Fagot, J. & Parron, J.(2010). Relational Matching in Baboons (Papio papio) with Reduced Grouping Requirements. Journal of Experimental Psychology : Animal Behaviour Processes, 36, 184-193

Relational Matching in Baboons (Papio papio) with Reduced Grouping Requirements

Joël Fagot and Carole Parron

CNRS

Institut de Neurosciences Cognitives de la Méditerranée et
Université de la Méditerranée, Marseille, France

Corresponding author : Dr. Joël Fagot, INCM, 31 ch Joseph Aiguier 13402 Marseille cedex 20, France.

Email : Fagot@incm.cnrs-mrs.fr. Phone: 33 (0)4 91 16 43 06

Abstract

Analogical reasoning is a corner stone of human cognition with unknown phylogenetic origins. Recent animal studies have suggested that only apes can solve the 2- by 2-items relational matching (RMTS) analogy problem, with potential benefits of language- (Premack, 1983) or token-training procedures (Thompson, Oden, & Boysen, 1997). In this study, six baboons were initially trained in a RMTS task in which the SAME and DIFFERENT relations were exemplified by compound stimuli made of two adjacent patches of colors. Learning occurred in this task with a first set of colors, and transferred in probe trials with new colors (Experiment 1). Manipulation of the size of the sample or comparison stimuli (Experiment 2) showed that the performance was not controlled by strict consideration of the surface of the color patches, suggesting cognitive flexibility. Performance deteriorated to chance level when a gap was introduced in between the two elemental features composing the SAME or DIFFERENT displays (Experiment 3), but this effect of gap size could be overcome by training (Experiment 4). It is suggested that monkeys share with humans and apes the ability to judge relations between relations, even in absence of language- or token training, but that this ability was previously hidden by local mode of processing limiting consideration of the stimuli as pairs in the RMTS task, rather than as independent objects.

Relational Matching in Baboons (Papio papio) with Reduced Grouping Requirements

An important challenge of comparative cognition is to track the evolution of the cognitive functions, for a better understanding of what makes us humans. In that research effort, comparative psychologists are very interested by tasks influenced by human as language, to understand what cognition can be in absence of language. One good example of such tasks is the analogy problem called “relational matching”. Relational matching problems can be illustrated as follows: given one first set of two stimuli, say A and B, which of the following two comparison pairs, say CC or DE, match the sample? Because there is no perceptual similarity between the sample and comparisons, this problem cannot be solved on perceptual cues. Solving this relational matching to sample task (RMTS) both requires conceptualization of the same and different abstract relations, and the ability to form abstract equivalence judgments about relations between relations. Hence, DE goes with AB in our example because both pairs illustrate the abstract relation of “differentness”. A great deal of studies supports the claim that this form of relational thinking is in humans promoted by language competencies (e.g., Gentner & Christie, 2008)

The RMTS task described above was firstly used in the comparative literature to study analogical reasoning in chimpanzees (Premack, 1983). Premack reported that language trained chimpanzees could pass the task, while language naive chimpanzees could not. From these results and others from related analogy problems (Gillan, Premack, & Woodruff, 1981), Premack (1983) claimed that responsiveness to a relation between relations requires the use of an 'abstract code' which can only be provided by language training. Thus, for Premack, only humans beyond infancy and chimpanzees with language expertise should be able to solve this kind of analogy problem.

To further assess the contribution of “language” to relational thought, Thompson, Oden & Boysen (1997) compared the RMTS performance of a language trained chimpanzee (i.e., Sarah) to that of three chimpanzees with a history of token training but no language training *per se*, in addition to a fifth naïve chimpanzee. Language and token trained chimpanzees achieved similar performance, therefore ruling out the idea that language training is mandatory for successful RMTS performance. Thompson and collaborators argued from these results that token training provides the necessary representational units to symbolize the SAME and DIFFERENT relations, allowing the animal to re-encode the task as a simple matching to sample task using the symbols as entries (Thompson et al., 1997).

In a different study, Thompson and collaborators also asked language and token naïve infant chimpanzees to handle pairs of SAME or DIFFERENT objects mounted together on a display board (Oden, Thompson, & Premack, 1990). After this familiarization period, they were proposed a new pair of objects, instantiating the same identity/nonidentity relation as the initial pair. The second pair was manipulated less when it showed the same relation as the first one, but habituation did not occur when the two pairs showed the opposite relations. Interestingly, these same chimpanzees remained unable to more explicitly judge the equivalence of these relations in the RMTS tasks (Oden et al., 1990; Thompson & Oden, 1996). From the authors, infant chimpanzees can perfectly detect, code, and store abstract identity relations, and are even implicitly sensitive to relations between relations, but this ability would become overtly functional only after language (Premack, 1983) or token training (Thompson et al., 1997).

Vonk (2003) recently reconsidered the role of token training in a study using one gorilla and four orangutans, all naïve with either token or language training procedures. In Vonk’s (2003) study, the apes were requested to match pairs of colored geometrical shapes considering the SAME or DIFFERENT relations instantiated by each pair. Two stimulus dimensions were manipulated in the task (the shape and color of each stimulus), and the participants were asked to matched SAME/DIFFERENT stimulus pairs when the number of shared dimensions between the sample and comparisons was manipulated. Although perceptual features contributed to matching performance, the gorilla and three of the four orangutans remained successful in the most abstract version of the task (Vonk, 2003). That result suggests that language or token training might not be a necessary condition for successful RMTS performance. However, Flemming and co-authors questioned the abstract nature of the process used by the successful apes in Vonk (2003). For them, the apes may have selected the alternative that was perceptually less similar to the sample, with limited consideration of the relations between relations (Flemming, Beran, Thompson, Kleider, & Washburn, 2008). The exact contribution of language of token training procedures to RMTS performance remains to be elucidated, which warrants further studies on this issue.

