, and within these clusters the trials alternated which female avatar was on the right side vs. left side along with which name was called.

We used Adobe Premiere Pro (Adobe Systems, San Jose, CA) to produce the trial videos that included images and recordings, which automatically progressed after the experimenter pressed the start key. In each trial, the images were presented on the screen after a 0.5 s presentation of a black screen with fixation cross in the center (intended to center apes' gaze before trial onset). After the avatar images were presented on the monitor screen for 1 s, the recording of a familiar caretaker calling the name of one of the groupmates played for a duration of 2 s (the names were called twice while the images on the screen were continuously presented). The trial images remained on the screen for 3 s after the end of the recording, for a total of 6.5 s per trial (Figure 2). Within a cluster, trials progressed one immediately following the other for a total duration of 13 s per two-trial cluster (including the two 0.5 s fixation crosses at the start of each 6 s trial). Groupmates' names were called by caretakers who were most familiar with the apes (at least five years of experience working with the group). The Edinburgh chimpanzee and Kumamoto Sanctuary bonobo populations heard the voice of a familiar female caretaker, while the Planckendael bonobo and Kumamoto Sanctuary chimpanzee populations heard the voice of a familiar male caretaker. To replicate a natural sounding call, we instructed the caretakers to call the ape names as they would when calling apes to move from one enclosure to the next. Recordings were captured in quiet settings free from any background noise or disruption. The recordings were played through two speakers that were placed directly behind the monitor on which the participants saw the trial images (one speaker on either side of the monitor, see Figure 1). The speakers were connected to the experimenter laptop via USB, and the volume was set on the speakers so that it was loud enough for the participants to hear the sounds being played, but not too loud so that participants were not startled (the same speaker volume was used throughout the duration of the experiment). All participants were familiar with sounds being played from speakers through previous research experience. Because participation was voluntary (i.e., apes could walk away from the experimental set-up at any time), the number of clusters administered within a day varied between one and six, depending on the duration of apes' attendance and attention at the testing set-up. After administering all trials in the predetermined order, we verified that there was at least one fixation toward either the target or distractor image (see Data Scoring and Analysis below) for each trial. After subjects completed their originally

assigned trial order, trials that yielded zero fixations to either image were repeated until we had data for a full set of trials for each subject. In total, we tested 420 trials (20 missing trials due to persistent lack of interest from three individuals).

Data Scoring and Analysis

Areas of interest (AOIs) were defined in Tobii Studio as 700 x 700 pixel areas around each of the two images in each trial; thus each AOI included the image plus a 50 pixel buffer on each side of the image. Fixations were calculated using Tobii Studio's I-VT Filter, and then fixation data were exported frame-by-frame from Tobii Studio into TSV files. We summed total fixation duration within each AOI across the total trial, and for each second of the trial (i.e., for seconds 0 – 1, 1 – 2, 2 – 3, etc.) for a total of 6 binned fixation durations. To measure biases in looking towards the target avatar versus the distractor avatar, we next calculated both raw difference scores (i.e., sum of fixations to target avatar minus sum of fixations to distractor avatar) and a proportional Differential Looking Score (DLS):

$$\frac{\text{Target minus Distractor fixation time}}{\text{Sum of fixation time to Target} + \text{Distractor}}$$

Both raw and proportional scores were used as dependent variables for each second within each trial. We additionally calculated DLS for the overall baseline (0-1s) and test (1-6 s) portions of the trial but did not do this for raw difference scores because it would be inappropriate to compare raw differences scores across periods of differing duration. However, where both measures are valid, we used both as they capture different information. Raw difference scores provide a direct measure of the difference in looking time to the target avatar relative to the distractor avatar and also captures variation in overall looking duration. However, to control for variation in overall looking time (see Lonsdorf et al., 2019), which may differ across individuals and wane throughout a trial, we also used the DLS. This proportional score, in contrast, amplifies strongly biased looks even during periods when overall looking duration is low. The raw difference scores and DLS variables range from (-3 to 3) and (-1 to 1), respectively.

We anticipated that recognition of others' names could manifest in two possible looking bias patterns: Apes might show sustained biases toward the target avatar throughout the test window (i.e., the 5 s from the time the recording started until 3 s after the recording ended) or they might show looking biases only during specific periods of the test window, such as during or immediately following utterance of the groupmate's name. Looking time to the target individual was chosen as a dependent variable because it is the most widely used spontaneous measure of cross-modal recognition in animals (Sliwa et al., 2011). Thus, we tested both predictions in separate models. In **Model 1**, we examined attentional biases in the baseline window (the 1 second period immediately before the recording was played) and the subsequent 5 s test window. In exploratory **Model Set 2**, we examined attentional biases across the test window's five 1 s time bins (each time bin after the start of the call). In exploratory **Model 3,** we tested attentional biases between trial phases: the baseline phase (0 – 1 s), the call phase (1 – 3s), and the post-call phase (3 – 6 s). This set of models was designed to first measure whether there were any broad differences in looking times between the baseline window before any auditory information was presented and the test window when participants were presented with the auditory information. Next, the exploratory models were designed to provide more fine-grained detail about the timing of apes' gaze patterns before, during, and after the presentation of the auditory stimuli. As this combined eye tracking and playback paradigm is the first of its kind to be used to explore cross-modal perception in great apes, these exploratory models were designed to capture potentially unknown details of apes' gaze patterns and any changes over time as they simultaneously viewed images and heard corresponding auditory information. The design of these analyses was partially inspired by previous studies, such as Sliwa et al. (2011), which analyzed overall looking times and also conducted a binned time series analysis.

General Modeling Approach

We fitted linear mixed effects models for DLS for **Model 1** and **Model 3**, and for both dependent variables (i.e., raw difference looking time and DLS) for exploratory **Model Set 2** using the statistical software R (version 4.0.2; R Core Team 2020). We only fit DLS for **Model 1** and **Model 3** because the baseline and test windows were different lengths (1 s and 5 s, respectively), and thus it was not possible to fit this model with raw difference scores. We fitted simple linear mixed models using the *lmer* function from the 'lme4' package (Bates et al., 2007). We used the original intervals for both the raw difference score [-3,3] and DLS [-1,1], so that it was possible to determine if these scores were significantly different from zero (a model intercept of zero signifies no bias toward the target or distractor avatar). The value of the model intercept reflects the values of the reference category. Since these scales are centered around zero, a significant model intercept term indicates that the scores from the reference category differ significantly from zero.

