Primates (2006) 47: 350–354
DOI 10.1007/s10329-006-0182-z

ORIGINAL ARTICLE

Ikuma Adachi · Hiroko Kuwahata · Kazuo Fujita
Masaki Tomonaga · Tetsuro Matsuzawa

Japanese macaques form a cross-modal representation of their own species in their first year of life

Received: 3 August 2005 / Accepted: 23 January 2006 / Published online: 25 April 2006

Abstract We tested whether infant Japanese macaques (*Macaca fuscata*) have a cross-modal representation of their own species. We presented monkeys with a photograph of either a monkey or a human face on an LCD monitor after playing back a vocalization of one of those two species. The subjects looked at the monitor longer when a human face was presented after the monkey vocalization than when the same face was presented after human vocalization. This suggests that monkeys recall and expect a monkey's face upon hearing a monkey's voice.

Keywords Cross-modal representation · Japanese macaque · (*Macaca fuscata*) · Natural concept

Introduction

Although a variety of non-human species have been shown to form various natural concepts (e.g., Herrnstein and Loveland 1964; Herrnstein et al. 1976; Cerella 1979; Yoshikubo 1985), previous studies have not demonstrated an important aspect of human concepts, namely, their multi-modal nature. Human natural concepts contain exemplars in various sensory modalities. For example, our concept of dogs includes their visual appearances, their smells, their vocalizations, etc. More importantly, in humans, each exemplar of a concept naturally activates other exemplars of the same concept across other sensory modalities. For instance, we activate visual images of dogs when we hear their barking without seeing the animals.

In a previous study (I. Adachi et al., submitted), we demonstrated that dogs possess a multi-modal representation of their owner that has a cross-modal function described above. We used an expectancy violation procedure in which we presented to dogs a vocalization followed by a photograph of a face, either matching or mismatching in personal identity with the vocalization. Our hypothesis was that if the subjects activate a visual representation of the person upon hearing the voice, they should be surprised and look at the photograph for longer if a mismatching face is presented than if a matching face is presented. The results suggested that dogs actively generated visual representations of their owner upon hearing the voice of the owner.

The next question we need to ask concerns the generality of this finding. Three considerations seem particularly important. One is how widespread the existence of such cross-modality of concepts might be in the animal kingdom. Non-human primates have been shown to form associations across sensory modalities. For example, a female chimpanzee was successfully trained to match noises of objects like castanets or voices of familiar trainers to their respective photographs (Hashiya and Kojima 1999, 2001). More recently, Ghazanfar and Logothetis (2003) found that rhesus monkeys looked longer at videos of monkeys showing the facial expression that matched a simultaneously presented vocalization than at a non-matching video in a preferential looking procedure. Thus, it is quite likely that non-human primates spontaneously activate visual images of animate and inanimate objects when they hear the noise or voice associated with the objects.

The second consideration is how this ability develops. Human infants learn arbitrary auditory–visual pairings

I. Adachi (✉) · H. Kuwahata · K. Fujita
Department of Psychology, Graduate School of Letters,
Kyoto University, Yoshida-Honmachi, Sakyo,
Kyoto 606-8501, Japan
E-mail: iadachi@bun.kyoto-u.ac.jp
Tel.: +81-75-7532759
Fax: +81-75-7532759

I. Adachi · H. Kuwahata
Japan Society for the Promotion of Science, Kyoto, Japan

M. Tomonaga · T. Matsuzawa
Section of Language and Intelligence, Primate Research Institute,
Kyoto University, Kyoto, Japan

between voices and faces, from the age of 3 months (Brookes et al. 2001). The ontogeny of cross-modal recognition is not well known in non-humans. Gunderson et al. (1990) examined cross-modal recognition abilities of 6-week-old infant pigtailed macaques (*Macaca nemestrina*). The animals were given an object to explore tactually in a darkened room. Following this familiarization period, they were visually presented the tactually familiar object along with a novel object, and looking time to each stimulus was recorded. The infants showed a significant novelty response, which appeared to be based on the discriminability of the test objects. The results are similar to those obtained for human infants tested using the same paradigm. More studies are needed to examine whether such cross-modality is adapted to the senses other than those tested and how it changes during early development.

