Cross-modal individual recognition in domestic horses (*Equus caballus*)

Leanne Proops[1], Karen McComb[1], and David Reby

Centre for Mammal Vocal Communication Research, Department of Psychology, School of Life Sciences, University of Sussex, Brighton BN1 9QH, United Kingdom

Individual recognition is considered a complex process and, although it is believed to be widespread across animal taxa, the cognitive mechanisms underlying this ability are poorly understood. An essential feature of individual recognition in humans is that it is cross-modal, allowing the matching of current sensory cues to identity with stored information about that specific individual from other modalities. Here, we use a cross-modal expectancy violation paradigm to provide a clear and systematic demonstration of cross-modal individual recognition in a nonhuman animal: the domestic horse. Subjects watched a herd member being led past them before the individual went of view, and a call from that or a different associate was played from a loudspeaker positioned close to the point of disappearance. When horses were shown one associate and then the call of a different associate was played, they responded more quickly and looked significantly longer in the direction of the call than when the call matched the herd member just seen, an indication that the incongruent combination violated their expectations. Thus, horses appear to possess a cross-modal representation of known individuals containing unique auditory and visual/olfactory information. Our paradigm could provide a powerful way to study individual recognition across a wide range of species.

animal cognition | vocal communication | social behavior | playback experiment | expectancy violation

How animals classify conspecifics provides insights into the social structure of a species and how they perceive their social world (1). Discrimination of kin from nonkin, and of individuals within both of these categories, is proposed to be of major significance in the evolution of social behavior (2, 3). Individual recognition can be seen as the most fine-grained categorization of conspecifics, and there is considerable interest in discovering the prevalence and complexity of this ability across species. While individual recognition is generally believed to be widespread (4), there is much debate as to what constitutes sound evidence of this ability (5). To demonstrate individual recognition, a paradigm must show that (*i*) discrimination operates at the level of the individual rather than at a broader level, and (*ii*) there is a matching of current sensory cues to identity with information stored in memory about that specific individual. Numerous studies to date have provided evidence for some form of social discrimination of auditory stimuli, but how this is achieved remains unclear. It is of considerable interest to establish whether any animal is capable of cross-modal integration of cues to identity, as this would suggest that in addition to the perception and recognition of stimuli in one domain, the brain could integrate such information into some form of higher-order representation that is independent of modality.

A number of species have been shown to make very fine-grained discriminations between different individuals (6–8). For example, in the habituation–dishabituation paradigm, subjects that are habituated to the call of one known individual will dishabituate when presented with the calls of a different known individual. What is unclear from this result is whether discrimination occurs because listeners simply detect an acoustic difference between the two calls or because, on hearing the first call, listeners form a multi-modal percept of a specific individual and then react strongly to the second call not only because of its different acoustic properties, but also because it activates a multi-modal percept of a different individual. Alternative approaches that go some way to addressing this issue have shown subjects to be capable of associating idiosyncratic cues with certain forms of stored information such as rank category or territory (9–12). However, paradigms to date do not allow researchers to adequately assess whether this form of recognition is cross-modal.

One rigorous way to demonstrate cross-modal individual recognition is to show that an animal associates a signaler's vocalization with other forms of information they have previously acquired in another modality that are uniquely related to that signaler. By presenting a cue to the identity of a familiar associate in one modality and then, once that cue is removed, presenting another familiar cue, either congruous or incongruous in another modality, we can assess whether the presentation of the first cue activates some form of preexisting multi-modal representation of that individual, creating an "expectation" that the subsequent cue will correspond to that associate.