For **Models 1 and 3**, we used a significance threshold of .050 when reporting p-values, and report trends between .050 and .100, given that such effects may still be biologically meaningful (Stoehr, 1999). Exploratory **Model Set 2** used the same significance threshold, except for interpretation of the intercept term. As we used the intercept term to conduct separate significance tests for each of the five test window time bins after the start of the call, we therefore applied a Bonferroni correction to the alpha level such that we used a significance threshold of .010 when reporting p-values (Bonferroni correction: $p = .050/(5$ separate tests)), and here we report trends as between $p = .010$ and .020. We first used likelihood ratio tests to compare the fit of the full model against the null model, which included only the random effects and control predictor (see Tables 1 - 3 for full model sets and comparisons). We ran the *vif* function before running each model to determine whether any model effects had collinearity, and this function indicated that none of the models' effects were collinear. Before observing the results of the models, we visually inspected plots of the models' residual values against fitted values and *q-q* plots to confirm that each model met the assumptions of normally distributed and homogeneous residuals. Next, we used the Anova function with type III sum of squares in the 'car' package to obtain *p*-values for individual terms within these models (Fox et al., 2012). We inspected all models to ensure conformity of assumptions of normality and homogeneous distribution of residuals, and the absence of collinear predictors.

Model 1: Attentional Biases in Baseline Versus Total Test Window

Model 1 tested the prediction that, after hearing the recording, apes will exhibit a *sustained* increase in attention towards the image of the groupmate whose name was called as compared to the image of the groupmate whose name was not called. Each trial was represented by two rows of data - one for the baseline window and another for the test window. We included the categorical test predictor of baseline vs. test window to determine whether subjects showed stronger biases in the test window than in the baseline. We included z-transformed trial number as a single continuous control predictor to account for any potential effects of habituation across trials. As random intercepts, we included subject identity (to account for repeated measures from each subject), trial ID (unique name for each trial to account for baseline and test biases within individual trials), and stimulus dyad (i.e., the IDs of the groupmates in the images, combined as a single dyad measure to account for potential random variability in preferences for specific individuals). We included z-transformed trial number as a random slope for the subject and avatar dyad random intercepts. In the model output, the intercept term will be significant if, within the reference category, the population average of the dependent measure differs significantly from zero. Because our dependent measure was centered around zero, we used the intercept to assess whether apes showed a significant bias in attention, with positive estimates reflecting a bias toward the target and negative estimates reflecting a bias toward the distractor. By adjusting the reference category between baseline and test windows, we were able to assess biases within each window separately.

Exploratory Model 2: Comparison of Attentional Biases Across Each One-Second Time Bin

Exploratory **Model Set 2** tested the prediction that apes will show more fleeting biases only in specific periods of the test window. Each trial was represented by six rows of data - one for each 1-second bin (the first being the baseline). We planned this more detailed analysis to provide a test of attentional biases at a smaller timescale within the trial, along with the general analysis of the trial as a whole sequence in **Model 1**. This model was identical to **Model 1** except that we included time bin as the single categorical predictor rather than the baseline vs. test factor, and we included trial number as a random slope for the avatar dyad random intercept. As in Model 1, we adjusted the reference category and used the intercept term to identify significant biases in attention within individual time bins (we used baseline as the default reference category to identify significant differences relative to baseline).

Exploratory Model 3: Comparison of Attentional Biases Between Species and Populations

Exploratory **Model 3** tested whether there are differences in attentional biases during the baseline and test phase between species and/or populations, given that we have previously found population differences between Kumamoto Sanctuary apes and Edinburgh/Planckendael apes in eye-tracking studies (Lewis et al., 2021; Lewis et al., 2023). We ran this model only with DLS, as DLS could control for the differences in trial phase lengths between the baseline and test window. We included a three-way interaction between the test predictor of baseline vs. test window, population (Kumamoto Sanctuary apes vs. Edinburgh and Planckendael apes), and species to determine whether specific populations or species showed stronger biases in the test window than in the baseline. Z-transformed trial number was included as a single continuous control predictor to account for any potential effects of habituation across trials. As random intercepts, we included subject identity, trial ID, and stimulus dyad. We included z-transformed trial number as a random slope for the subject and avatar dyad random intercepts. Because our dependent measure was centered around zero, we again used the intercept to assess whether apes showed a significant bias in attention, with positive estimates reflecting a bias toward the target and negative estimates reflecting a bias toward the distractor. By adjusting the reference category between baseline and test windows, we were able to assess biases within each window separately.

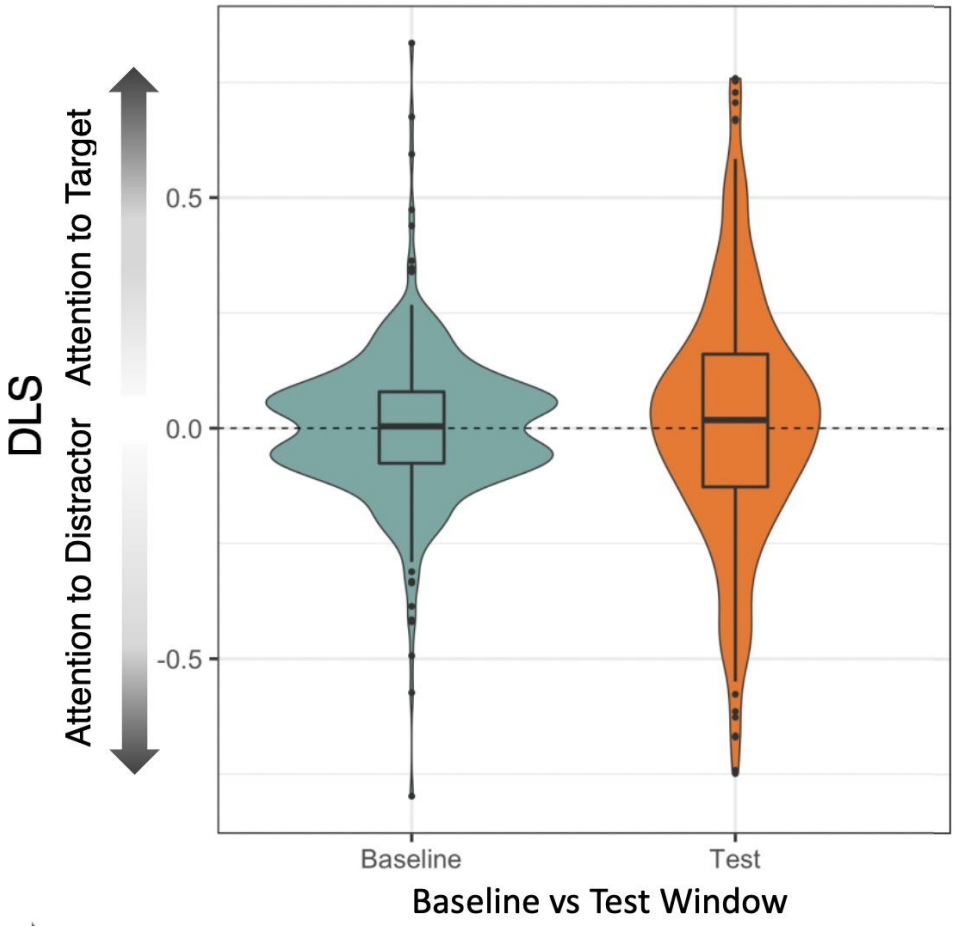
Results

Model 1: Attentional Biases in Baseline Versus Test Windows

The full-null model comparison for Model 1 was not significant (DLS, $\chi 2 = 0.579$, $p = .447$). However, we explored the contribution of the individual test predictors as their simultaneous exclusion from the null model in the full-null model comparison masks their individual effects (Aberson, 2002). Although apes exhibited slightly stronger biases during the test window (estimate = 0.032 ± 0.678 (SE)) than during the baseline (estimate = 0.002 ± 0.151 (SE)), this effect was not significant. When the reference category was adjusted to individually test the intercept of the baseline window and test window, attentional biases did not differ significantly from chance for either window (Table 1, Figure 3).