In a comparative perspective, it is now clearly established that monkeys and pigeons share with ape the capacity to perceive SAME/DIFFERENT relations. Accurate SAME/DIFFERENT discriminations were repeatedly obtained in tasks requiring consideration a variety of perceptual cues (Blaisdell & Cook, 2005; Cook, Kelly, & Katz, 2003; Neiworth & Wright, 1994; Sands, Lincoln, & Wright, 1982; Wasserman, Young, & Fagot, 2001; Young & Wasserman, 1997). They were also found in a more conceptual task in which baboons had to judge if two objects of a pair belong to the same functional (food vs. non food) category (Bovet & Vauclair, 2001). Capuchin monkeys can moreover identify the positive cup among a set of three cups of different sizes, considering the relative size of the positive stimulus within a first set of cups (Kennedy & Fragaszy, 2008). In spite of these remarkable achievements, there is to our knowledge no evidence that a non-ape animal species can solve Premack’s (1983) second order 2- by 2-item RMTS task, suggesting a profound disparity between apes and non apes animals in conceptualization power (Fagot, Wasserman, & Young, 2001; Flemming, Beran, & Washburn, 2007; Thompson & Oden, 1996, 2000). The most successful RMTS performance was obtained by Wasserman and collaborators, who reported that baboons could match 2-items displays with 16-items arrays (Fagot et al., 2001), and pigeons could match 16-item arrays with 16-items (SAME or DIFFERENT) arrays (Cook & Wasserman, 2007). However, explicit 2- by 2-item RMTS tests led to chance performance (Fagot et al., 2001; Flemming et al., 2007). In addition, and unlike infant chimpanzees, use of a familiarization/novelty handling procedure failed to reveal an implicit sensitivity to the relational (identity/nonidentity) properties of the stimulus pairs (Thompson & Oden, 1996).

Why pigeons and monkeys can only solve the RMTS with multiple-item arrays remains puzzling. One the one hand, it could be considered that due to limited conceptual abilities, non-ape species may require redundant illustrations of the SAME and DIFFERENT relations to perceive them (Wasserman, Young, & Cook, 2004). One the other hand, multi-element displays may provide extra perceptual cues for categorization which are unavailable in 2-item displays (Flemming et al., 2007). Previous attempts to identify the contribution of perceptual cues to matching performance have shown that the processing of multi-element arrays involve an analysis of the high spatial frequencies of the icons (Cook & Wasserman, 2006; Wasserman, Young et al., 2001), but is relatively independent of their orientation (Cook & Wasserman, 2007); spatial alignment (Wasserman, Fagot, & Young, 2001) or size (Cook & Wasserman, 2007). However, performance positively correlates with the entropy of the arrays of icons (Fagot et al., 2001; Wasserman, Fagot, & Young, 2000; Young, Wasserman, & Garner, 1997). According to Flemming et al. (2007), that correlation with entropy suggests that monkeys consider the perceptual variance of the stimuli rather than the abstract identity-nonidentity relations. In sum, the comparative literature reveals a clear disconnect between humans and apes on one side, and the other animal species on the other side, regarding the ability to solve the 2- by 2-item RMTS task, but the origin of this difference between primate groups remains highly uncertain.

In this context, the present research was aimed at further exploring possible perceptual influences on SAME/DIFFERENT conceptualization in monkeys. Its main originality is to relate two literatures which have been so far considered as independent, namely the literature on analogical reasoning reported above and that recently growing literature on selective attention and global/local processing in animals.

When tested with large geometrical shapes made of smaller ones, baboons (Deruelle & Fagot, 1998; Fagot & Deruelle, 1997; Parron & Fagot, 2007), macaques (Hopkins & Washburn, 2002), capuchins (De Lillo, Spinozzi, Truppa, & Naylor, 2005; Spinozzi, De Lillo, & Salvi, 2006; Spinozzi, De Lillo, & Truppa, 2003) and pigeons (Cavoto & Cook, 2001) tend to process compound stimuli locally, which is in sharp contrast which the global processing mode traditionally found in our western culture (Navon, 1977). In monkeys, the processing of the global structure of such stimuli is controlled by the distance separating the local elements (Deruelle & Fagot, 1998; Spinozzi et al., 2006). In addition, when chimpanzees are compared to either baboons or macaques, using the same stimuli, they show less sensitivity to element separation than monkeys (Fagot & Tomonaga, 1999; Hopkins & Washburn, 2002).

Being able to solve the 2- by 2-item RMTS task requires a process by which the visual displays is mentally organized as a series of pairs of objects serving as cognitive units, each pair illustrating a SAME or DIFFERENT relation. The comparative literature on global local processing suggests that monkeys and pigeons may both focus their attention so strongly on the individual objects, due to element separation, that they would process the 2-item as sets of conceptually independent objects, rather than as pairs. We suggest that this hypothesis can explain the inability of monkeys (Fagot et al., 2001; Flemming et al., 2007) to process the relation between relations in the 2- by 2-item RMTS task. It can also explain positive effect of icons redundancy in multi-item displays, considering that the high density of icons may favor grouping processes, and therefore the processing of these displays as SAME or DIFFERENT cognitive units.