The third consideration concerns the kinds of concept that incorporate this cross-modality aspect. Dogs diverged from the common ancestor with wolves somewhere between 35,000 and 100,000 years ago (Vilá et al. 1997). It is thought that dogs have been selected for sophisticated skills in interacting and communicating with humans during their long history of domestication and close cohabitation. Domestication may have enhanced dogs' abilities to form cross-modal representation of their owners.

In the present study, we extended our approach in the directions indicated above; we used infant Japanese macaques (*Macaca fuscata*) as subjects and explored their concept of species. We used the same expectancy violation procedure as was used in the previous study of dogs. This procedure requires no intensive training of subjects and thus is particularly useful for comparative and developmental approaches.

Recognizing one's own species is fundamental not only to reproduction but also coherence and stability of social groups. It is already known that several macaque species, including Japanese macaques, have a preference for viewing photographs of their own species (Fujita 1987; Fujita and Watanabe 1995; Fujita et al. 1997). This suggests that macaque monkeys appear to have a concept of their own species that incorporates visual aspects. However, visual information is sometimes of limited value, especially for forest-living species. A cross-modal representation that allows visually biased species to activate a representation in their favored modality from information received in other sensory modalities such as audition or olfaction would be highly advantageous. Thus, we may expect their concept of species to extend to the auditory modality because they naturally hear vocalizations of conspecifics in association with their visual images. However, this question has never been asked experimentally.

The main purpose of this study was to examine whether infant Japanese macaques, an Old World monkey species, would show evidence of a multi-modal representation of species (their own species and human) which would be activated upon perceiving the appropriate vocalization. By doing this, we attempted to answer the three questions introduced above, namely: (1) whether human-like cross-modality of concepts is shared only by highly domesticated animals such as dogs; (2) when and how such multi-modal concepts develop in primate infants; and (3) whether such concepts extend to things other than those that have special importance for animals, such as owners for dogs.

Methods

Subjects, stimuli, and apparatus

We used 15 infant Japanese macaques in their first year (range 26–152 days, mean ± SD 99 ± 38.3), born in a large social group (Takahama group) at the Primate Research Institute, Kyoto University, Inuyama, Japan. We tested them when the whole group was temporarily caught for a regular health and physical checkup in summer (four younger infants—range 26–66 days, mean ± SD 42.75 ± 16.8) and fall (11 older infants—range 99–152 days, mean ± SD 119.81 ± 15.16). The smaller number of younger subjects was because we tested them only if they were fully awake during the test session. The group contained nearly 60 individuals in an outdoor enclosure. The subjects consequently had extensive experience of seeing and hearing conspecifics, but much less experience with humans.

The following four test stimuli were prepared: (1) a photograph of an unfamiliar adult female Japanese macaque against an ivory-colored background (PM), (2) a photograph of an unfamiliar human male (PH) against the same background, (3) a vocalization by a female Japanese macaque (VM), and (4) a vocalization by the human (VH). The monkey vocalization was a "coo-call", typically used to solicit social contact with other individuals. The human vocalization was "ooi", which Japanese people typically use to attract another's attention. The two vocalizations were of approximately equal duration.

The stimuli were prepared as follows. All vocalizations were stored on the computer in WAV format, with a sampling rate of 44,100 Hz and a sampling resolution of 16-bit. The amplitude of the vocalizations sounded equivalent to human ears. We took a digital full-face photograph of each stimulus individual and stored the photograph on the computer in JPEG format sized at 450×550 (W×H) pixels, or ca. 20×25 cm on the 18.1-inch LCD monitor (SONY SDM-M81) used for presentation. We set up the apparatus as shown in Fig. 1 in an experimental room at the Primate Research Institute. Briefly, the LCD monitor was located about 50 cm from the subject's face. At the start of the test, a black opaque screen (50×70 cm) was placed in front of the monitor to prevent the subject from seeing the monitor. A digital camcorder (SONY DCR-TRV-30), located behind the monitor, recorded the subject's behavior. Presentation of the stimuli was controlled by a Visual Basic 5.0 pro-