In our study, horses were shown 1 of 2 herd mates who was then led past them and disappeared behind a barrier. After a delay of at least 10 s, the subjects were then played either a call from that associate (congruent trial) or a call from another familiar herd mate (incongruent trial) from a loudspeaker placed close to the point of disappearance. Each of 24 horses participated in a total of four trials: a congruent trial where they saw stimulus horse 1 and heard stimulus horse 1, an incongruent trial where they saw stimulus horse 1 and heard stimulus horse 2, and, the balanced counterparts, a congruent trial where they saw stimulus horse 2 and heard stimulus horse 2 and an incongruent trial where they saw stimulus horse 2 and heard stimulus horse 1 (Fig. 1). Four different pairs of stimulus horses were presented across subjects with six subjects being exposed to each pair. Each subject was presented with a different call exemplar for each associate, so that a total of 48 exemplars were used in the playbacks. The order of trials was counterbalanced across subjects.

If horses have some form of cross-modal representation of known individuals, and this representation contains unique auditory, visual, and potentially olfactory information, we predicted that the presentation of mismatched cues to identity would violate their expectations. This violation of expectation would be indicated by a faster response time and a longer looking

See Commentary on page 669.

[1]To whom correspondence may be addressed: l.c.proops@sussex.ac.uk or karenm@ sussex.ac.uk.
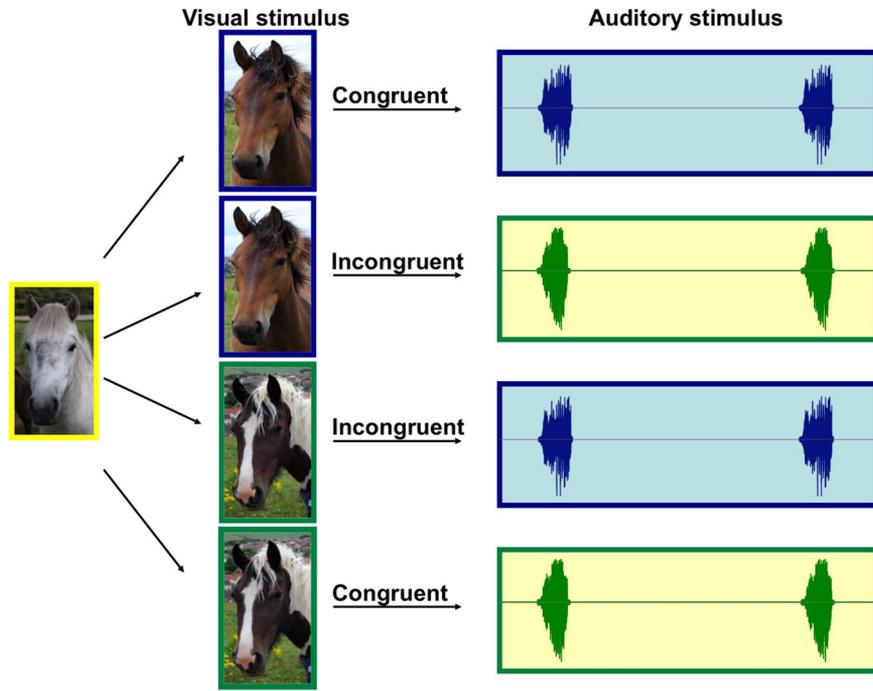
PSYCHOLOGY

Fig. 1. Diagrammatic representation of the experimental paradigm, as applied to one of our 24 subjects. Each subject receives a balanced set of four trials (detailed in text).

time in the direction of the vocalization during incongruent compared to congruent trials (13).