Four experiments will be presented below. Experiment 1 will demonstrate successful 2- by 2-item RMTS performance in baboons, when each stimulus pair is made by adjacent elements. To our knowledge, this is the first demonstration of successful 2- by 2-item relational matching in a non-ape species. Experiment 2 will further demonstrate that baboons' relational matching performance with adjacent pairs survives manipulation of the size of the stimulus. Experiment 3 will demonstrate that their performance deteriorates drastically when a small separation is introduced between the 2-item sample displays. Experiment 4 will finally show that this effect of gap size can be overcome with training. From these results, we will argue that monkeys have (at least rudimentary) skills for relational matching, but that these skills were masked in earlier studies by a too high demand in terms of grouping.

Experiment 1

This first experiment proposed a 2- by 2-item RMTS task to baboons, in which the SAME or DIFFERENT stimuli consisted in adjacent color patches. It was hypothesis that the use of adjacent stimuli will alleviate the need for grouping, therefore favoring SAME/DIFFERENT relational matching performance.

Method

Participants. They were 4 males (i.e., B03, B05, B07, B09) and 2 females (B06, B08) 20-years-old Guinea baboons (Papio papio) living in large (6 X 4 m) indoor/outdoor enclosures within the C.N.R.S (Marseille) facility. These baboons have a long experimental history. They have already been tested in a variety of computerized tasks using a joystick, including relational matching experiments (Fagot et al., 2001; Wasserman et al., 2000; Wasserman, Young et al., 2001). They also received two previous (unpublished) experiments using the same automatic touch-screen operant system as used here. Each baboon had a 2.1 X 8 mm subcutaneous glass tag implanted in each forearm for automatic identification during testing.

Apparatus. The test employed an automatic operant conditioning system, called ALDM (Fagot, submitted), provided *ad libitum* in the outdoor enclosure of the baboons where they lived in social groups. The main particularity of this learning device is to identify the subjects automatically once they enter the test system, allowing self testing on a voluntary basis, while maintained in a social group, and on a 24h/day schedule. ALDM comprised a freely accessible test chamber (.7 X .7 X .8 m) which rear side was maintained opened. The test chamber was fitted in its innermost front side with a (7 X 7 cm) view port and two hand ports (8 X 5 cm). Looking through the view port allowed vision of a 17 inches LCD touch monitor installed at eye level 25 cm from the view port. Introducing one hand through one hand port permitted actions on the touch screen. Two antennas fixed around each view port automatically read the ID number of each participant when it introduced its forearm through the view port. Numeric identification signals from the arm tags served to trigger the computer controlled stimulus, presentation and to assign behavioral measures (stimulus choices and response times) to each participant. The test equipments other than the screen were concealed from view and inaccessible to the experimental subject. Because the test system was made up with waterproof opaque Perspex material, stimuli could be presented in well standardized conditions, irrespective of external weather fluctuations and lighting.

Grains of dry wheat served as rewards. They were delivered inside the test booth by a homemade food dispenser. The research was controlled by a test program developed by the first author with Eprime language (V 1.2, Psychology Software Tools, Pittsburg, US). The main feature of this program is to allow independent tests regimen for each baboon, irrespective of the order in which baboons presented themselves in the test booth. To prevent social inhibition from the dominant subject of the group, a time out of 15 to 30 minutes was imposed to each subject after each training or testing session.

Stimuli. The SAME or DIFFERENT stimuli consisted in two adjacent (200 X 100 pixels) patches of colors, as illustrated in Figure 1. The two patches were displayed with the same color or with different colors, in order to be representative of the SAME or DIFFERENT category. A detectable two pixel wide black line delineated the adjacent border of the two color patches, to prevent viewing the SAME stimuli as a single large stimulus rather than as a stimulus made in two parts. Ten colors presumed to be perceptually different (e.g., brown, green, light red, yellow, etc..) for this trichomate species were used to create the training stimuli. Ten new distinct colors served for the test stimuli. There was no control of color brightness to provide a maximum number of cues for stimulus classification.

Training procedure. The screen turned grey at the onset of each trial, in order to provide a cue indicating that the program is ready for the self identification procedure. Each trial started when the baboon introduced one hand in the hand port to identify itself. That action triggered the test trial assigned to that baboon and the immediate display of a (3 X 3 cm) yellow fixation cross on the bottom part of the screen. Touching the fixation stimulus induced the central presentation of a compound sample stimulus from either the SAME or DIFFERENT category. The baboon was requested to touch the sample stimulus in response to this display. That action triggered the immediate presentation of two comparison stimuli, one in each hemi-screen, with 400 pixels separating them. One of the comparison stimulus was different from the sample, but of the same category. That stimulus will be hereafter considered as S+. The other one (i.e., S-) was from the other category. The task was to indicate by a hand touch the comparison S+ stimuli showing the same abstract (SAME or DIFFERENT) relation as the sample.