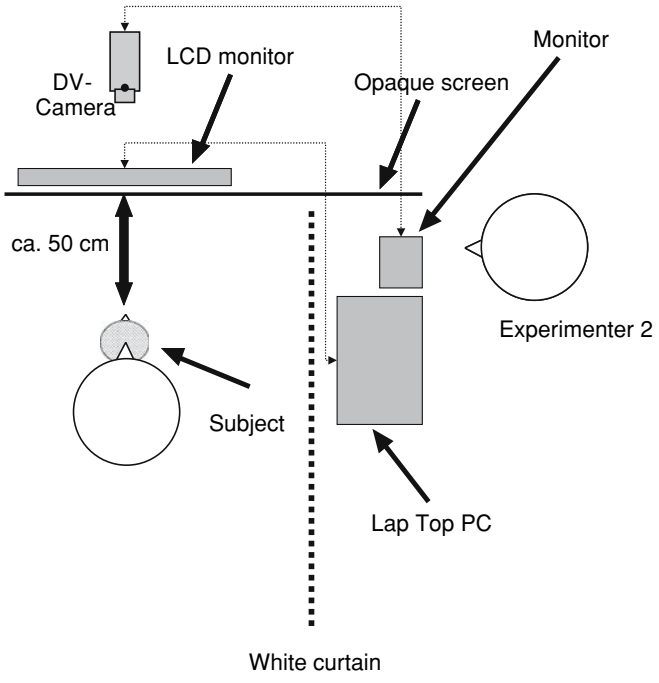

Fig. 1 A schematic drawing of the apparatus used to test Japanese macaques (*Macaca fuscata*). The experimental area was separated into two parts by a white curtain. Experimenter 1 held the subject, while Experimenter 2 operated the personal computer to present stimuli, and removed the screen so that the subject could see the visual stimuli on the monitor

gram on a laptop personal computer (Dell Inspiron 4100, with Pentium III 1.2 GHz).

Procedure

We used an expectancy violation procedure often used to test human infants for inferences about causality between external events (e.g., Wynn 1992). Typically, the subjects are shown one event several times and then a second event. It is assumed that the subjects should be surprised at the second event if it contradicts what they expect based on the first. Our previous experiment on dogs has shown that this procedure is useful for investigating multi-modal representations (I. Adachi et al., submitted).

Each trial consisted of the following events. One experimenter (Experimenter 1) held the monkey in a towel on his or her lap in front of the LCD monitor, and remained silent, stationary, and passive throughout trials. A second experimenter (Experimenter 2), who observed the subject via a 2.5-inch television monitor connected to the camcorder behind the LCD monitor, started the trial when the subject appeared calm and alert, and oriented toward the LCD monitor. Each trial consisted of two phases. The first was the voice phase and the second was the photograph phase. In the voice phase, one of the two vocalizations was played back through the speakers installed in the monitor, every 2 s for a total of three vocalizations. The duration of each vocalization was about 600 ms. The photograph phase began immediately after the final vocalization. Experimenter 2 smoothly removed the opaque screen to reveal a face on the LCD monitor. This experimenter was always positioned behind the curtain (Fig. 1) and was ignorant of the precise face shown on the monitor. The photograph phase lasted 15 s. Behaviors of the subjects in this phase were video-recorded for later analysis. We used the opaque screen to facilitate subjects' understanding that something was hidden behind the screen. Each subject was given the following four types of test trials: (1) a VM–PM trial, in which the monkey photograph appeared after the monkey vocalization; (2) a VH–PH trial, in which the photograph of the human appeared after his vocalization; (3) a VH–PM trial, with the monkey photograph following the human vocalization; and (4) a VM–PH trial, in which the photograph of the human followed the monkey vocalization. The vocalization and the photograph matched in the former two trials but mismatched in the latter two.