Results

Behavioral responses in the 60 s following the onset of the playbacks are shown in Fig. 2. As predicted, horses responded more quickly to incongruent calls than to congruent calls (response latency: $F_{1,20} = 5.136$, $P = 0.035$); they also looked in the direction of the stimulus horse more often and for a longer time in the incongruent trials (number of looks, $F_{1,20} = 4.730$, $P = 0.042$; total looking time, $F_{1,20} = 5.208$, $P = 0.034$). There was no significant difference in the duration of the first look (duration of first look: $F_{1,20} = 2.802$, $P = 0.110$). These responses did not differ significantly from the first presentation of the calls in trials one and two to the second presentation in trials three and four (response latency, $F_{1,20} = 0.900$, $P = 0.354$; number of looks, $F_{1,20} = 0.019$, $P = 0.891$; total looking time, $F_{1,20} = 0.455$, $P = 0.508$; duration of first look, $F_{1,20} = 0.118$, $P = 0.735$). Neither did subject results differ significantly according to which stimulus horse pair they were presented with (response latency, $F_{3,20} = 1.278$, $P = 0.350$; number of looks, $F_{3,20} = 1.707$, $P = 0.198$; total looking time, $F_{3,20} = 2.098$, $P = 0.309$; duration of first look, $F_{3,20} = 1.996$, $P = 0.147$). There were no significant interactions between the factors. Horses called in response to playbacks during 12 of the 96 trials and calling was not obviously biased toward congruent or incongruent trials (subjects called in 5 congruent compared to 7 incongruent trials).

For each of the four behavioral responses, the scores for the congruent trials were also subtracted from those for the incongruent trials to produce overall recognition ability scores for each subject. Unlike some species, where social knowledge appears to be greater in older animals (7), this research showed no evidence that the ability to recognize the identity of the callers improved with age in an adult population (response latency, $r_{22} = -0.015$, $P = 0.944$; total looking time, $r_{22} = -0.014$, $P = 0.947$; number of looks, $r_{22} = -0.057$, $P = 0.793$; duration of first look, $r_{22} = 0.203$, $P = 0.341$). Neither were there any differences in the recognition abilities of male and female

horses (response latency, $F_{1,22} = 0.021$, $P = 0.885$; total looking time, $F_{1,22} = 0.153$, $P = 0.700$; number of looks, $F_{1,22} = 0.002$, $P = 0.967$; duration of first look, $F_{1,22} = 0.070$, $P = 0.794$).

Discussion

Overall, horses responded quicker, and looked for a longer time, during trials in which the familiar call heard did not match the familiar horse previously seen, indicating that the incongruent combination violated their expectations. Given that the stimulus horse was out of sight when the vocal cue was heard, our paradigm requires that some form of multi-modal memory of that individual's characteristics had to be accessed/activated for this result to be obtained. This is the first clear empirical demonstration that in the normal process of identifying social companions of its own species, a nonhuman animal is capable of cross-modal individual recognition.

The ability to transfer information cross-modally was once thought to be unique to humans (14). At the neural level, however, areas responsible for the integration of audiovisual information have now been located in the primate brain. Images of species-specific vocalizations or the vocalizations themselves each produce activation of the auditory cortex and higher-order visual areas and areas of association cortex that contain neurons sensitive to multi-modal information (15–17). This neural circuitry corresponds closely to areas in the human brain that support cross-modal representation of conspecifics and in which differences in activity have been found for presentation of congruent vs. incongruent face–voice pairs (16, 18). Such similarities between human and animal brain function suggest that the possession of higher-order representations that are independent of modality is not unique to humans. Indeed, Ghazanfar (19) considers the neocortex to be essentially multisensory in nature, implying that some degree of cross-modal processing is likely to be widespread across mammal taxa.