The colors used to draw the stimuli were randomly selected on a trial to trial basis from the first set of 10 training colors, with the constraint that there was not a single color shared by the sample, S+ or S-. This aspect of the procedure promotes consideration of the relational dimension of the task, rather than perceptually based identity matching. Note that repetition of the displays was unlikely. Up to 20160 distinct SAME and 241920 distinct DIFFERENT configurations can be created from the initial set of 10 colors (see note 1). These large numbers of possibilities make strategies based on the rote learning of specific configurations of color likely inefficient. Correct responses were indicated by a 3 second green screen, a high frequency tone, and were all food rewarded. Incorrect responses were indicated by a 3 second time out during which the screen turned green, a low frequency tone, and were never food rewarded. A correction procedure was also adopted: After an error, the next trial systematically showed a sample stimulus selected from the same category as in the error trial, but made from a new random selection of its constituting colors. A maximum of three correction trials were presented in a succession, if errors perseverated. The screen systematically turned black for an inter-trial interval of 6 seconds after each trial. It then turned back grey, indicating that the subject can now process the next self identification phase, and thus the next trial. All training stimuli were presented on a black background.

The training phase involved repetition of series of 100 randomly ordered trials (correction trials excluded), which were completely balanced considering the (SAME or DIFFERENT) sample category, and the location of S+ in the left or right hemi-screen. Trials within a session were presented in random order, and sessions were repeated until a criterion of 80% correct or more was achieved in two consecutive sessions. The program automatically switched the successful baboons to the test program, once that training criterion was reached.

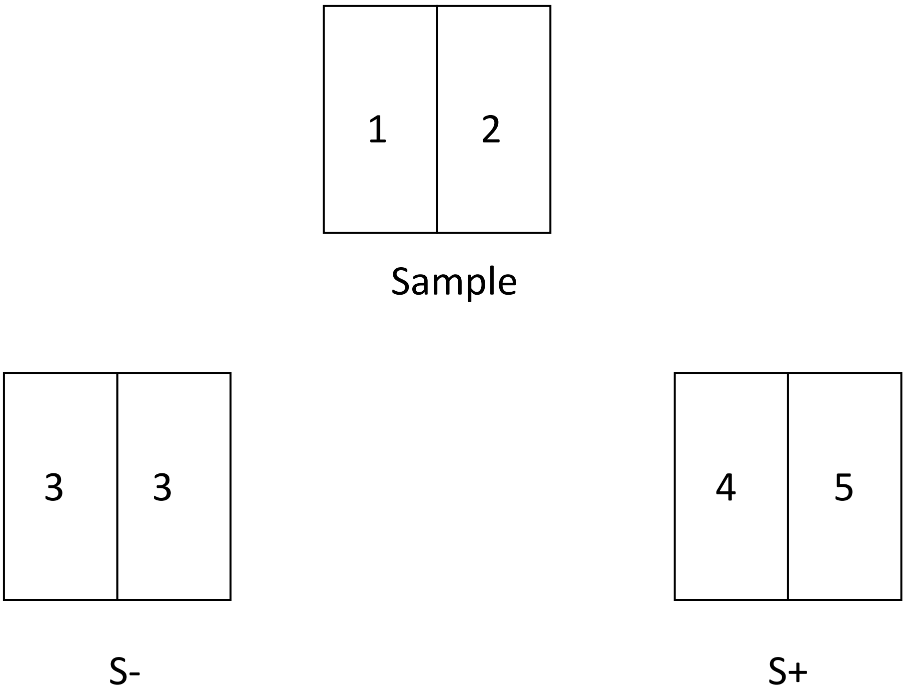

Figure 1: Stimulus configuration used in Experiment 1. Colors are indicated by numbers for illustrative purposes. The trial illustrated here is a DIFFERENT trial during which the baboon to select the stimulus comparison (S+) showing the same DIFFERENT relation as the compound sample. The tasks is a sequential matching task, the sample is presented first, in the middle of the screen, immediately followed by the displays of the two comparison forms.

Testing. It involved three consecutive sessions of 128 randomly ordered trials. These sessions consisted in 80 baseline trials, identical to the training trials described above, mixed with 48 probe trials using the second set of 10 new colors. Probe trials were completely balanced considering the sample category (SAME or DIFFERENT) and left/right location of S+. They were randomly reinforced on an 80% basis to match the percentage of rewards obtained at the end of training, after the baboons have reached a learning criterion of 80% correct. No correction procedure was used for the probe trials. The program automatically switched the subject to an unrelated two choice discrimination task, once testing was completed with the desired baboons.

Results

A total of 37412 training trials were run for the group; 34.7% of them corresponding to the correction trials. Figure 2 depicts the learning curves, as computed without the correction trials. All baboons but one (i.e., B07) learned the task to criterion. Learning was achieved in 19 (B03), 39 (B05), 64 (B06); 42 (B08) and 47 (B09) 100 trial sessions (mean = 42.2 sessions; SD = 16.2%). B07 could only maintain a ceiling performance in the 65-70% correct range, but lost interest in the task after 38 sessions.

