Short communication

Cross-modal integration and conceptual categorization in baboons

Julie Martin-Malivel, Joël Fagot *

Center for Research in Cognitive Neurosciences, 31 Chemin Joseph Aigier, 13402 Marseille Cedex 20, France

Received 15 September 2000; received in revised form 25 January 2001; accepted 25 January 2001

Abstract

This study investigates concept formation and cross-modal integration in baboons. Response times were recorded in a categorical task involving discrimination between human and baboon vocalizations. We show that a brief presentation of human or baboon prime pictures conceptually related to the target sound shortened response speed of one baboon. Cross-modal priming effects were replicated with degraded pictures, and were also found in a sample of humans. Cross-modal priming demonstrates that this baboon had formed amodal abstract concepts of the human and baboon categories. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Intermodal; Concept; Categorization; Picture perception; Monkey; Baboon

1. Introduction

It is well known that animals have the ability to group objects into categories. To mention just two of many examples, both pigeons and cebus monkeys discriminate human from non-human pictures, and generalize their behavior to previously unseen exemplars of the two classes [4,9]. What remains uncertain is the kind of cognitive structure that supports this ability [11]. Several authors argued that discrimination may reflect learning to attend one or several physical dimensions of the stimuli [12], prototype learning [1] or stimulus generalization from a few memorized exemplars [2,13]. Others argued that categorization transcends perceptual resemblance when dissimilar stimuli are associated to the same behavior (‘secondary stimulus generalization’ hypothesis [15]). Alternatively, categorization performance could be controlled by more conceptual mechanisms [10] relatively independent of the perceptual characteristics of the stimuli or previous associations to motor responses. In this paper, we report two cross-modal experiments suggesting the existence of abstract concepts in monkeys.

2. Experiment 1

2.1. Method

2.1.1. Subjects, apparatus and stimuli

Subjects were two 12-year old baboons (Papio papio) hereafter called B15 and B09. Baboons were familiar with the set-up and go/no-go testing procedure, but had never been trained nor tested on cross-modal tasks. Four keepers working with baboons were also tested. Keepers were selected as subjects because they were highly familiar with the baboon species and its vocalizations. The auditory stimuli consisted in human or baboon vocalizations. Human vocalizations comprised instances of two familiar French words (‘allez’ and ‘descends’) that are commonly pronounced during everyday experiments. A screaming call and a contact call comprised the baboon vocalizations. Sounds were approx. 0.8- to 1.5-s in duration, and 70 db. All vocalizations were stereo digital recordings of humans and baboons unknown by the subjects. The visual stimuli consisted in pictures of one or several humans or one or several baboons. Models on pictures differed in behavior and posture, resulting in an heterogeneous set of images. To equate backgrounds, both humans and baboons were photographed within an enclosure for ba-
boons. Pictures were $8 \times 6$ cm and sustained $8 \times 6$ degrees of visual angle. They had a $640 \times 480$ pixel definition and 256 possible colors.

Baboons were tested inside a cage facing two loudspeakers, a joystick, and a 14-in. color monitor driven by a Pentium 133 PC computer (for a more technical details, see [7]). Humans were tested at a table on which the monitor and the joystick were placed. For the two species, viewing distance was maintained constant (49 cm) by way of facemasks.

2.1.2. Testing procedure

Fig. 1 illustrates the go/no-go testing procedure. A test trial consisted in the following sequence: (1) manipulation of a joystick to point a cursor on a warning stimulus shown in the center of a computer screen; (2) 120 ms display of either a baboon or a human picture; (3) 170 ms delay; (4) display of either a human or a baboon vocalization; (5) 3-s delay during which go responses could be given by moving the joystick. With that procedure, auditory targets from each category were thus preceded by a prime image from either the same or different category. These two distinct types of trials will hereafter be referred to as the same category (S-trial) and different category trials (D-trial).

B15 and two humans were requested to move the joystick consecutive to a human vocalization (go response), and to refrain from moving it consecutive to a baboon vocalization (no-go response). The opposite rule was followed for B09 and the other humans. Baboons’ correct go or no-go responses were food reinforced. Both species received an auditory feedback after the trial, indicating if the response was correct or erroneous. Presentation of the prime image immediately after the warning stimulus had been pointed to assured that the image was perceived [16]. Note however that the task contained no explicit request for an attentional processing of the prime, responses being given considering the target sounds only, irrespective of the prime category. Both response times (RT) and scores were recorded. RTs were recorded from the onset of the auditory targets. The experiment consisted in three test sessions of 120 randomly ordered trials. Each session involved 60 different pictures (30 per category) and 120 different vocalizations (60 per category). Within a session, each prime image was presented once prior to a human vocalization and once prior to a baboon vocalization.
2.1.3. Preliminary training

Each subject received training sessions using the same procedure as in Fig. 1, except that sounds only were presented. Subjects firstly learned to categorize 60 different sounds (30 for the human and baboon category) until they performed 80% correct or more within a session. Learning auditory discrimination was laborious for the baboons and took several months. Humans achieved learning criterion in one training session only.