At the behavioral level, a number of species have recently proved capable of integrating multisensory information in a socially relevant way. Nonhuman primates can process audiovisual information cross-modally to match indexical cues (20)
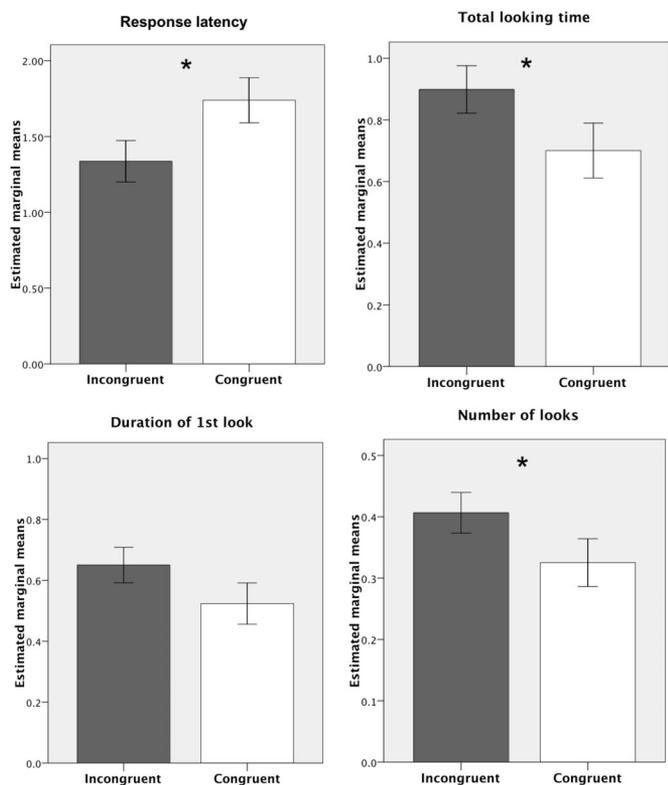
Proops *et al.*

Fig. 2. Estimated marginal means ± SEM for the behavioral responses of subjects during incongruent and congruent trials (*, $P < 0.05$).

and number of vocalizers (21), to match and tally quantity across senses (22) and associate the sound of different call types with images of conspecifics and heterospecifics producing these calls (23–25). Research aimed specifically at investigating the categorization of individuals has shown that hamsters (*Mesocricetus auratus*) are capable of matching multiple scent cues to the same individual (26) and that certain highly enculturated chimps (*Pan troglodytes*) can, through intensive training, learn to associate calls from known individuals with images of those individuals (27–29). Some species have also been shown to spontaneously integrate auditory and visual identity cues from their one highly familiar human caretaker during interspecific, lab-based trials (30, 31).

Here we demonstrate cross-modal individual recognition of conspecifics in a naturalistic setting by providing evidence that horses possess cross-modal representations that are precise enough to enable discrimination between, and recognition of, two highly familiar associates. The use of two stimulus horses randomly chosen from the herd indicates that our subjects are capable of recognizing the calls of a larger number of familiar individuals, an essential feature of genuine individual recognition (5, 8). If one associate had been slightly more familiar or preferred causing subjects to respond primarily on the basis of differing levels of familiarity, this would have produced biases in favor of the cues of certain individuals that would have been detected in the pattern of the results.

Conducting cross-modal expectancy violation studies in a controlled yet ecologically relevant setting provides the opportunity to reevaluate and extend the findings of field studies by formally assessing the cognitive processes at work in these situations. Such studies have demonstrated that elephants (*Loxodonta africana*) keep track of the whereabouts of associates by using olfactory cues (32) and that some primates can distinguish between the sound of congruous and incongruous rank inter-

actions and react to acoustic information in ways that suggest they may match calls to specific individuals (9, 33, 34). Our results indicate that cross-modal individual recognition may indeed underpin the complex classification of conspecifics reported and potentially provides a practical and standardized method through which this possibility could be tested directly.

Understanding the extent and nature of abilities to form representations across species is key to understanding the evolution of animal communication and cognition and is of interest to psychologists, neuroscientists, and ethologists. Our demonstration of the spontaneous multisensory integration of cues to identity by domestic horses presents a clear parallel to human individual recognition and provides evidence that some nonhuman animals are capable of processing social information about identity in an integrated and cognitively complex way.