Figure 1

Biases in Attention Toward Target Versus Distractor Avatar, in the Baseline and Test Windows



Note. Baseline window captures patterns of attention in the first second of the trial, before the recording of the caretaker calling the target avatar's name is played. The test window captures patterns of attention for the full five seconds after the start of the recording until the end of the trial. Black dashed line denotes chance (equal looking to target and distractor avatars). Boxes denote the interquartile range (IQR, from 25th percentile to 75th percentile), and middle lines denote medians.

Table 1

DLS Model 1 Output

Factor	Estimate	S.E.	$\chi 2$	Df	*p*-value
Baseline vs. Test	-0.029	0.033	0.789	1	.374
Trial Number	-0.012	0.021	0.344	1	.558

Note. Predictors of biases in attention toward target and distractor conspecific faces (with test window as the reference category). DLS was used as the dependent measure. Subject, avatar dyad, and trial ID were included as random intercepts. Trial number was included as a random slope for the subject and avatar dyad random intercepts. Estimates and SEs are taken from the model summary; $\chi 2$ values, degrees of freedom and *p* values are taken from the Anova (type III sum of squares) output.

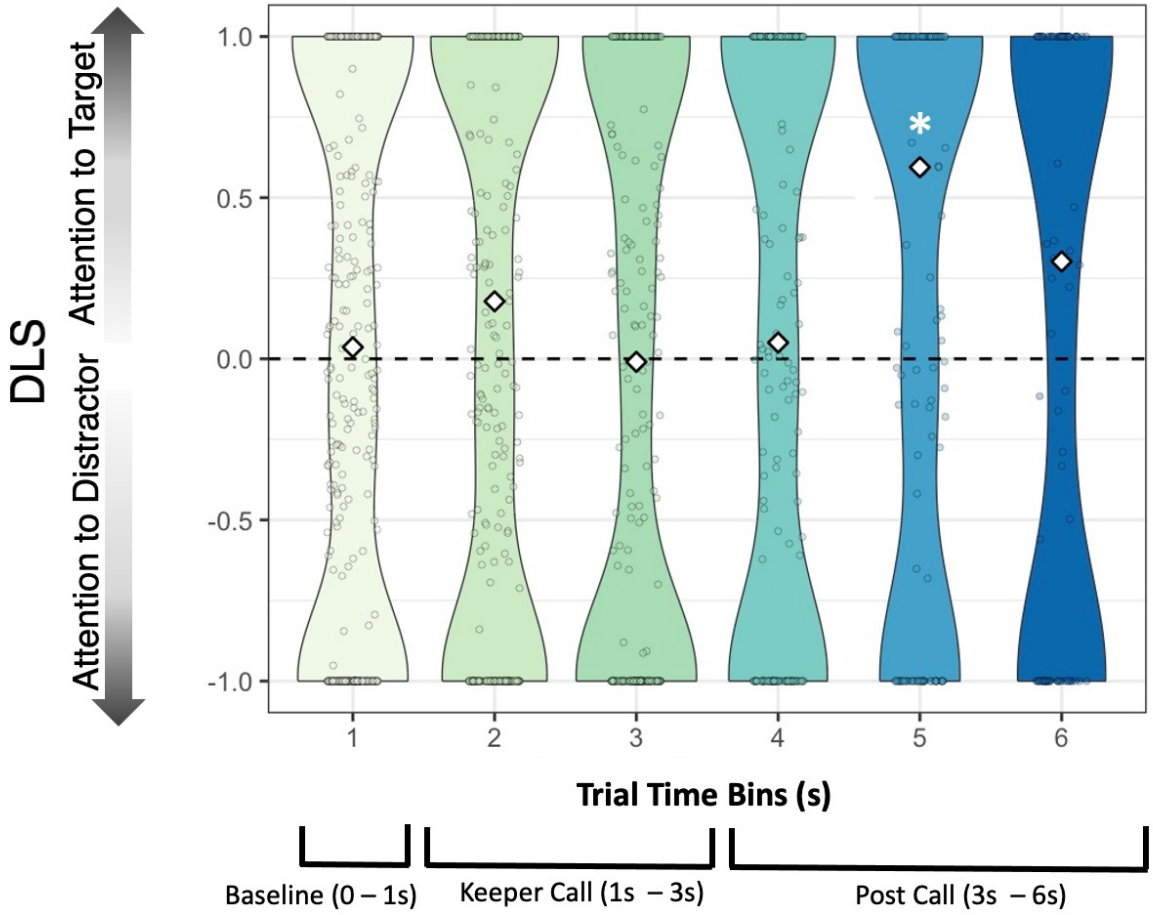
Exploratory Model 2: Attentional Biases Across One-Second Time Bins

The full-null model comparison for exploratory Model 2 was not significant for either the DLS ($\chi 2 = 6.277$, $p = .280$) or difference scores ($\chi 2 = 8.229$, $p = .144$). However, we again explored the contribution

of the individual test predictors as their individual effects may have been masked by their simultaneous exclusion from the null model in the full-null model comparison (Aberson, 2002). Examination of the model intercept with each time bin as a reference category indicated that subjects showed a significant bias toward the target avatar in time bin 5 in the raw differences scores model (estimate = 0.097, $\chi 2 = 7.245$, $p = .007$), and a marginally significant bias in time bin 5 in the DLS model (estimate = 0.178, $\chi 2 = 5.914$, $p = .015$) (see Tables 2 and 3 and Figure 4). Thus, during time bin 5, which occurs in the middle of the post-call window, apes spent approximately 18% more time looking towards the target avatar as compared to the distractor avatar.

Figure 4

Biases in Attention Toward Target vs. Distractor Avatar, in Each 1 s Time Bin



Note. Baseline window captures patterns of attention in the first second of the trial, before the recording of the caretaker calling the target avatar's name is played. Black dashed line denotes chance (equal looking to target and distractor avatars). Diamonds denote medians for each time bin. *$p < .020$ (alpha Bonferroni corrected for the five trial time bins).