These four trials were presented in semi-random order with the restriction that the same vocalization was not repeated on consecutive trials. The intertrial interval was about 5 min, during which subjects were run on other experiments involving non-social stimuli. We hypothesized that, if the subject generated a visual image of the appropriate species upon hearing a vocalization, it would be surprised at the mismatch in the latter two types of test trials (VH–PM and VM–PH trials), and thus would look at the photograph for longer than in the other two types of test trials.

Results

After the experiments, the videos of trials were captured on a personal computer and converted to MPEG file format (30 frames per second). A coder who was blind to the stimuli recorded the duration of subjects' looks at the monitor in the photograph phase. A second coder scored data for eight randomly sampled subjects to check the reliability of coding. The correlation between total looking time for each trial measured by the two coders was highly significant (Pearson's $r = 0.941$, $n = 32$, $p < 0.01$).

We calculated total looking time in each trial for each subject. Figure 2 shows the mean duration of looking at the monitor in the photograph phase for each condition averaged for all subjects, and summarized data on matched versus mismatched trials regardless of face stimuli. Looking times were analyzed by means of a 2×2 repeated-measures analysis of variance with photographs (monkey or human) and conditions (match or mismatch) as factors. There was no significant main effect of photograph ($F_{1,14} = 1.556$, $p = 0.233$), but a significant main effect of condition ($F_{1,14} = 6.245$, $p = 0.026$) and a significant interaction between the two factors ($F_{1,14} = 15.054$, $p = 0.002$).

As a post-hoc analysis, we compared looking time between match and mismatch conditions in the trials where the same face was presented: that is, between VH–PM and VM–PM and between VH–PH and VM–PH, using conservative paired *t*-tests with alpha set at 0.025 because there were two comparisons. The *t*-tests revealed no significant difference between the two conditions VM–PM and VH–PM: left two bars ($t_{14} = 0.484$, $p > 0.1$) in which the monkey face was presented. In contrast, there was a highly significant difference between VH–PH and VM–PH trials (right two bars) ($t_{14} = -5.014$, $p < 0.001$), in which the human face was presented (Fig. 2).

To examine possible developmental changes, we compared younger ($n = 4$) and older ($n = 11$) infants (Fig. 3). Their looking times were analyzed by means of a 2×2×2 analysis of variance with photographs (monkey or human), conditions (match or mismatch), and age groups (younger or older) as factors. This analysis again revealed no significant main effect of photograph ($F_{1,13} = 0.244$, $p > 0.630$), but a significant main effect of condition ($F_{1,13} = 5.416$, $p = 0.037$) and a significant interaction between photograph and condition $F_{1,13} = 9.899$, $p = 0.008$). No interactions with age groups were significant, although the main effect of age groups approached significance ($F_{1,1} = 3.596$, $p = 0.080$). Looking time of younger infants tended to be shorter than older infants overall, but not as a function of stimulus. However, with only four infants for younger group, studies are called for.

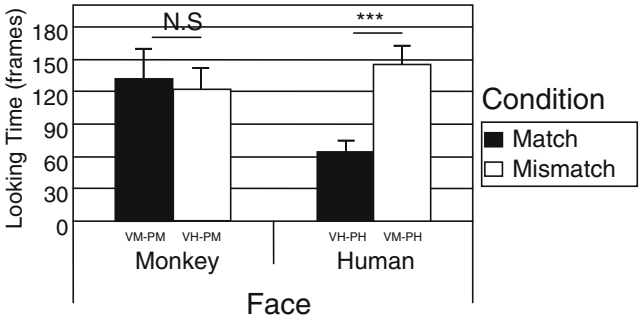

Fig. 2 Duration of looking at the monitor in the photograph phase for each condition averaged for all subjects. *Triple asterisks* indicate $p < 0.001$ by a paired *t*-test

