Looking back: The “representational mechanism” of joint attention in an infant chimpanzee (*Pan troglodytes*)

SANAE OKAMOTO

Department of Psychology, Graduate School of Environmental Studies, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan

MASAYUKI TANAKA and MASAKI TOMONAGA

Section of Language and Intelligence, Department of Behavioral and Brain Sciences, Primate Research Institute, Kyoto University, Inuyama Aichi 484-8506, Japan

Abstract: An infant chimpanzee's joint visual attention to objects behind him was investigated. A recent study has shown that a 13-month-old infant chimpanzee can follow human social cues including glancing (Okamoto, Tomonaga, Ishii, Kawai, Tanaka, & Matsuzawa, 2002a). In humans, 12-month-olds do not follow gaze to objects behind them but 18-month-olds do (Butterworth & Jarrett, 1991). In the present study, from 13 months old, the infant chimpanzee had been tested to look at one of two identical objects, which an experimenter indicated by pointing or head turning. The objects were set in front of or behind the subject. In our series of experiments, we used moving or stationary objects as targets. Moreover, the experimenter manipulated a computer at the onset of each block of trials. The results show that by the age of 20 months, the infant reliably followed the experimenter's cues and looked back to the target behind him. Moving targets elicited more responses than stationary targets, and the subject showed more follow responses after having seen the experimenter manipulating the computer.

Key words: joint attention, “representational mechanism,” looking back, social cues, chimpanzee infant.

By the end of their first year, human infants are sensitive to information specifying where others are looking. Scaife and Bruner (1975) demonstrated that infants as young as 2-months-old tend to follow an adult’s gaze. The ability to follow the gaze of other individuals is a critical component of joint visual attention, defined as looking toward the object of

---

1 This research was financially supported by the Cooperative Research Program of the Primate Research Institute, Kyoto University to S. Okamoto, Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan to O. Takenaka (#10CE2005), T. Matsuzawa (#12002009), and M. Tomonaga (#13610086), and the Grant-in-Aid for the 21st Century COE Program (A2 to Kyoto University). We thank O. Takenaka for his generous support of this study, and T. Matsuzawa, K. Bard, and J. Call for their valuable comments on this study. We wish to express our thanks for the critical reading of the manuscript to J. Barth. Thanks are also due to Y. Mizuno, A. Ueno, N. Nakashima, T. Imura, M. Uozumi, S. Hirata, M. Hayashi, Y. Fukiura and the other members of the Language and Intelligence Section for their support of the research project. Especially we thank N. Maeda, K. Kumazaki, A. Kato, J. Suzuki, S. Goto, C. Hashimoto, and K. Matsubayashi for their daily care of the chimpanzees.

2 Correspondence concerning this article should be sent to: Sanae Okamoto, Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. (Email: okamoto@eva.mpg.de)
Joint attention is considered as an early social cognitive ability leading to the later development of the ability to infer others’ mental states (cf. Baron-Cohen, 1995; Tomasello, 1995).

Butterworth and colleagues (Butterworth & Cochran, 1980; Butterworth & Jarrett, 1991; Butterworth, 1991) propose a naturalistic approach of joint visual attention with three successive mechanisms that develop in human infants between the ages of 6–18 months. At 6 months, infants progress gradually from responding to the head movements of others to orienting in the same general direction within their visual field (ecologic mechanism). At this age, however, infants terminate their search at the first salient object in their scan path. By 12 months, infants are able to localize the particular object at which the other is looking (geometric mechanism). They also found that infants establish joint attention to objects within their visual field before they do so for objects outside their visual field. That is, infants younger than 18 months cannot yet represent their whole environment, some region of which might be visible to another person. By 18 months, infants can follow someone else’s gaze into space that is outside their own initial visual field (representational mechanism). On the other hand, Corkum and Moore, who advocate an empiric and parsimonious approach, assumed that when infants follow the adults’ gaze, they often see interesting objects and events, and hence learn to use the gaze direction of others as a cue to where such events might be located (Corkum & Moore, 1995, 1998; Moore, 1999). That is, social learning drives joint attention, though learning is constrained by certain causal and social sensitivities.