A transfer test was then proposed with 60 new sounds of the two categories. Humans showed perfect scores, and each baboon performed above chance for the human as well as for the baboon categories of sounds, $\chi^2$ tests, all $P < 0.01$. B09 and B15 achieved 77.2% and 79.4% correct on average for the two categories, and B15 achieved 88%. Positive transfers to novel sounds demonstrate the use of open-ended categorization rules.

With the aim of familiarizing the subjects with the priming procedure, 60-trial sessions involving the procedure illustrated in Fig. 1 were run prior to the test. These sessions employed 20 different landscapes devoid of humans or animals as primes. B09 and B15 needed 10 and 3 sessions, respectively, to achieve 80% or better in that last training phase. All humans performed above 80% accuracy at the first session.

2.2. Results and discussion

Each human and baboon performed significantly above chance with both human and baboon’s vocalizations, $\chi^2$ tests, all $P < 0.001$. B09 and B15 achieved 77.2% and 79.4% correct on average for the two categories, respectively, and humans performed from 98 to 99.7% correct. For each individual, correct RTs departing from the mean by two S.D.s or more (4.8% of the total) were discarded for statistical analyses, and the remaining RTs of each individual baboon were submitted to an analysis of variance (ANOVA) in which each target sound was introduced as the random factor, and the categorical consistency (S-trial vs. D-trial) and the test session (1, 2 or 3) served as the independent variables.

The analysis of B15’s data revealed a significant effect of categorical consistency. For this baboon, RTs were 35 ms faster on average when the prime and target were from the same category (i.e. S-trials) than when they were from different categories (D-trial), $F(1, 146) = 3.94, P < 0.05$. That difference is not accounted for by a speed accuracy trade-off, as revealed by a non significant difference between the number of errors in S- compared to D-trials (8.9% vs. 15.6%), $\chi^2 (1) = 0.23$, ns. Interestingly, a similar ANOVA conducted on false alarm trials ($N = 49$) revealed a trend for reduced RTs in S- compared to D-trials (difference = 128 ms), $F(1, 43) = 3.22, P < 0.08$. The main effect of session and the session by categorical consistency interaction was not significant for that analysis. As pictures of baboons were presented as primes in S- false alarm trials, reduced RTs in these trials demonstrate that the presentation of a baboon prime picture did not lengthen B15’s response behavior. This finding suggests that the effect of categorical consistency observed in correct go trials does not reflect an attentional perturbation induced by presentation of a baboon pictures.

Considering B09, no significant effect emerged on the percentage of errors (S-trials = 25.6%, D-trials = 20%, $\chi^2 (1) = 0.54$, ns). Similarly, the ANOVA on correct RTs failed to reveal any significant effect or interaction considering either the correct go-trials (S-trials = 1068 ms, D-trials = 1072 ms, $F(1,123) = 0.03$, ns), or the false alarm trials RTs (S-trials = 1072 ms, D-trials = 1007 ms, $F(1, 28) = 2.22$, ns).

One interesting question is to know if categorical priming in B15 reflects an interference in D-trials or a facilitation in S-trials. Our testing procedure provides no direct cue to answer that question, because there was no neutral condition during testing. Indirect evidence might be derived, however, from a comparison between mean RTs in the last two training sessions involving landscapes and mean RTs in the S- or D-trials of the test sessions. Inspection of these three means suggests facilitation rather than interference, because RTs differences between landscapes and D-trials was much smaller ($M = 7$ ms) than RT difference between landscape and S-trials ($M = 42$ ms).1

Humans also showed reduced RTs in S- compared to D-trials (Fig. 2). That effect emerged, for each individual human, when RTs were submitted to the same type of trial (S, D) by test session (s1, s2, s3) ANOVA as for baboons (all $P < 0.01$). Note that the main effect of test session, and the categorical consistency by test session interaction never reached significance level. Interestingly, the size of the effect for the group of four

---

1 Response time differences were not statistically tested because landscape-trials were proposed prior to the S- and D-trials, resulting in a possible effect of test order.
humans was of 41 ms, which is very similar to B15 in the same test (i.e. 35 ms).