Table 2

DLS Model 2 Output

Factor	Estimate	S.E.	$\chi 2$	df	*p*-value
TimeBin(1)	.033	0.033	0.596	1	.440
TimeBin(2)	0.075	0.047	2.548	1	.110
TimeBin(3)	-0.027	0.051	0.278	1	.598
TimeBin(4)	0.058	0.060	0.908	1	.341
TimeBin(5)	0.178	0.073	5.914	1	.015
TimeBin(6)	0.069	0.090	0.600	1	.439
Trial Number	-0.026	-0.026	0.992	1	.319

Note. Predictors of biases in attention toward target and distractor conspecific faces, with each time bin as a reference category. DLS was used as the dependent measure. Subject, avatar dyad, and trial ID were included as random intercepts. Trial number was included as a random slope for the avatar dyad random intercept. Each TimeBin(x) line in the table represents the results of the model intercept with the respective time bin as the reference category. P values between .01 and .02 are underlined. Estimates and SEs are taken from the model summary; $\chi 2$ values, degrees of freedom and P values are taken from the Anova (type III sum of squares) output.

Table 3

Raw Difference Model 2 Output

Factor	Estimate	S.E.	$\chi 2$	df	*p*-value
TimeBin(1)	0.004	0.856	0.033	1	.855
TimeBin(2)	0.030	0.203	1.650	1	.198
TimeBin(3)	-0.008	0.026	0.119	1	.730
TimeBin(4)	0.035	0.029	1.349	1	.245
TimeBin(5)	0.097	0.036	7.245	1	**.007**
TimeBin(6)	-0.009	0.044	0.013	1	.911
Trial Number	-0.012	0.014	0.839	1	.359

Note. Predictors of biases in attention toward target and distractor conspecific faces, with each time bin as a reference category. Raw difference was used as the dependent measure. Subject, avatar dyad, and trial ID were included as random intercepts. Trial number was included as a random slope for the avatar dyad random intercept. Each TimeBin(x) line in the table represents the results of the model intercept with the respective time bin as the reference category. Bonferroni corrected *p* values less than .010 are bolded. Estimates and SEs are taken from the model summary; $\chi 2$ values, degrees of freedom and *p* values are taken from the Anova (type III sum of squares) output.

Exploratory Model 3: Species and Population Differences

The full-null model comparison for exploratory Model 3 was not significant ($\chi 2 = 6.277$, $p = .967$). However, we again explored the contribution of the individual test predictors as their individual effects may have been masked by their simultaneous exclusion from the null model in the full-null model comparison (Aberson, 2002). The three-way interaction between the test predictor of baseline vs. test window, population (Kumamoto Sanctuary apes vs. Edinburgh and Planckendael apes), and species was not significant, and therefore we dropped this three-way interaction and reran the model with two-way interactions between each test predictor. These two-way interactions were also not significant, so we dropped them and reran the model with each as an individual test predictor. The full-null model comparison for this reduced exploratory Model 3 was not significant ($\chi 2 = 0.802$, $p = .849$). The individual test predictors were also not significant (Baseline vs. test: estimate = 0.029, $\chi 2 = 7.245$, $p = .007$; Population: estimate = 0.005, $\chi 2 = 0.014$, $p = .906$; Species: estimate = -0.006, $\chi 2 = 0.020$, $p = .887$), indicating that there were no significant differences in attention in the baseline vs. test window between species or population (see Table 4).

Table 4

Exploratory DLS Model 3 Output

Factor	Estimate	S.E.	$\chi 2$	df	*p*-value
BaselinevTest (Test)	0.029	0.034	0.749	730.356	.387
Population	0.005	0.045	0.014	33.298	.906
Species	-0.006	0.042	0.020	28.949	.887
Trial Number	-0.012	0.022	0.271	23.409	.603

Note. Predictors of biases in attention toward target and distractor conspecific faces between baseline and test windows (with test window as the reference category). DLS was used as the dependent measure. Subject, avatar dyad, and trial ID were included as random intercepts. Trial number was included as a random slope for the subject and avatar dyad random intercepts. Estimates and SEs are taken from the model summary; $\chi 2$ values, degrees of freedom and *p* values are taken from the Anova (type III sum of squares) output.

Discussion

In this study, we conducted a combined eye-tracking and playback experiment to investigate whether non-language-trained chimpanzees and bonobos could cross-modally integrate auditory and visual perception to recognize the caretaker-given names of their groupmates. We found only minimal evidence that apes preferentially attend to the named individual after hearing the name uttered twice, but no evidence that apes show sustained (5 s) attention toward their groupmate after that individual's name has been called. Apes showed preferential attention to the named individual but only in time bin 5, mid-way through the post-call phase of the trial. Thus, evidence in support of the hypothesis that apes recognize the heterospecific, caretaker-given names of their groupmates and integrate this information into their cross-modal representation of those individuals is limited. From this evidence, we are unable to conclude that apes are able to use social vocal labels to direct attention toward a static image of a named individual (avatar). However, the limited evidence that apes may do so warrants further investigation and provides predictions for future studies about the temporal manifestation of these attention-based effects.

We believe our study is the first to combine eye tracking with a playback design to measure nonhuman animals' cross-modal perception of social vocal labels. We carefully controlled for the identities of the avatars in this experiment. Across trials presenting the same pair of avatars (within-subjects), each individual was the target (named) and distractor (unnamed) an equal number of times. We also used a relatively large number of groupmates as avatars within each population, and tested across four different ape populations, demonstrating that these limited effects generalize across individuals and populations. The use of eye-tracking also allowed us to eliminate all of the movement or behavioral cues that could trigger biased attention and would be present if the experiment was performed in a live group setting.

Interestingly, we find limited effects only when comparing DLS and difference scores across one-second time bins but not when comparing DLS across baseline and test windows. This inconsistency could owe to the overall attentional patterns observed in our experiment. The most pronounced biases revealed by the time bin models appear toward the end of the trial at a time when overall attention has waned. As a result, these biases could be less detectable in the total test window, where the smaller raw fixation durations from the biased windows are drowned out by the larger raw fixation durations during earlier time bins within the test window. This pattern of results underscores the importance of using eye movement metrics that capture different information, in order to thoroughly characterize behavioral and cognitive phenomena (Hopper et al., 2021; Lewis & Krupenye, 2022). Our exploratory analyses also provide time-based hypotheses for future studies aimed at replicating or extending these findings. As we only found significant attentional biases towards the groupmate whose name was called in time bin 5, in the middle of the post-call window, it may be the case that it takes apes a couple of seconds to fully process spoken names after they are heard and/or that the effect is only detectable in the most engaged apes who continue to attend to the photos this far into the trial. However, it also remains possible that this is a spurious effect, and thus future research should investigate the specific timelines of social vocal label processing in great apes. A

previous study found that a chimpanzee had an event-related potential (ERP) response to the sound of her own caretaker-given name being called, however this response was also delayed as compared to human ERP responses (Hirata et al., 2011; Ueno et al., 2009). Thus, it may be that chimpanzees have delayed initial responses to social vocal labels as compared to humans.