This ability is found not only in humans but also in some nonhuman animals, especially in primates. Following the gaze direction of conspecifics is an important ability for social primates because it allows individuals to take advantage of the visual experience of group mates that spot interesting or important objects and events, such as food, predators, or significant social interactions. A variety of primate species visually track the gaze direction of conspecifics to external objects (chimpanzees, mangabeys, and macaques; Emery, Lorincz, Perrett, Oram, & Baker, 1997; Tomasello, Call, & Hare, 1998; Tomonaga, 1999). Moreover, chimpanzees also follow the gaze direction of humans. They can do this on the basis of eye direction alone, independent of head direction (Povinelli & Eddy, 1996; Itakura & Tanaka, 1998), even when the target is located above and/or behind them (Call, Hare, & Tomasello, 1998; Itakura, 1996; Povinelli & Eddy, 1997). In a previous study with an infant chimpanzee, Okamoto et al. (2002a) found that the subject showed reliable following responses to the object that was indicated by various cues, including glancing alone by the age of 13 months. In this previous study, however, since the target objects were within the subject’s visual field, the ability to follow cues to a target outside the subject’s view was not tested. Tomasello, Hare, and Fogleman (2001) reported that chimpanzees less than 3–4 years-old do not look outside their own visual field when using the experimenter’s head turn cue. In their study, however, the target objects were not presented outside the subject’s visual field.

Since joint attention has an important role in the development of social animals, it is also important to understand how the actions of others elicit infants’ joint attention. Along the lines of Corkum and Moore’s empiric and parsimonious account, in another study with human subjects, pointing cues elicited more episodes of joint attention than looking alone, and distinctive and complex targets elicited more episodes of joint attention than identical targets (Deák, Flom, & Pick, 2000). The authors also found that infants looked more at front than at back targets, but there was also an effect of magnitude of head turn. They also suggested that human infants’ joint attention to targets behind them is affected by the distinctiveness and complexity (i.e., interesting) of the targets. Thus, environmental factors also affect the infant’s joint attention.

The present longitudinal study was conducted to clarify the ability of an infant chimpanzee to follow the experimenter-given cues to targets.
outside his visual field, the representational mechanism in Butterworth’s terminology. Moreover, we manipulated two factors to investigate what affects the chimpanzee’s joint attention to objects outside his visual field: distinctiveness and memory of targets.

**Methods**

**Subject**
One male chimpanzee infant, Ayumu, served as the subject from 13 months of age. He had been reared with his mother since birth at the Primate Research Institute, Kyoto University. Both live in a community of 14 chimpanzees in an enriched outdoor compound (Ochiai & Matsuzawa, 1998). He participated in a variety of tests investigating the development of cognitive abilities (Matsuzawa, 2001a,b; Hirata & Celli, 2003; Sousa, Okamoto, & Matsuzawa, 2003; Tomonaga, Tanaka, & Matsuzawa, 2003). He maintained his free-feeding body weight throughout the present study. Care and use of the chimpanzee adhered to the Guide for the care and use of laboratory primates of the Primate Research Institute (1986).

**Apparatus**
The experiment was conducted in the chimpanzee experimental booth (2 m × 2 m × 2.4 m) at the Primate Research Institute. The experimental apparatus consisted of a feeder and two types of objects that served as targets (Figure 1). The feed cylinder was made of one transparent acrylic board and two cylinders (length: 10 cm). It was placed at the middle of a panel in the experimental booth. Two cylindrical holes (4 cm) through which food rewards were given were 50 cm above the floor. In the experiment, we used two types of objects as targets (Figure 2a,b).

**Close Object.** Two identical stationary toys (mean size 2 cm) were presented between the subject and the experimenter. Each of them was attached to one end of an acrylic bar (length: 28 cm). Forty different toys (such as little bears, little colored ball, etc.) were presented across the trials to keep Ayumu interested.

**Distant Object.** Two identical stationary and large toys (mean size 25 cm × 25 cm: such as balloon toys, cartoon mask, etc.) or two identical screens (16 cm × 21 cm) of mobile computers were presented behind the subject. In the beginning of the experiment, we presented a stationary blue screen, and then, we presented one of 32 moving color images (geometric objects, animals, animated cartoons, etc.), in the moving targets condition or just continued to show the stationary blue screen in the stationary targets condition.

Prior to the experiment, the subject had been taken into the experimental booth with his mother since he was 11-days-old (e.g., Okamoto et al., 2002a). Thus, at the beginning of the experiment he was already very familiar with both the experimental booth and the experimenter. The behavior of the subject was video-taped (Figure 2a,b).