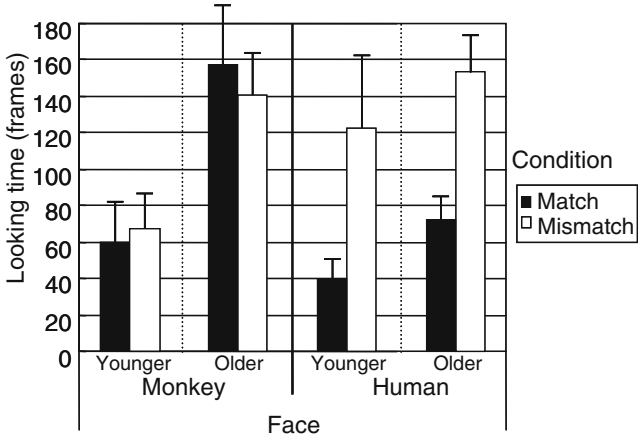

Fig. 3 Duration of looking at the monitor in the photograph phase for each condition for both age groups

Discussion

The present results show that infant monkeys in their first year of life have already formed a multi-modal representation of their own species that incorporates at least visual and auditory senses. More importantly, they not only associate auditory and visual information but also activate the visual representation upon hearing a corresponding vocalization. We found no evidence of differences between the two age groups. It is surprising that very young infants around 1-month-old have already formed this concept; additional tests with younger subjects should explore whether they innately have such concepts of conspecifics or not.

One possible confounding factor is that the experimenter, who held the subjects, might have unwittingly provided the latter with cues. However, this is unlikely because the experimenter was asked to be silent and passive throughout, and was ignorant of the stimuli presented because he was asked to look at the small monitor on the cam-coder to keep the subjects on film.

We have demonstrated that a non-human primate species, the Japanese macaque, has a multi-modal representation of their own species. Thus, this ability is shared by at least two non-human groups, namely primates and carnivores. Comparative studies among species from more taxa on this multi-modal aspect of concepts would reveal how it has evolved. So far, demonstrations have been limited to auditory and visual modalities; conceivably, species that rely more upon other sensory modalities may show cross-modality in other combinations of senses. As described in the Introduction, Gunderson et al. (1990) found that 6-week-old infant pigtail macaques visually discriminated a novel object from an object that they had previously explored tactually. In their experiment, it is unclear whether the subjects generated the visual images before the two objects were presented to the subjects. However, if we combine their results and ours, it appears that macaque monkeys have the ability to form multi-modal concepts incorporating visual, auditory, and tactile senses.

On the other hand, we found no evidence that the infant monkeys had formed such a representation of humans. Thus, we cannot say that their multi-modal concepts extend to things other than those that have special importance for animals. This asymmetrical result may have emerged because generating an appropriate visual image in response to the human vocalization might be more difficult due to the infant monkeys' limited exposure to humans. In a previous study (Adachi

et al. 2003), we found that recognition of biological motion was affected by visual experience. Enclosure-reared macaque monkeys, who were reared in the same environment as the monkeys in the present study, recognized the biological motion of a macaque but not that of a human. In contrast, cage-reared monkeys with extensive visual experience of humans showed the opposite tendency.

One possibility is that an innate mechanism for species recognition might limit the formation of a multi-modal concept of humans in macaques. In fact, although the monkeys in Adachi et al. (2003) recognized biological motion of the species with which they had the most extensive experience, there was a difference in how this recognition developed. Cage-reared monkeys came to prefer human biological motion from the age of 8–15 weeks, whereas the enclosure-reared group preferred macaque biological motion at all ages tested from 0 to 25 weeks. This difference suggests that some innate factors in the development of biological motion perception might interact with experience. Similar processes might influence monkeys' formation of cross-modal representations of species.

Another possible explanation of the asymmetrical results is that the subjects' preference for conspecifics, as in the series of studies by Fujita and colleagues (e.g., Fujita 1987; Fujita and Watanabe 1995; Fujita et al. 1997), would make them look at the stimulus for so long that it could overshadow a difference in looking time between match and mismatch conditions when the conspecific face was presented. However, this seems unlikely because no main effect of face was found.