Materials and Methods

Study Animals. Twenty-four horses, 12 from Woodingdean livery yard, Brighton, U.K., and 12 from the Sussex Horse Rescue Trust, Uckfield, East Sussex, U.K., participated in the study. Ages ranged from 3 to 29 years (12.63 ± 1.56) and included 13 gelded males and 11 females. At both sites subjects live outside all year in fairly stable herds of ≈30 individuals. The horses from Woodingdean yard are privately owned and are brought in from the herd regularly for feeding; some of them are ridden. The horses at the Sussex Horse Rescue Trust are checked once a day but remain with the herd most of the time. Four horses from each study site were chosen to be "stimulus horses"; these were randomly selected from the horses for which we had recorded a sufficient number of good quality calls. The horses chosen as subjects were unrelated to the stimulus horses, had no known hearing or eyesight problems, and were comfortable with being handled. Only horses that had been part of the herd for at least 6 months could participate to ensure that the subjects would be familiar with the appearance and calls of the stimulus horses. The subjects had not been used in any other studies.

Call Acquisition. We recorded the long distance contact calls (whinnies) of herd members ad libitum. Whinnies are used by both adult and young horses when separated from the group (35–37). All calls were recorded in a situation where the horses had been isolated either from the herd or from a specific herd mate or when the horses were calling to their handler around feeding time. Recordings were made of both herds between February and September, 2007 by using a Sennheisser MKH 816 directional microphone with windshield linked to a Tascam HD-P2 digital audio recorder. Calls were recorded in mono at distances between 1 and 30 m, with a sampling frequency of 48 kHz and a sampling width of 24 bit. Six good-quality recordings, taken on at least two separate occasions, were randomly chosen for each stimulus horse as auditory stimuli. This enabled us to present each subject with a unique call from each stimulus horse to avoid the problem of pseudoreplication (38).

Playback Procedure. Subjects were held by a naive handler on a loose lead rope during trials to prevent them from walking away or approaching their associates. For each subject, one of two possible "stimulus horses" was held for ≈60 s a few meters in front of them. The stimulus horse was then led behind a barrier and from this point of disappearance, after a delay of at least 10 s, the subjects were played two identical calls with a 15-s interval between each call. Previous work has shown that horses have a short-term spatial memory of <10 s in a delayed response task (39); thus the delay in our recognition task ensured that some form of stored information had to be accessed. The call played was either from the stimulus horse just seen (congruent trial) or from the other stimulus horse (incongruent trial). The subjects were given the four counter-

balanced combinations with inter-trial intervals of at least 4 days to prevent habituation (see Fig. 1).

Four pairs of stimulus horses were used, being presented to six subjects each. Presentation of trials and stimulus horses were counterbalanced in that each stimulus horse was used in trials one and two (one congruent, one incongruent) for three subjects and in trials three and four for the other three subjects. The order of congruent and incongruent trials was counterbalanced across subjects with the constraint that the calls of the stimulus horses were presented alternately.

Vocalizations were played from a Liberty Explorer PB-2500-W powered speaker attached to a Macbook Intel computer. Intensity levels were normalized to 75 dB ($\pm$ 5 dB) and calls were broadcast at peak pressure level of 98 dB peak SPL at 1 m from the source (taken as the average output volume of subjects recorded previously). Responses were recorded by using a Sony digital handycam DCR-TRV19E video recorder. Handlers were naive to the identity of the callers. They were asked to hold the horses in front of the video camera on as loose a rope as possible, allowing the horses to graze and move around freely on the rope. The handlers remained still, looking at the ground and did not interact with the subjects.

Behavioral and Statistical Analysis of Responses. Videotapes were converted to .mov files and analyzed frame by frame (frame, 0.04 s) on a Mac G4 powerbook, by using Gamebreaker 5.1 video analysis software (40). The total time subjects spent looking in the direction of the speaker in the 60 s following the onset of the playbacks was recorded. The onset of the look was defined as the frame at which the horse's head began to move toward the speaker, having previously been held in another position. A fixed look toward the speaker was then given and the end of the look was taken to be the frame at which the horse began to move his head in a direction away from the speaker. A look was defined as being in the direction of the call if the horse's nose was facing a point within 45° to the left or the right of the speaker. The duration of the first look toward the speaker, the total number of looks, and the total looking time were recorded. Latency to respond to the call was also measured and defined as the number of seconds between the onset of the call and the beginning of the first look in the direction of the speaker. For subjects that did not respond, a maximum time of 60 s was assigned.