**Procedure**

**Gaze following task.** The general procedure was almost the same as in the previous study (see Okamoto et al., 2002a). At first, the subject
came to the experimental booth with his mother. After his mother began working on another task [one of the visual discrimination tasks that had her facing a computer system (Kawai & Matsuzawa, 2000) at another side of the booth], the experiment for the subject began. At the onset of each trial, the experimenter sat on the floor outside the booth, 20 cm behind the objects in the “Close Object” condition (see Figure 1). After the subject approached the front of the experimenter, the experimenter looked at the center of the acrylic board with two food cylinders with her hands resting on her knees. As soon as the subject looked at the experimenter’s face, the experimenter presented a cue to the target object for three seconds, followed by the presentation of the food reward (a piece of fruit) through a cylinder nearest to the target object, independently of the subject’s response. One session was conducted once a week.

Preliminary test. Before shifting to the main experiment, the preliminary test was conducted when the subject was between 13 and 20 months old. This period was a part of another experiment (follow-up training in Okamoto et al., 2002a). This test had started to maintain the subject’s motivation in the experimental setting and test his behavior to the experimenter-given cues. In this phase, we mainly used “Close Object” as target conditions like in the previous study (see the phase 2 of experiment 1 in Okamoto et al., 2002a). We used the same kinds of experimenter-given cues (see Okamoto et al., 2002a), including tapping on the object (10 trials), pointing to the object (10 trials, see Main test), head turning to the object (20 trials, see Main test), glancing toward the object without head movement (8 trials), and no social cue (12 trials). In this phase, each session comprised approximately 60 trials. Moreover, during each month we used the “Distant Object” set-up to test the emergence of looking back behavior presenting the Distant Point cue (see below) as in Phase 1 of the Main test period. The subject received 16 “Distant Point” trials in total.

Main test. At 21 months old, the main test was introduced. The subject was tested to look at one of two identical objects behind him (Figure 1), which the experimenter indicated by giving social cues. In this period, four
types of experimenter-given cues were used (Figure 3). The first two cues were baseline cues: Close Point, where the experimenter gazed at and pointed to the Close Object target with her tip of the index finger, and Close Head turn, where the experimenter turned her head and gazed towards the Close Object target. The other two cues were experimental cues: Distant Point, where the experimenter gazed at and pointed to the Distant Object target with the tip of her index finger, and Distant Head turn, where the experimenter turned her head and gazed towards the Distant Object target. These cues were presented only one time during each trial with the action performed at normal speed as in daily human interaction. (It should be mentioned that it is likely that the subject had previously seen human pointing or head turning during his daily interaction with humans). The distance between the finger or the experimenter’s head and the target object was approximately 230 cm for Distant Point and 245 cm for Distant Head turn, respectively. Irrespective of the subject’s response to experimental cues, no food reward was presented. The subject also received the control (non-cued) trials in Phase 3 (see below). In this condition, the experimenter looked down and put down her hand after attracting the subject’s attention. The food reward was given through either the left or right cylinder according to a predetermined order irrespective of the subject’s response.

**Experimental design.** In the main test, two factors were manipulated: incentive and memory of targets. The main test consisted of four phases. Phase 1: Two stationary and largely identical toys (mean size 25 cm) were presented as Distant Objects. Phase 2a: Two identical computer screens showing moving screen savers served as Distant Objects. The screen savers were activated according to a predetermined time setting (1, 2, 3, 4, 5 or 6 min) at the same time. The order of the time settings was randomized. Approximately 20 s after activating the screen savers, the experimenter indicated one of the Distant Objects by giving one of the experimental cues. Phase 2b: The computers with stationary screen savers served as targets. The computers were not activated. The experimenter indicated one of the Distant Objects by using one of the experimental cues according to the predetermined time setting as in Phase 2a. Sessions of Phases 2a and 2b were randomly alternated. Phase 3: Instead of the
experimental cues, a control condition was used with a moving screen saver according to the predetermined time setting as in Phase 2. Phase 4: This phase is the same as phase 2a. It was considered as the recovery condition for investigating the effect of the cues. In the latter three phases, the experimenter manipulated the computer in front of the subject at the onset of each block of trials.

Each session comprised approximately four blocks of trials. One block consisted of about 20 trials (18 baseline cues: 9 trials each, experimental cues: 2 trials each). During almost every month (at 22, 23, 24, 25, 27, 28, and 29 months of age), we administered only one probe trial to test the subject’s follow responses by using the Distant Head turn cue. In all phases, the position of the target object (left or right) and the order of the conditions were randomized using Gellermann’s (1933) random sequences. The target appeared equally often on the left and right sides.