Further studies of monkeys with more extensive experience with humans are called for to explore whether the formation of a cross-modal representation is specific to recognition of own species or generalizable to others. Also, more studies need to be extended to non-social objects to ask whether these abilities are limited to objects in the social domain.

Finally, a weakness of our experiment is that we used only one photo and one vocalization for each of the two stimulus species. Thus, the multi-modal concept we have demonstrated might concern "monkeys" rather than "own species". Further tests with more exemplars are needed before we can safely conclude that the monkeys have a multi-modal concept of their own species, although such tests may be difficult in practice. However, the concept we have shown could not be individual-specific because both photos and vocalizations were unfamiliar to the subjects during the experiment. Nor can the concept be a consequence of association learning because there was no pairing between stimuli that would result in the formation of such association learning.

Acknowledgements This study was supported by Research Fellowships of the Japan Society of the Promotion of Science (JSPS) for Young Scientists to Ikuma Adachi and Hiroko Kuwahata, the Grants-in-Aid for Scientific Research Nos. 13410026 and 17300085 from JSPS, Japan, to Kazuo Fujita, and by the 21st Century COE Program, D–10, from Ministry of Education, Culture, Sports, Science, and Technology to Kyoto University. It was also supported by the Cooperative Research Program of the Primate Research Institute, Kyoto University. We also thank James R. Anderson for his editing of the manuscript. We would like to express our thanks to all members of the Center for Human Evolution Modeling Research, Primate Research Institute, Kyoto University for their assistance during this study and for the management of the subjects' health. The experiments complied with *The Guide for the care and use of laboratory primates*, Primate Research Institute, Kyoto University.

References

Adachi I, Fujita K, Kuwahata H, Ishikawa S (2003) Perception of biological motion in infant macaques. In: Tomonaga M, Tanaka M, Matsuzawa T (eds) Cognitive and behavioral development in chimpanzees (in Japanese). Kyoto University Press, Kyoto, pp. 333–336

Brookes H, Slater A, Quinn PC, Lewkowicz DJ, Hayes R, Brown E (2001) Three-month-old infants learn arbitrary auditory–visual pairings between voices and faces. Infant Child Dev 10:75–82

Cerella J (1979) Visual classes and natural categories in the pigeon. J Exp Psychol Hum Percept Perform 5:68–77

Fujita K (1987) Species recognition by five macaque monkeys. Primates 28:353–366

Fujita K, Watanabe K (1995) Visual preference for closely related species by Sulawesi macaques. Am J Primatol 37:253–261

Fujita K, Watanabe K, Widarto TH, Suryobroto B (1997). Discrimination of macaques by macaques: the case of Sulawase species. Primates 38:233–245

Ghazanfar AA, Logothetis NK (2003) Neuroperception: facial expressions linked to monkey calls. Nature 423:937–938

Gunderson VM, Rose SA, Grant-Webster KS (1990) Cross-modal transfer in high- and low-risk infant pigtailed macaque monkeys. Dev Psychol 26:576–581

Hashiya K, Kojima S (1999) Auditory–visual intermodal matching by a chimpanzee (*Pan troglodytes*). Primate Res 15:333–342

Hashiya K, Kojima S (2001) Acquisition of auditory–visual intermodal matching to sample by a chimpanzee (*Pan troglodytes*): comparison with visual–visual intramodal matching. Anim Cogn 4:231–239

Herrnstein RJ, Loveland DH (1964) Complex visual concept in the pigeon. Science 146:549–551

Herrnstein RJ, Loveland DH, Cable C (1976) Natural concepts in pigeons. J Exp Psychol Anim Behav Process 2:285–302

Vilá C, Savolainen C, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK (1997) Multiple and ancient origins of the domestic dog. Science 276:1687–1689

Wynn K (1992) Addition and subtraction by human infants. Nature 358:749–750

Yoshikubo S (1985) Species discrimination and concept formation by rhesus monkeys (*Macaca mulatta*). Primates 26:285–299