The videos were coded blind in a random order. Twenty videos (20.8%) were scored by a second coder, providing an inter-observer reliability of 0.968 ($P < 0.0001$) for total looking time, 0.992 ($P < 0.0001$) for response latency, 0.907 ($P < 0.0001$) for number of looks, and 0.995 ($P < 0.0001$) for duration of first look, measured by Spearman's rho correlation. The distributions of scores for total looking time, duration of first looks, and number of looks were positively skewed and so were $\log_{10}$ transformed to normalize the data. The distribution of the response latency scores was bimodal and so a fourth root transformation was performed. Results were analyzed by using $2 \times 2 \times 4$ mixed-factor ANOVAs with condition (congruent/incongruent) and trial (trials 1 and 2 using stimulus horse 1/trials 3 and 4 using stimulus horse 2) as within-subject factors and stimulus pair (which pair of stimulus horses were presented) as a between-subjects factor. Each dependent variable (total look duration, duration of first look, number of looks, and response latency) was analyzed in a separate ANOVA. The effect of age and sex was investigated by subtracting the congruent responses from the incongruent responses and adding up the total for each subject to obtain a measurement of the magnitude of difference in responding to congruent and incongruent trials for each subject, i.e., (incongruent trial 1 − congruent trial 1) + (incongruent trial 2 – congruent trial 2). A score of 0 would indicate no difference in the behavior of a subject across the trial types, a positive score would indicate a larger response to the incongruent trials, and a negative score would indicate a greater response to the congruent trials. This measurement of the degree of recognition by each subject was calculated for the four behavioral responses and was correlated with age of subjects, by using Pearson's correlation coefficient. The recognition score for male and female subjects was compared by using a one-way ANOVA.