One limitation of our study is that it required ape participants to integrate multimodal information from *three* separate sources (audio from one human caretaker, and images of two different groupmates). Reliable responses may have been hindered by apes' confusion about the artificial production of the social vocal label and whether they should look for the named individuals in real life versus on the screen. Although nonhuman primates have previously been shown to use cross-modal perception and integrate auditory and visual information to identify others, these signals have only ever included a maximum of *two* separate sources (i.e., a vocalization from an individual, combined with a static image of that same individual). Future research could include similar paradigms with adjustments to the stimulus presentation or the technology used to present stimuli, for example testing responses via touchscreens. In addition, while eye-tracking allowed us to control for many key variables (e.g., movement of target and distractor apes), it is also possible that our static screen-based paradigm was insufficiently naturalistic and that subjects would show stronger responses in live experiments. Further directions could test whether these responses are more pronounced in live action tasks using motion-tracking to measure head movements and body positions in response to hearing social vocal labels, rather than more artificial and potentially less motivating screen-based tasks. Apes may have failed to associate others' names with static images of their faces perhaps because the production of the name was somewhat artificial (playback through a speaker), the images of the faces were static, or a combination of both. In addition, the ape participants did not produce the names themselves but were instead required to understand these social vocal labels that were assigned and produced by humans, which may have limited apes' association, recall, and/or comprehension of these labels. We also have a limited understanding of the developmental patterns of these abilities. Although this study included a large number of participants compared to other experimental studies with great apes (N = 24), there were only a few individuals in our study within younger age groups. Thus, future studies with a larger number of younger ape participants could clarify the development of comprehension of vocal labels in great apes.

Although previous research demonstrates that great apes gain knowledge and make predictions and decisions by attending to third-party interactions (de Waal, 1982; Krupenye et al., 2016; Krupenye & Hare, 2018; Slocombe et al., 2010; Wittig et al., 2014), it remains uncertain if chimpanzees and bonobos have an ability to spontaneously recognize social vocal labels like the proper, caretaker-given names of their groupmates. If the marginal effect that was found in time bin 5 was robust and replicable, this would point to the possibility that the capacity to understand spoken language may have evolved before the ability to produce complex human language, which is foundational for communication in cooperative interactions (Levinson, 2006). This is somewhat evidenced by the bonobo Kanzi, who was heavily trained in a language-enriched environment and could recognize some spoken proper names (Rabinowitz, 2016; Savage-Rumbaugh et al., 1986, 1993). Similarly, research with dogs suggests that they have an ability to spontaneously learn many vocal labels of categories, and with training some exceptional dogs can learn a number of object-names (Fugazza & Miklósi, 2020; Fugazza et al., 2021). However, the apes included in the present study were not trained to learn specific words or any other type of language comprehension, and thus it may be that these species have the biological potential to learn social vocal labels with extensive training, but are unable to comprehend these vocal labels without training. If nonhuman great apes do not have the capacity to spontaneously comprehend spoken proper names without formal training, it could be that this ability is unique to humans and perhaps dependent on natural, complex language. Further research on the evolutionary timeline of the comprehension of social vocal labels is crucial for clarifying the stratified evolution of modern human communication.

If captive apes do recognize others' names and use others' names to direct their attention to third-party interactions, then they would likely be able to build third-party knowledge from employing their comprehension of social vocal labels. This builds on previous findings from other primates, which indicate that primates do attend to and learn from third-party interactions (Cheney & Seyfarth, 1990a, 1990b, 2007;

Slocombe et al., 2010). There is even some evidence that primates can integrate vocal information with third-party knowledge to direct attention and behavior to relevant individuals or events (Cheney & Seyfarth, 1980; Wittig et al., 2014). Upon hearing playback recordings of infant screams (and without any other apparent cues), vervet monkeys looked to the infants' respective mothers. Wittig and colleagues also found that, after observing a fight and hearing recordings of aggressive vocalizations from a bystander, chimpanzees looked longer and moved away more often when the calls were from the former opponent's bond partner compared to when they were from the former opponent's non-bond partner (Wittig et al., 2014). This creative use of gaze to study social knowledge suggests that the capacity to integrate vocal information and third-party knowledge to recognize and discriminate between groupmates is present in monkeys as well as apes. In our study, apes showed a limited understanding of spoken proper names - utterances with no evolved relevance to them. The findings by Cheney and Seyfarth - and work showing that primates also match the voices and faces of familiar individuals (Hashiya, 1999; Kojima et al., 2003; Sliwa et al., 2011) - suggest that this capacity may build on adaptive cognitive mechanisms for linking vocal information (including third-party vocal information) with multimodal representations of individual identity. Finally, Tagaki and colleagues found that household cats look longer at an image of a cat with whom they cohabitate after hearing another cat's name called (incongruent condition), and look less long at the image if they hear the same cat's name called (congruent condition) (Takagi et al., 2022). Although results from café cats did not match, and the design only required cats to attend to a single individual rather than multiple individuals, it suggests that some nonprimate animal species may also have the ability to cross-modally match social vocal labels and specific individuals. Thus, there is a possibility that some animal species are able to integrate visual and auditory information of others, but captive primates do not spontaneously learn the spoken proper names that their caretakers use to identify their groupmates.

Previous research indicates that bottlenose dolphins themselves produce social vocal labels that link vocal and visual information of familiar individuals through the use of "signature whistles" (Bruck et al., 2022; Janik & Sayigh, 2013; Quick & Janik, 2012). Signature whistles are distinctive calls that each individual uniquely develops in the first months of life, and bottlenose dolphins copy the signature whistles of conspecifics. Bottlenose dolphins most often use signature whistles when approaching others especially in groups, which suggests that they use these signature whistles as part of a beginning sequence in interaction. Thus, bottlenose dolphins likely use these signature whistles to address distinct individuals much in the same way that humans use proper names, and therefore have a flexible, cross-modal, socially-specific communication system (Bruck et al., 2022; Janik & Sayigh, 2013; Quick & Janik, 2012). In addition, marmosets have recently been found to use distinct phee calls to vocally label their groupmates, use similar phee calls to label others in family groups, and accurately perceive and respond to specific phee calls that are directed at them (Oren et al., 2024). At present they are the only primate species found to potentially use social vocal labels (although see Jaakkola, 2025), and thus it is unknown whether this capacity is homologous or has evolved separately in dolphins, marmosets, and humans. In addition, these species have been found to use cross-modal perception to identify others using *conspecific* production of social vocal labels, whereas the present study aimed to test great apes' cross-modal perception of social vocal labels produced by a *heterospecific* human caretaker. Therefore, further research with additional primate and nonprimate species comparing these multiple types of cross-modal perception would help to address this current phylogenetic puzzle and help bolster our understanding of the evolution of vocal communication, cross-modal perception, and complex language.