In brief, Experiment 1 showed that categorization of vocalizations was affected by perception of conceptually related pictures. This effect was obtained in one baboon and in the four human subjects. Experiment 2 assessed the reliability of our findings. It verified if facilitation would still emerge in B15 when the prime images are degraded in several ways.

3. Experiment 2

3.1. Methods

That experiment used 160 novel vocalizations (80 per category) and 80 novel prime images (40 per category). For each category, the set of primes comprised 10 plain color images, 10 cutouts of the human or baboon models pasted on a uniform grey background, 10 plain grey-scale images, and 10 grey-scale cutouts of the models pasted on a uniform grey background. B15 received six test sessions of 160 randomly ordered trials. Each vocalization was presented once within a session, and each prime image was presented once prior to a human vocalization, and once prior to a baboon vocalization.

3.2. Results and discussion

B15 made only 19.5% of errors. ANOVAs were computed on scores and RTs with the prime type (color, color-cutout, grey-scale, grey-scale-cutout), testing session (s1, s2, s3, s4, s5, s6), and categorical consistency (S-trial vs. D-trial) serving as the unique three factors. There was no significant main effect or interaction for scores (all Ps > 0.1, ns). The analysis of RTs was restricted to correct go trials, because there were too few false alarm trials in each condition (i.e. from 9 to 16). That ANOVA showed a main effect of categorical consistency, \( F(1, 348) = 24.31, \ P < 0.001 \), showing shorter RTs in S- than in D-trials (mean difference = 47 ms). This effect of categorical consistency occurred with the four types of primes (Fig. 3) as revealed by a non significant main effect of prime type \( F(3, 348) = 1.43, \ ns \) and non-significant prime type by categorical consistency interaction. There was also a main effect of test session \( F(5, 348) = 8.53, \ P < 0.001 \), showing a progressive lengthening of RTs with repeated sessions \( (M = 741, 759, 807, 816, 804, \ and \ 833 \ ms \ for \ session \ 1--6, \ respectively) \).

4. General discussion

Our two experiments demonstrated that presentation of a prime image affected the processing of conceptually related sounds in one baboon (B15) and in humans. This result shows that neither B15 nor the humans exclusively processed the perceptual attributes of the stimuli. Clearly, cross-modal effects cannot be accounted for by the feature [12], prototype [1] or exemplar theories [13], which posit that categorization is exclusively controlled by the perceptual characteristics of the stimuli. Alternatively, our results cannot be explained to satisfaction by the ‘secondary stimulus generalization’ account [15]: there was no common response associated to the pictures and the sounds prior to the experiment, the pictures being presented for the first time during the test. Moreover, reward delivery was not correlated with prime category, which prevented learning of prime-target associations on the basis of reinforcement contingencies.

Cognitive processes at a higher level might explain cross-modal facilitation. We propose that both B15 and the humans developed amodal abstract concepts of the human and baboon categories, and referred to these concepts while responding in the task. In our experiments, differences between S- and D-trials emerged because the prime and target stimuli were conceptually equivalent in S-trials, and conceptually different in D-trials. Importantly, subjects did not receive any advance training on cross-modal tasks prior to the test reported here. Therefore, conceptual representations of the two categories cannot be products of experimental manipulation, and were probably already built-in at the beginning of the research.

Previous studies showed cross-modal integration in monkeys or apes see [3] for an example. In apes, cross-modal matching was demonstrated with concrete objects [5,8] and symbols [14]. All these studies employed well-identified social or non-social objects that
had to be recognized across modalities. The current study also adds a new finding to this literature: cross-modal interference might also occur across classes of objects, not only specific objects.

Of course, one important limitation of this study is that conceptual categorization was found in only one baboon. Why B09 did not show this ability remains unclear. Failure to demonstrate cross-modal facilitation in this animal does not necessarily indicate a complete lack of conceptualization. It is possible, for instance, that this baboon did not recognize humans or baboons on pictures, as suggested by recent publications on picture/object equivalence [6]. Alternatively, difficulties for cross-modal integration could also account for the B09’s performance. Further experiments will be necessary to identify the contribution of these factors in B09. Because cross-modal effects were consistently observed in B15 and in humans, we presume that conceptual categorization is possible in a variety of individuals and species, and will be revealed more systematically if tested by adequate cross-modal categorical procedures.

Acknowledgements

We thank B. Arnaud, R. Fayolle and M. Chiambretto for technical assistance.

References