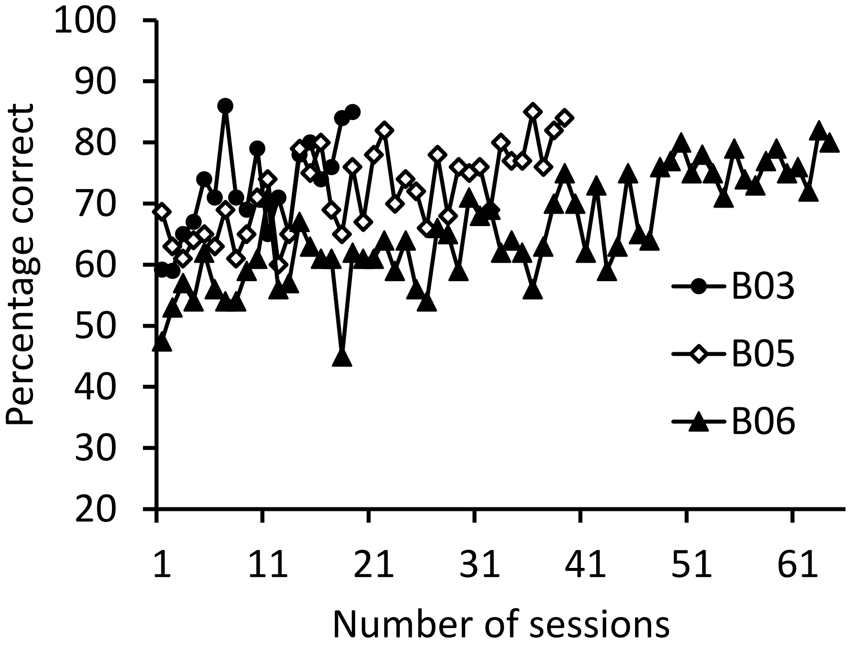

Figure 2a

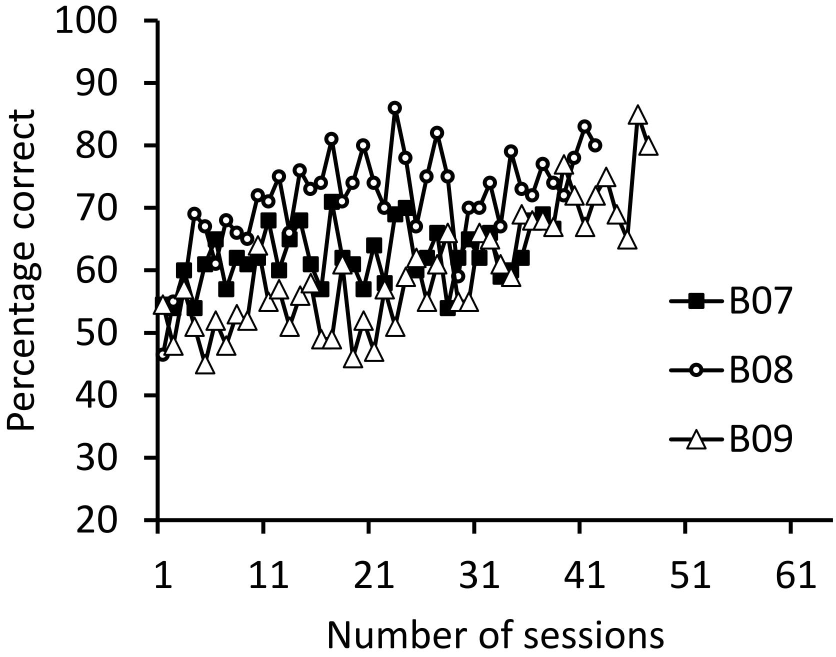

Figure 2b

Figure 2: Learning curves in Experiment 1.

Figure 3 illustrates performance in probe trials for all baboons, except B07 who never reached the learning criterion. A first one-way analysis of variance (ANOVA) verified if average performance depended on the sample category (SAME or DIFFERENT). Average performance for SAME trials did not differ from that of the DIFFERENT trials (mean SAME = 75.8%; mean DIFFERENT = 77.3 %; F(1,4)= .072, p>.05). The next analyses verified if individual performance in probe trials, summed across categories, exceeded chance level (i.e., 50% correct). It showed that all baboons were reliably above chance in probe trials (one-tailed binomial tests, all ps < .05).

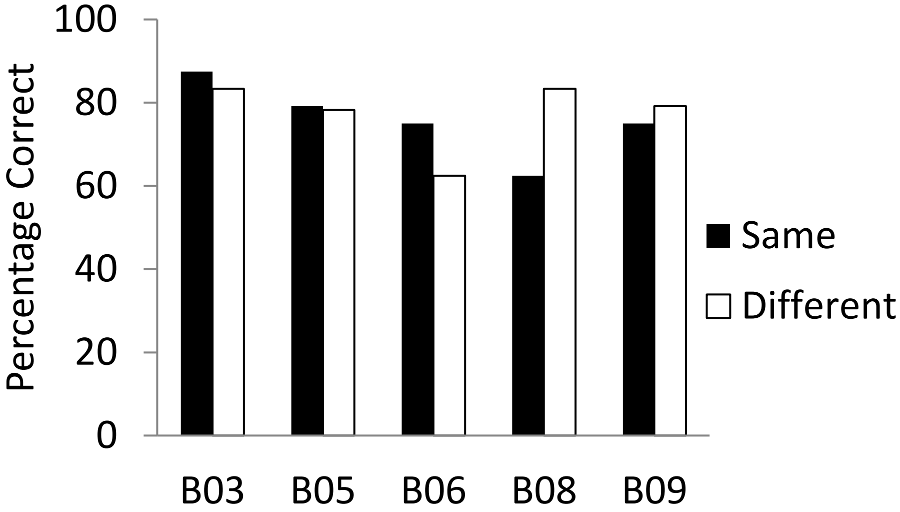

Figure 3: Individual performance in probe trials. Individual performance was systematically above chance, as revealed by one-tailed binomial tests (p<.05)