The results from the present study raise important questions about the sociolinguistic precursors shared between humans and our primate relatives and the evolutionary trajectory of language and communication in our own great ape lineage. Here, we find only very limited evidence that nonhuman apes associate vocal labels with familiar individuals, and thus cannot conclude that they can comprehend social vocal labels. This limited evidence suggests that further investigation to determine the extent of nonhuman great apes' cross-modal perception would be informative. Additional research using novel paradigms, expanded ape populations, and a greater number of nonprimate species will be necessary to decode the phylogenetic patterns of the comprehension and use of both conspecific and human-given vocal labels. If these capacities are indeed shared with nonhuman great apes, they are likely to serve as evolutionary

foundations of human communication that preceded the emergence of complex spoken language in our species.

Acknowledgments

We thank Edinburgh Zoo, Zoo Planckendael, and Kumamoto Sanctuary for permission and support conducting this research. At Edinburgh Zoo, we thank Kate Grounds, Donald Gow and the Budongo Research Unit caretakers; at Planckendael Zoo, we thank Marjolein Osieck, Zjef Pereboom and the bonobo caretakers; and at Kumamoto Sanctuary, we thank Satoshi Hirata for providing photos and Kim Livingstone, Naruki Morimura, Yutaro Sato, Hanling Yeow, and James Brooks for their assistance in conducting this research. We thank Steve Worthington for his assistance with R coding and statistical analyses. We are grateful to the Royal Zoological Society of Scotland (RZSS) for core financial support to the RZSS Edinburgh Zoo's Budongo Research Unit where this project was carried out. We are grateful to the RZSS, Zoo Planckendael, and Kumamoto Sanctuary keeping and veterinary staff for their care of animals and technical support throughout this project. Finally, we thank all of the chimpanzees and bonobos who participated in this study.

Data Accessibility: Data are available on GitHub: https://github.com/LauraLewis15/Ape-Eye-Tracking-Names-Data

Author Contributions: L.S.L., J.M.G.S., J.D., J.C., and C.K. designed the experiment with input from co-authors, L.S.L. prepared the stimuli, L.S.L. and F.K. conducted the experiment, L.S.L. and C.K. analyzed the data, L.S.L. drafted the manuscript, and all authors provided feedback on the manuscript.

Funding: This research was supported by a Harvard Mind Brain Behavior Interfaculty Initiative Graduate Student Award and a Harvard GSAS Predissertation Summer Fellowship to L.S.L; Japan Society for the Promotion of Science KAKENHI Grants 19H01772 and 20H05000 to F.K.; European Research Council Synergy Grant 609819 SOMICS to J.C.; and European Commission Marie Skłodowska-Curie Fellowship (MENTALIZINGORIGINS), Templeton World Charity Foundation grant (TWCF-2021-20647), and CIFAR Azrieli Global Scholars grant to C.K. Edinburgh Zoo's Budongo Research Unit is core supported by the Royal Zoological Society of Scotland (Registered charity number: SC004064) through funding generated by its visitors, members and supporters, and by the University of St Andrews (Registered charity number: SC013532) who core supports the maintenance and management costs of the research facility.

Conflict of Interest: We declare no competing interests.

References

Aberson, C. (2002). Interpreting null results: Improving presentation and conclusions with confidence intervals. *Journal of Articles in Support of the Null Hypothesis*, *1*(3), 7.

Adachi, I., & Fujita, K. (2007). Cross-modal representation of human caretakers in squirrel monkeys. *Behavioural Processes*, *74*(1), 27–32. https://doi.org/10.1016/j.beproc.2006.09.004

Adachi, I., Kuwahata, H., & Fujita, K. (2007). Dogs recall their owner's face upon hearing the owner's voice. *Animal Cognition*, *10*(1), 17–21. https://doi.org/10.1007/s10071-006-0025-8

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Bruck, J. N., Walmsley, S. F., & Janik, V. M. (2022). Cross-modal perception of identity by sound and taste in bottlenose dolphins. *Science Advances*, *8*(20), eabm7684. https://doi.org/10.1126/sciadv.abm7684

Cäsar, C., & Zuberbühler, K. (2012). Referential alarm calling behaviour in New World primates. *Current Zoology*, *58*(5), 680–697. https://doi.org/10.1093/czoolo/58.5.680

Carvajal, L., & Krupenye, C. (2025). Mental representation of the locations and identities of multiple hidden agents or objects by a bonobo. *Proceedings of the Royal Society B: Biological Sciences.*

Cheney, D. L., & Seyfarth, R. M. (1980). Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour*, *28*(2), 362–367. https://doi.org/10.1016/S0003-3472(80)80044-3

Cheney, D. L., & Seyfarth, R. M. (1990a). *How Monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press.

Cheney, D. L., & Seyfarth, R. M. (1990b). The representation of social relations by monkeys. *Cognition*, *37*(1–2), 167–196. https://doi.org/10.1016/0010-0277(90)90022-C

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon Metaphysics: The Evolution of a Social Mind*. University of Chicago Press.

Davenport, R. K., Rogers, C. M., & Russell, I. S. (1973). Cross modal perception in apes. *Neuropsychologia*, *11*(1), 21–28. https://doi.org/10.1016/0028-3932(73)90060-2

de Waal, F. (1982). *Chimpanzee Politics: Power and Sex Among Apes*. The Johns Hopkins University Press.