1. Cheney DL, Seyfarth RM (1982) Recognition of individuals within and between groups of free-ranging vervet monkeys. *Am Zool* 22(3):519–529.
2. Hamilton WD (1963) The evolution of altruistic behavior. *Am Nat* 97:354–356.
3. Trivers RL (1974) Parent-offspring conflict. *Am Zool* 14(1):249–264.
4. Sayigh LS, *et al.* (1999) Individual recognition in wild bottlenose dolphins: a field test by using playback experiments. *Anim Behav* 57:41–50.
5. Tibbetts EA, Dale J (2007) Individual recognition: it is good to be different. *Trends Ecol Evol* 22:529–537.
6. Janik VM, Sayigh LS, Wells RS (2006) Signature whistle shape conveys identity information to bottlenose dolphins. *Proc Natl Acad Sci USA* 103(21):8293–8297.
7. McComb K, Moss C, Durant SM, Baker L, Sayialel S (2001) Matriarchs as repositories of social knowledge in African elephants. *Science* 292(5516):491–494.
8. Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim Behav* 51:1007–1015.
9. Seyfarth RM, Cheney DL, Bergman TJ (2005) Primate social cognition and the origins of language. *Trends Cogn Sci* 9(6):264–266.
10. Husak JF, Fox SF (2003) Adult male collared lizards, *Crotaphytus collaris*, increase aggression toward displaced neighbours *Anim Behav* 65:391–396.
11. Peake TM, Terry AMR, McGregor PK, Dabelsteen T (2001) Male great tits eavesdrop on simulated male-to-male vocal interactions. *Proc R Soc Lond Ser B Biol Sci* 268(1472):1183–1187.
12. Stoddard PK (1996) Vocal recognition of neighbours by territorial passerines. *Ecology and Evolution (Lawrence, Kans) of Acoustic Communication in Birds*, eds Kroodsma DE, Miller EH (Cornell University Press, Ithaca, NY), pp 356–374.
13. Spelke ES (1985) Preferential looking methods as tools for the study of cognition in infancy. *Measurement of Audition and Vision in the First Year of Postnatal Life*, eds Gottleib G, Krasnegor N (Ablex, Norwood, NJ), pp 323–363.
14. Ettlinger G (1967) Analysis of cross-modal effects and their relationship to language. *Brain Mechanisms Underlying Speech*, ed Darley FL (Grune & Stratton, New York), pp 53–60.
15. Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci* 28(17):4457–4469.
16. Gil-Da-Costa R, *et al.* (2004) Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. *Proc Natl Acad Sci USA* 101(50):17516–17521.
17. Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. (Translated from English) *J Neurosci* 26(43):11138–11147 (in English).
18. Campanella S, Belin P (2007) Integrating face and voice in person perception. (Translated from English) *Trends Cogn Sci* 11(12):535–543 (in English).
19. Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10(6):278–285.
20. Ghazanfar AA, *et al.* (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr Biol* 17:425–430.
21. Jordan KE, Brannon EM, Logothetis NK, Ghazanfar AA (2005) Monkeys match the number of voices they hear to the number of faces they see. *Curr Biol* 15(11):1034–1038.
22. Jordan KE, Maclean EL, Brannon EM (2008) Monkeys match and tally quantities across senses. *Cognition* 108(3):617–625.
23. Ghazanfar AA, Logothetis NK (2003) Neuroperception: facial expressions linked to monkey calls. *Nature* 423(6943):937–938.
24. Evans TA, Howell S, Westergaard GC (2005) Auditory-visual cross-modal perception of communicative stimuli in tufted capuchin monkeys (Cebus apella). *J Exp Psychol-Anim Behav Process* 31(4):399–406.
25. Parr LA (2004) Perceptual biases for multimodal cues in chimpanzee (Pan troglodytes) affect recognition. *Anim Cogn* 7(3):171–178.
26. Johnston RE, Peng A (2008) Memory for individuals: hamsters (Mesocricetus auratus) require contact to develop multicomponent representations (concepts) of others. *J Comp Psychol* 122(2):121–131.

27. 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.

28. Bauer HR, Philip MM (1983) Facial and vocal individual recognition in the common chimpanzee. *Psychol Rec* 33:161–170.

29. Izumi A, Kojima S (2004) Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes). *Anim Cogn* 7(3):179–184.

30. Adachi I, Fujita K (2007) Cross-modal representation of human caretakers in squirrel monkeys. *Behav Processes* 74(1):27–32.

31. Adachi I, Kuwahata H, Fujita K (2007) Dogs recall their owner's face on hearing the owner's voice. *Anim Cogn* 10(1):17–21.

32. Bates LA, *et al.* (2008) African elephants have expectations about the locations of out-of-sight family members. *Biol Lett* 4:34–36.

33. Cheney DL, Seyfarth RM (1997) Reconciliatory grunts by dominant female baboons influence victims' behaviour. *Anim Behav* 54:409–418.

34. Cheney DL, Seyfarth RM (1980) Vocal recognition in free-ranging vervet monkeys. *Anim Behav* 28(May):362–367.

35. Tyler SJ (1972) The behaviour and social organisation of the New Forest ponies. *Anim Behav Monogr* 5:85–196.

36. Kiley M (1972) The vocalizations of ungulates, their causation and function. *Z Tierpsychol* 31:171–222.

37. Feist JD, McCullough DR (1976) Behavior patterns and communication in feral horses. *Z Tierpsychol-J Comp Ethol* 41(4):337–371.

38. Kroodsma DE (1989) Suggested experimental-designs for song playbacks. *Anim Behav* 37:600–609.

39. McLean AN (2004) Short-term spatial memory in the domestic horse. *Appl Anim Behav Sci* 85(1–2):93–105.

40. Sportstec (GamebreakerWarriewood, NSW, Australia), 5.1.