Discussion

Successful RMTS performance was obtained in Experiment 1, when the abstract relations to be matched were illustrated by compound stimuli made of adjacent patches of colors. Although that performance suggests consideration of SAME/DIFFERENT relations, it might be proposed that the baboons used a perceptual strategy that considered the size of the color patches as a cue (for an example of responses controlled by stimulus size, see Peissig, Kirkpatrick, Young, Wasserman, & Biederman, 2006) Under this hypothesis, the baboons would follow that kind of strategies: “if the sample shows a big patch of identical color (i.e., for the same stimuli), then select the comparison showing a similarly big patch, if it showed two small patches, then select the comparison stimuli showing two small patches”. This hypothesis was tested in Experiment 2, using sample and comparison compound stimuli of different sizes.

Experiment 2

Experiment 2 verified in which extend baboons' performance in the 2-item RMTS task depends on the total surface area occupied by the two identical color patches in SAME trials, or by the elemental color patches in the DIFFERENT trials.

Methods

Participants. One of our baboons (B09) suddenly died from sickness. The group of participants therefore consisted in the four remaining baboons which were successful in Experiment 1.

Test procedure. The experiment involved two test phases which used the stimulus configuration illustrated in Figure 4a (test phase 1) and 4b (test phase 2). In test phase 1, the SAME and DIFFERENT sample stimuli were made of two adjacent 50 X 200 pixel color patches. Each sample patch was therefore reduced in width by 50% in comparison to Experiment 1. The comparison stimuli remained unchanged in that phase, and therefore used 100 X 200 pixel color patches. Test phase 2 used comparison S+ and S- stimuli reduced by 50% in width, and sample stimuli of the same size as in Experiment 1.

Test phases 1 and 2 followed an identical design. They both consisted in three sessions of 96 trials, including 80 baseline trials (as in Experiment 1) mixed with 16 probe randomly reinforced trials (at a 80% rate) with a 50% size reduction of the sample (test phase 1) or comparison stimuli (test phase 2). Probe trials within a session were fully balanced considering the variables of stimulus category (SAME, DIFFERENT) and location of S+ (left, right hemi-screen). For all baboons, test phase 1 was run prior to test phase 2.

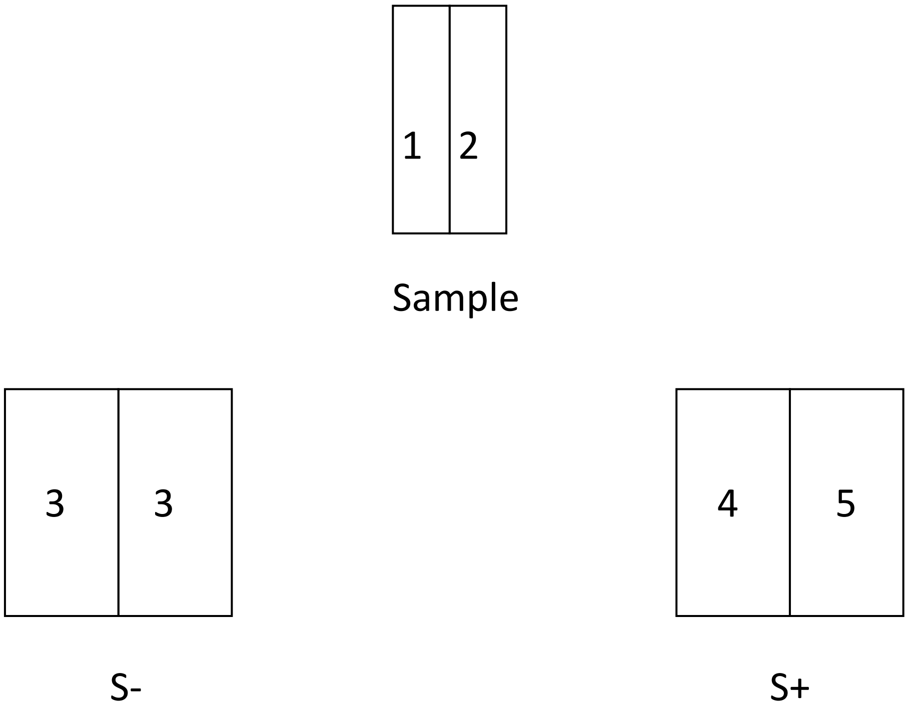

Figure 4a

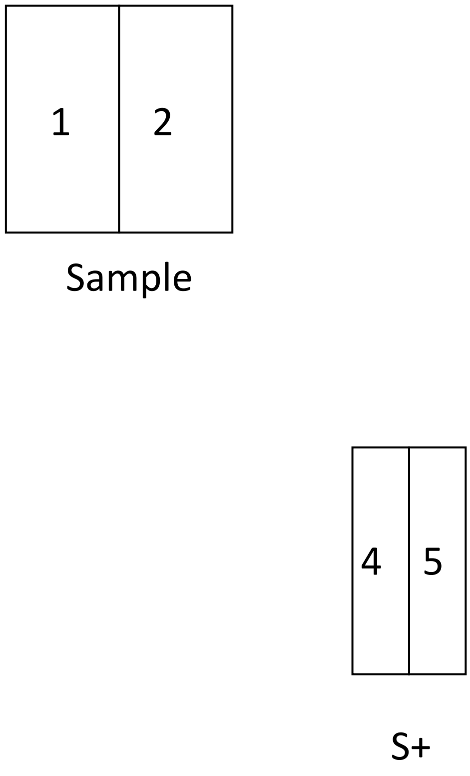

Figure 4b

Figure 4. Illustration of the test procedure adopted in the first (Figure 4a) and second (Figure 4b) test phase of Experiment 2.