Doherty, E., Davila-Ross, M., & Clay, Z. (2023). Multimodal communication development in semiwild chimpanzees. *Animal Behaviour*, *201*, 175–190. https://doi.org/10.1016/j.anbehav.2023.03.020

Evans, T. A., Howell, S., & Westergaard, G. C. (2005). Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(4), 399–406. https://doi.org/10.1037/0097-7403.31.4.399

Faragó, T., Pongrácz, P., Range, F., Virányi, Z., & Miklósi, Á. (2010). 'The bone is mine': Affective and referential aspects of dog growls. *Animal Behaviour*, *79*(4), 917–925. https://doi.org/10.1016/j.anbehav.2010.01.005

Fox, John, et al. (2012). *Package 'car' (2012).* Vienna: R Foundation for Statistical Computing. https://cran.microsoft.com/snapshot/2017-06-17/web/packages/car/car.pdf

Fugazza, C., Dror, S., Sommese, A., Temesi, A., & Miklósi, Á. (2021). Word learning dogs (*Canis familiaris*) provide an animal model for studying exceptional performance. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-93581-2

Fugazza, C., & Miklósi, Á. (2020). Depths and limits of spontaneous categorization in a family dog. *Scientific Reports*, *10.1*, 1–9. https://doi.org/10.1038/s41598-020-59965-6

Genty, E., & Zuberbühler, K. (2014). Spatial reference in a bonobo gesture. *Current Biology*, *24*(14), 1601–1605. https://doi.org/10.1016/j.cub.2014.05.065

Ghazanfar, A. A., & Logothetis, N. K. (2003). Facial expressions linked to monkey calls. *Nature*, *423*(6943), 937–938. https://doi.org/10.1038/423937a

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour*, *32*(1), 182–193. https://doi.org/10.1016/S0003-3472(84)80336-X

Graham, K. E., Badihi, G., Safryghin, A., Grund, C., & Hobaiter, C. (2022). A socio-ecological perspective on the gestural communication of great ape species, individuals, and social units. *Ethology Ecology & Evolution*, *34*(3), 235–259. https://doi.org/10.1080/03949370.2021.1988722

Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLOS Biology*, *16*(2), e2004825. https://doi.org/10.1371/journal.pbio.2004825

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. https://doi.org/10.1038/nature06288

Hashiya, K. (1999). Auditory-visual intermodal recognition of conspecifics by a chimpanzee (*Pan troglodytes*). *Primate Research*, *15*(3), 333–342. https://doi.org/10.2354/psj.15.333

Hashiya, K., & Kojima, S. (2001). Acquisition of auditory–visual intermodal matching-to-sample by a chimpanzee (Pan troglodytes): Comparison with visual—visual intramodal matching. *Animal Cognition*, *4*(3), 231–239. https://doi.org/10.1007/s10071-001-0118-3

Hirata, S., Matsuda, G., Ueno, A., Fuwa, K., Sugama, K., Kusunoki, K., Fukushima, H., Hiraki, K., Tomonaga, M., & Hasegawa, T. (2011). Event-related potentials in response to subjects' own names: A comparison between humans and a chimpanzee. *Communicative & Integrative Biology*, *4*(3), 321–323. https://doi.org/10.4161/cib.4.3.14841

Hopper, L. M., Gulli, R. A., Howard, L. H., Kano, F., Krupenye, C., Ryan, A. M., & Paukner, A. (2021). The application of noninvasive, restraint-free eye-tracking methods for use with nonhuman primates. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01465-6

Hurford, J. R., & Hurford, J. R. (2007). *The Origins of Meaning: Language in the Light of Evolution*. Oxford University Press: Oxford.

Jaakkola, K. (2025). Do marmosets really have names? *Learning & Behavior*. https://doi.org/10.3758/s13420-024-00662-z

Janik, V. M., & Sayigh, L. S. (2013). Communication in bottlenose dolphins: 50 years of signature whistle research. *Journal of Comparative Physiology A*, *199*(6), 479–489. https://doi.org/10.1007/s00359-013-0817-7

Kano, F., & Call, J. (2014). Cross-species variation in gaze following and conspecific preference among great apes, human infants and adults. *Animal Behaviour*, *91*, 137–150. https://doi.org/10.1016/j.anbehav.2014.03.011

Kano, F., Hirata, S., Call, J., & Tomonaga, M. (2011). The visual strategy specific to humans among hominids: A study using the gap–overlap paradigm. *Vision Research*, *51*(23), 2348–2355. https://doi.org/10.1016/j.visres.2011.09.006

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: A comparative eye-tracking study. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1664), 1949–1955. https://doi.org/10.1098/rspb.2008.1811

Kawaguchi, Y., Kano, F., & Tomonaga, M. (2019). Chimpanzees, but not bonobos, attend more to infant than adult conspecifics. *Animal Behaviour*, *154*, 171–181. https://doi.org/10.1016/j.anbehav.2019.06.014

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Kisilevsky, B. S., Hains, S. M. J., Brown, C. A., Lee, C. T., Cowperthwaite, B., Stutzman, S. S., Swansburg, M. L., Lee, K., Xie, X., Huang, H., Ye, H.-H., Zhang, K., & Wang, Z. (2009). Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, *32*(1), 59–71. https://doi.org/10.1016/j.infbeh.2008.10.002

Kojima, S., Izumi, A., & Ceugniet, M. (2003). Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates*, *44*(3), 225–230. https://doi.org/10.1007/s10329-002-0014-8

Kripke, S. (1980). *Naming and Necessity*. Basil and Blackwell.

Krupenye, C., & Hare, B. (2018). Bonobos prefer individuals that hinder others over those that help. *Current Biology*, *28*(2), 280-286.e5. https://doi.org/10.1016/j.cub.2017.11.061

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech In infancy. *Science*, *218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425–2438. https://doi.org/10.1121/1.417951

Kulahci, I. G., Drea, C. M., Rubenstein, D. I., & Ghazanfar, A. A. (2014). Individual recognition through olfactory–auditory matching in lemurs. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1784), 20140071. https://doi.org/10.1098/rspb.2014.0071

Levinson, S. C. (2006). On the Human "Interaction Engine." In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (1st ed., pp. 39–69). Berg. https://doi.org/10.4324/9781003135517-3

Lewis, L., Kano, F., Stevens, J., DuBois, J., Call, J., & Krupenye, C. (2021). Bonobos and chimpanzees preferentially attend to familiar members of the dominant sex. *Animal Behaviour*, *177*, 193–206.

Lewis, L. S., & Krupenye, C. (2022). Eye-tracking as a window into primate social cognition. *American Journal of Primatology*, *84*(10), e23393. https://doi.org/10.1002/ajp.23393

Lewis, L. S., Wessling, E. G., Kano, F., Stevens, J. M. G., Call, J., & Krupenye, C. (2023). Bonobos and chimpanzees remember familiar conspecifics for decades. *Proceedings of the National Academy of Sciences*, *120*(52), e2304903120. https://doi.org/10.1073/pnas.2304903120

Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental Psychology*, *16*(6), 597–607. https://doi.org/10.1037/0012-1649.16.6.597

Lonsdorf, E. V., Engelbert, L. M., & Howard, L. H. (2019). A competitive drive? Same-sex attentional preferences in capuchins. *American Journal of Primatology*, e22998. https://doi.org/10.1002/ajp.22998

Macedonia, J. M., & Evans, C. S. (2010). Essay on contemporary issues in ethology: Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology*, *93*(3), 177–197. https://doi.org/10.1111/j.1439-0310.1993.tb00988.x

Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, *109*(18), 6862–6867. https://doi.org/10.1073/pnas.1113194109