**Data analysis**

The subject’s responses during cue presentation and after three seconds were categorized into two types for each target object. In Baseline cues using Close Object, the subject’s responses were categorized into “Follow” (the subject looked at the target object) and “no Follow” (all other behaviors including looking at the target object only after receiving a reward, looking at the experimenter’s face, not looking anywhere relevant, and looking at the object on the side opposite to the target object during cue presentation). In Experimental cues using Distant Object, responses were categorized into “Looking back” (the subject looked back to the Distant Object irrespective of left or right) and “no Looking back” (all other behaviors).

The main experimenter (SO) categorized the subject’s responses into these two types on the basis of video recordings. Especially, “Looking back” behavior was scored only when the subject’s face was visible to the observer on the video picture (from camera 2, see Figure 1). To assess interobserver reliability, two additional observers watched the video recordings and judged the subject’s response for a sample of four sessions. Observers were not informed which cues were presented. They judged whether the subject looked at the target to his left or right, behind him. Inter-observer reliability was computed by means of a Cohen’s kappa: mean kappa was 0.90.

**Results**

Table 1 shows the percentage of trials in which the subject made “Follow” responses in each condition in Preliminary and Main tests. The subject’s responses to the baseline social cues were consistently accurate during the present study.

Figure 4 shows the percentage of “looking back” responses in the Distant Point trials as a function of sessions and the subject’s age. The subject did not make any “looking back” responses to the Distant Point cue during the preliminary test. By the age of 21 months (Phase 1 in Main test), however, the subject began to make “Looking back” response although his performance fluctuated. The mean percentage of Phase 1 was 42.2%, significantly higher than preliminary test (Fisher’s exact probability test, $p = 0.0009$).

In Phase 2, the subject’s looking back response to the Distant Point cue increased. These responses decreased gradually from the beginning to the ending of this phase, however, presumably because the subject’s habituated to

<table>
<thead>
<tr>
<th>Table 1. Mean percentages of “follow” responses in the Close Object (baseline cue) trials in preliminary and main tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tap</td>
</tr>
<tr>
<td>Preliminary test</td>
</tr>
<tr>
<td>Main test</td>
</tr>
</tbody>
</table>

the targets gradually. In Phase 2b (stationary target conditions), the average response rate was as low as in Phase 1 (stationary target condition) (Phase 1 = 42.2%; Phase 2a = 73.1%; Phase 2b = 54.5%). There was a significant difference in the number of looking back responses when comparing the last four sessions of Phase 1 and the first four sessions of Phases 2a and 2b (Fisher’s exact probability test, both $p < 0.05$). More interestingly, the subject showed more looking back responses in Phase 2a (moving targets) than in Phase 2b (stationary targets: Wilcoxon’s signed rank test, $N = 5, Z = 2.023, p < 0.05$).

In Phase 3 (control condition), the subject showed considerably fewer looking back responses. On average, he responded to the object behind him on 11.5% of the trials. In Phase 4 (recovery condition) the subject’s looking back responses increased quickly. The mean percentage of Phase 4 was 83.7%.

We compared the performances between each experimental condition (Phases 1, 2a, 2b, and 4) and control condition (Phase 3) using Fisher’s exact probability tests with correction for multiple comparisons using the Bonferroni method. These analyses revealed significantly more follow responses to the Distant Point cue in each of the experimental phases than in the control phase (all $p < 0.05$).

We also gave seven probe trials with the Distant Head turn cue (1 trial in Phase 1, 3 in Phase 2, and 3 in Phase 4), and the subject showed looking back responses in all trials (binomial test using 0.5 as chance probability, $p = 0.008$).

Although he often turned his head or body to the left side (71.7% of total looking back responses), the side he turned his head to and the side the experimenter pointed to matched in 80.4% of total looking back responses (Fisher’s exact probability test, $p = 0.0001$). One more important result to be noted is, however, that the subject did not look at the experimenter again once he looked back behind him.

**Discussion**

In the present study, by the age of 21 months, the infant chimpanzee reliably followed the experimenter’s cues and looked back to the target behind him. Unfortunately, we only tested a single chimpanzee infant in this study. Thus, tests with more subjects are needed to verify the generality of the present results. Nevertheless, the present results have many implications for comparative research on joint attention.