Results and discussion

Scores were submitted to a Test Phase (phase 1, phase 2) by Sample Category (SAME, DIFFERENT) ANOVA. Performance was roughly similar in the two test phases (mean phase 1 = 62.5% correct; mean phase 2 = 61.5%), leading to a non significant effect of test phase, F(1-3) = .07, p>.05. Performance for SAME trials was numerically larger (mean=69.3%) than for DIFFERENT trials (mean 54.7%), but this difference was not significant, F(1-3) = 6.26, p>.05. Lack of signification is due to the fact that category differences were mostly attributable to B03 and B05 (B03: mean SAME = 72.9%; mean DIFFERENT = 54.2%; B05: SAME = 72.9; DIFFERENT = 43.7%; B06: SAME = 60.4%; DIFFERENT = 66.7%; B08: SAME = 64.6%; DIFFERENT = 60.4%). Finally, the Test Phase by Trial Category interaction was not significant, F(1-3) <.1; p>.05.

As there were no reliable effects of test phase, performance of each baboon was summed across test phases, to increase sample size, and compared to chance level by one-tailed binomial test (p<.05). One tailed tests were justified by unidirectional predictions. These tests showed that B03, B06 and B08 were all reliably above chance in Experiment 2 (see Figure 5). This result suggests positive transfers across stimulus sizes, and thus the processing of the abstract properties of the displays. Note however that the performance declined in comparison to Experiment 1, suggesting that the performance of the baboons was also controlled by perceptual factors related to stimulus size.

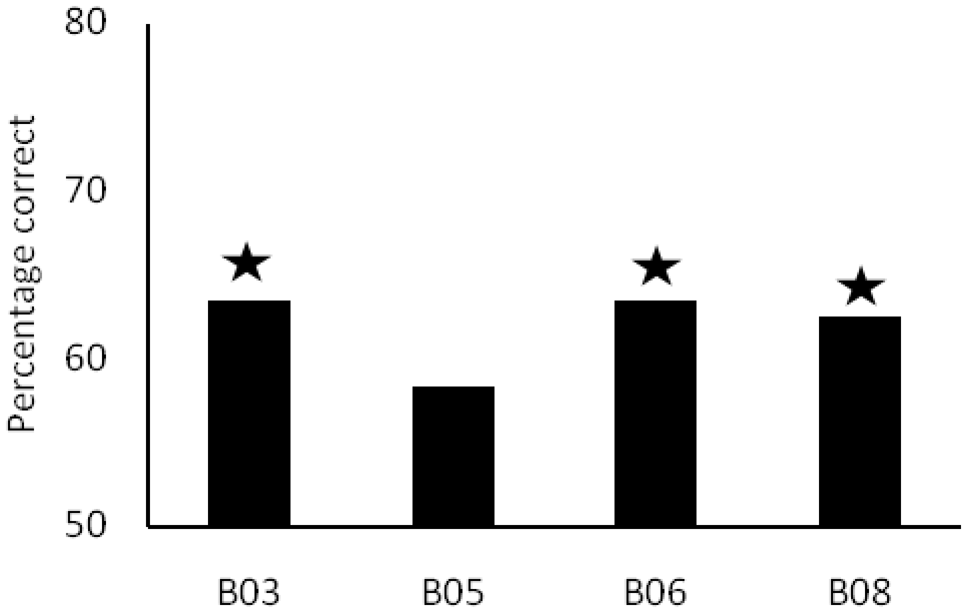

Figure 5: Average transfer performance in Experiment 2 and for each baboon. Stars indicate an above chance performance inferred from a one-tailed binomial test (p<.05).

B03 exhibited an asymmetrical performance in favor of the SAME trials (Same: 72.9% correct; DIFFERENT: 54.2%) trials. Because of that asymmetry in responding, its performance in the transfer tests is hardly amenable to clear-cut interpretations. Results are however more clear-cut for B06 and B08; who expressed an above chance performance in absence of a clear bias favoring the SAME or DIFFERENT category trials. For B06 and B08 at least, findings suggest that the relational matching rules acquired with SAME/DIFFERENT stimuli of the same size can transcend size variations.

Experiment 3

The next experiment assessed the RMTS performance obtained in Experiments 1-2 would survive when spatial gaps are introduced in between the two elemental color patches composing the sample configuration.

Methods

Participants and test apparatus. They were the same as in Experiment 2.

Stimuli. Two different types of stimuli were used. The first type consisted in two adjacent (100 X 200) color patches of either the same or different colors, following the same principle as in Experiment 1. The second type of stimuli was made with two non-adjacent (100 X 200 pixels) patches of colors, separated by a gap of either 10, 30 or 60 pixels (see Figure 6).