Nakamura, K., Takimoto-Inose, A., & Hasegawa, T. (2018). Cross-modal perception of human emotion in domestic horses (*Equus caballus*). *Scientific Reports*, *8*(1), 8660. https://doi.org/10.1038/s41598-018-26892-6

Oren, G., Shapira, A., Lifshitz, R., Vinepinsky, E., Cohen, R., Fried, T., Hadad, G. P., & Omer, D. (2024). Vocal labeling of others by nonhuman primates. *Science*, *385*(6712), 996–1003. https://doi.org/10.1126/science.adp3757

Pardo, M. A., Lolchuragi, D. S., Poole, J., Granli, P., Moss, C., Douglas-Hamilton, I., & Wittemyer, G. (2024). Female African elephant rumbles differ between populations and sympatric social groups. *Royal Society Open Science*, *11*(9), 241264. https://doi.org/10.1098/rsos.241264

Proops, L., McComb, K., & Reby, D. (2009). Cross-modal individual recognition in domestic horses (*Equus caballus*). *Proceedings of the National Academy of Sciences*, *106*(3), 947–951. https://doi.org/10.1073/pnas.0809127105

Quick, N. J., & Janik, V. M. (2012). Bottlenose dolphins exchange signature whistles when meeting at sea. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2539–2545. https://doi.org/10.1098/rspb.2011.2537

Rabinowitz, A. (2016). *Linguistic competency of bonobos (Pan paniscus) raised in a language-enriched environment* (Master's thesis, Iowa State University). https://doi.org/10.31274/etd-180810-5422

Rendall, D., Seyfarth, R. M., Cheney, D. L., & Owren, M. J. (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour*, *57*(3), 583–592. https://doi.org/10.1006/anbe.1998.1031

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., Rumbaugh, D. M., & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, *58*(3/4), i–252. https://doi.org/10.2307/1166068

Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General*, *115*(3), 211–235. https://doi.org/10.1037/0096-3445.115.3.211

Seyfarth, R. M., Cheney, D. L., & Bergman, T. J. (2005). Primate social cognition and the origins of language. *Trends in Cognitive Sciences*, *9*(6), 264–266. https://doi.org/10.1016/j.tics.2005.04.001

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behaviour*, *28*(4), 1070–1094. https://doi.org/10.1016/S0003-3472(80)80097-2

Sliwa, J., Duhamel, J.-R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice–face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences*, *108*(4), 1735–1740. https://doi.org/10.1073/pnas.1008169108

Slocombe, K. E., Kaller, T., Call, J., & Zuberbühler, K. (2010). Chimpanzees extract social information from agonistic screams. *PLOS ONE*, *5*(7), e11473. https://doi.org/10.1371/journal.pone.0011473

Sorrentino, C. M. (2001). Children and adults represent proper names as referring to unique individuals. *Developmental Science*, *4*(4), 399–407. https://doi.org/10.1111/1467-7687.00181

Stoehr, A. M. (1999). Are significance thresholds appropriate for the study of animal behaviour? *Animal Behaviour*, *57*(5), F22–F25. https://doi.org/10.1006/anbe.1998.1016

Takagi, S., Saito, A., Arahori, M., Chijiiwa, H., Koyasu, H., Nagasawa, M., Kikusui, T., Fujita, K., & Kuroshima, H. (2022). Cats learn the names of their friend cats in their daily lives. *Scientific Reports*, *12*(1), 6155. https://doi.org/10.1038/s41598-022-10261-5

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science,* *331*(6016), 10.1126/science.1199198. https://doi.org/10.1126/science.1199198

Tomasello, M., & Call, J. (1997). *Primate Cognition*. Oxford University Press.

Ueno, A., Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., Matsuda, G., Fukushima, H., Hiraki, K., Tomonaga, M., & Hasegawa, T. (2009). Brain activity in an awake chimpanzee in response to the sound of her own name. *Biology Letters*, *6*(3), 311–313. https://doi.org/10.1098/rsbl.2009.0864

Wittig, R. M., Crockford, C., Langergraber, K. E., & Zuberbühler, K. (2014). Triadic social interactions operate across time: A field experiment with wild chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1779), Article 1779. https://doi.org/10.1098/rspb.2013.3155.

Supplementary Materials

Table S1

Characteristics of Study Participants

Individual	Species	Sex	Date of Birth	Age (years)	Facility
Frek	Chimpanzee	M	10/21/93	25.69	Edinburgh Zoo
Louis	Chimpanzee	M	7/26/76	42.93	Edinburgh Zoo
Qafzeh	Chimpanzee	M	3/31/92	27.25	Edinburgh Zoo
Rene	Chimpanzee	M	2/21/93	26.36	Edinburgh Zoo
Velu	Chimpanzee	M	6/24/14	5.02	Edinburgh Zoo
Edith	Chimpanzee	F	4/11/96	23.22	Edinburgh Zoo
Kilimi	Chimpanzee	F	2/20/93	26.36	Edinburgh Zoo
Iroha	Chimpanzee	F	09/05/08	10.49	Kumamoto Sanctuary
Mizuki	Chimpanzee	F	12/16/96	22.22	Kumamoto Sanctuary
Hatsuka	Chimpanzee	F	06/20/08	10.70	Kumamoto Sanctuary
Misaki	Chimpanzee	F	01/14/99	20.14	Kumamoto Sanctuary
Natsuki	Chimpanzee	F	07/08/05	14.15	Kumamoto Sanctuary
Zamba	Chimpanzee	M	07/26/95	24.11	Kumamoto Sanctuary
Habari	Bonobo	M	1/29/06	13.04	Planckendael Zoo
Kikongo	Bonobo	M	1/29/14	5.04	Planckendael Zoo
Rubani	Bonobo	M	4/10/16	2.85	Planckendael Zoo
Djanoa	Bonobo	F	3/27/95	23.89	Planckendael Zoo
Nayoki	Bonobo	F	3/24/12	6.89	Planckendael Zoo
Vijay	Bonobo	M	12/28/03	15.44	Kumamoto Sanctuary
Junior	Bonobo	M	1/14/95	24.39	Kumamoto Sanctuary
Lolita	Bonobo	F	4/20/89	30.13	Kumamoto Sanctuary
Connie Lenore	Bonobo	F	2/3/82	37.34	Kumamoto Sanctuary
Ikela	Bonobo	F	11/27/91	27.52	Kumamoto Sanctuary
Louise	Bonobo	F	10/28/72	46.60	Kumamoto Sanctuary

Note. Age is age at time of testing

Table S2

Characteristics of Participant Groups

Facility	Total # of Individuals	# Males	# Females
Edinburgh Zoo – Chimpanzees	15	8	7
Planckendael Zoo – Bonobos	13	7	6
Kumamoto Sanctuary – Bonobos	6	2	4
Kumamoto Sanctuary – Chimpanzees	6	1	5