Our previous study (Okamoto et al., 2002a) and the present study clearly indicate that mechanisms of joint attention also emerge successively in an infant chimpanzee as in human infants (cf. Butterworth & Cochran, 1980; Butterworth & Jarrett, 1991). Moreover, factors such as the distinctiveness of targets also influenced the chimpanzee’s joint attention as in human infants (Deák, Flom, & Pick, 2000). The comparison of results between Phase 1 and

---

**Figure 4.** The percentage of “Looking back” responses in the Distant-Point condition as a function of session and the subject’s age.
Phase 2b suggest that the subject’s looking back behavior was facilitated by seeing the experimenter manipulating the computer. Some episodic memory of targets being manipulated may influence joint attention in the sense of increased expectancy of a subsequent event. Furthermore the attractiveness of subsequent events also affected the subject’s response (moving vs. stationary targets both of which were manipulated by the experimenter; Phase 2a vs. 2b). In Phase 2, we introduced 2a (moving target) first, and then, the 2a and 2b (stationary target) conditions alternately. It is possible that the mean responses of 2a are higher than that of 2b, because the more interesting condition was presented first and the conditions were not presented in a random order. However, as can been seen in Figure 4, the main tendency that the looking back response is higher in condition 2a than in the adjacent condition 2b is preserved. This tendency indicates that distinctive (e.g., attractive, interesting) targets elicited more looking back responses than identical ones. Moreover, we stress that the looking back behavior only occurred when the experimenter indicated the targets, even though the response rate decreased gradually. In other words, only the experimenter’s gesture was used as a trigger to look back to the target behind the chimpanzee infant. Additionally, since the side that the subject looked back to often matched the side the experimenter pointed to, we can suggest that the subject’s responses might represent a “representational mechanism.”

Looking back responses according to the other’s gaze cue is apparently evidence for a “representational mechanism” in chimpanzee infants. As noted in the Results section, however, the subject showed a lack of subsequent behavioral sequences after looking back which is a common observation in human infants. That is, the subject did not look at the experimenter again after looking back at the target. This result implies that there must be some substantial difference in “joint attention” between these two species. So that the question arises: are the present results really evidence for “Joint attention” in chimpanzee infants? Chimpanzee infants follow the other’s gaze, but do not interact with him/her. Carpenter, Nagell, and Tomasello (1998) and Tomasello (1999) described the three main types of joint attentional interaction as; check attention (9–12 months), follow attention (11–14 months), and direct attention (13–15 months). According to their distinction, the chimpanzee infant followed the other’s attention but did not direct the other’s attention to the external object. Similarly, Emery (2000) noted that joint attention is different from shared attention. “Joint attention” is the same as “gaze following” except that there is a focus of attention (such as an object). “Shared attention,” on the other hand, is a combination of mutual attention and joint attention, where the focus of attention for both individual A and B is on the object of joint focus and each other. That is to say, “Triad relationship-based” joint attention is an important component of social cognitive skills in human infants older than one year. On the basis of Emery’s definition, the present results suggest that the chimpanzee infant and human experimenter jointly attended to the object behind the infant, but did not share their attention with each other. More strictly speaking, the experimenter’s behavior was a mere trigger for the subject to initiate appropriate reactions such as searching behind him. As in previous studies, it remains unresolved whether chimpanzees attribute referential intent and visual experience to the signaler or merely follow gaze direction geometrically to specific locations (e.g., Povinelli & Eddy, 1996; Tomasello, Hare, & Agnetta, 1999).

Is it impossible for the chimpanzee to engage in real joint (shared) attention, that is, “triad relationship-based joint attention”? In a dyad interaction, infant chimpanzees do show “mutual gaze” when interacting with their mother from early infancy (cf. Bard, Myowa-Yamakoshi, Quinn, Tomonaga, & Matsuzawa, 2002; Okamoto, Kawai, Sousa, Tanaka, Tomonaga, Ishii, & Matsuzawa, 2002b). Furthermore, the present results show that some factors influence joint attention to objects outside the visual field in a chimpanzee infant as in human infants (Déak et al., 2000). In the future,
we should conduct more detailed comparative examinations concerning the developmental changes from dyad- to triad-based interactions involving gazing behaviors and factors affecting these interactions. Such studies will provide a clearer idea of visual communication including joint attention and the understanding of social-cognitive abilities in nonhuman primates.

**References**


(Received May 16, 2003; accepted March 6, 2004)