Test procedure. The baboons firstly received training sessions identical to those of Experiment 1, until they reached 80% correct in one training session. They were then automatically tested in 3 sessions of 128 randomly ordered trials (80 baseline trials mixed with 48 probe trials). Baseline trials in the test sessions used adjacent stimuli with a minimal gap of 2 pixels. They were thus the same as in training. Probe trials used a novel configuration of stimuli illustrated in Figure 6. The sample stimuli in the probe trials consisted in two (100 X 200 pixels) color patches which were now separated by 10, 30 or 60 pixels, while the comparison stimuli continued to be made of two adjacent color patches similar to those of baseline trials. Probe trials were fully balanced regarding sample gap size (10, 30 or 60 pixels), stimulus sample (SAME or DIFFERENT) category, and the (left or right) location of S+. All colors were randomly selected from color set 1, with the constraint that colors were not shared by the sample, S+ and S- stimuli. The other aspects of the procedure were as in Experiment 1.

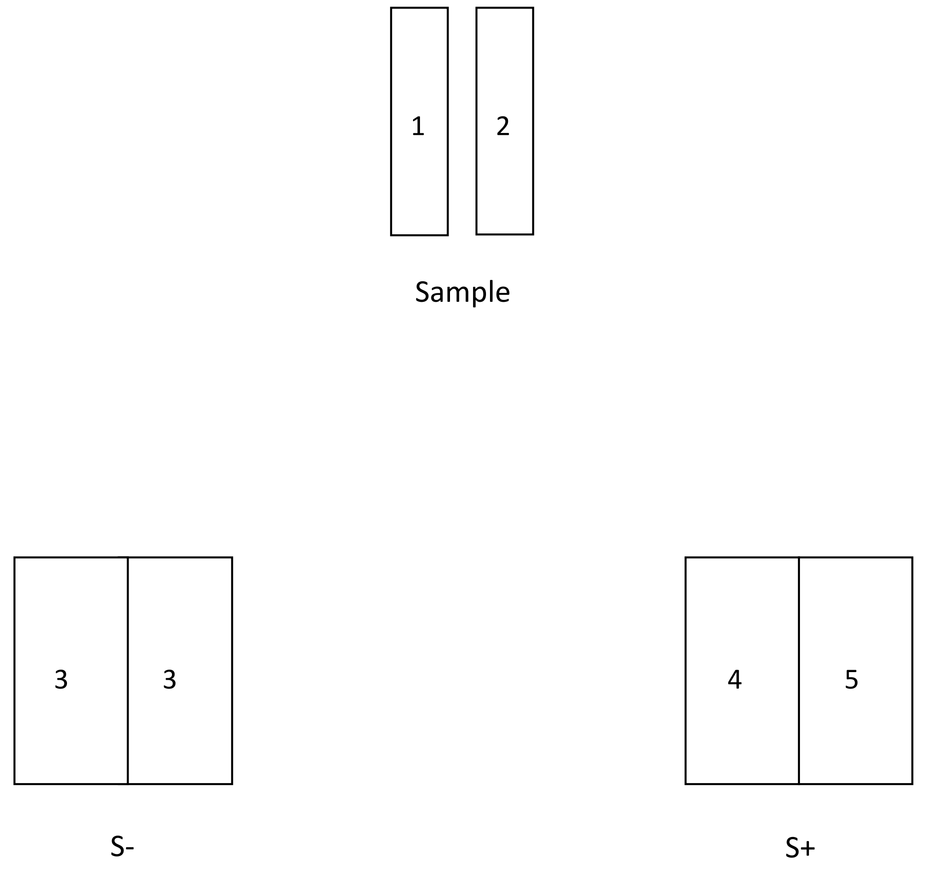

Figure 6. Illustration of the test procedure of Experiment 3. Note the gap separating the two color patches composing the sample stimuli. During testing, gap size were of 2 (baseline trials), 10, 30 or 60 pixels.

Results and discussion

Scores were analyzed with a two-way Sample Category (SAME, DIFFRERENT) by Gap Size (0, 10, 30, 60) ANOVA. Performance strongly decreased with gap size, (mean 2 pixels = 80%; mean 10 = 60,4%; mean 30 = 55,2%; mean 60 = 48,9%, F(3-9)=29.05, p<.0.001 (see Figure 7), but neither the main effect of Sample Category, F(1,3) = 0.21, p>.05, nor the Sample Category by Gap size interaction were significant, F(3,9)=0.83, p>.05. Post-hoc Tukey tests (p<.05) showed that performance with 2 pixel gaps (80% correct) was greater than performance in the other conditions of gap size. In addition, performance with a gap of 10

(60,42%) pixels exceeded that obtained with 60 pixels, but the 30 and 60 pixels conditions did not differ from each other (55,2% and 48,96%, respectively).

We finally verified if performance of each individual for each gap size were above chance (one-tailed binomial test, p<.05). All participants demonstrated an above chance performance with the quasi-adjacent (2 pixel gap) sample used in baseline trials (see Figure 7). For the other conditions of Gap Size, B03 and B05 could maintain a modest but nevertheless above chance performance with the smallest (10 pixels) gap size. By contrast, all participants were at chance in the two conditions using the largest (i.e., 30 and 60 pixels) gaps. The results are thus clear: The relational matching performance strongly deteriorated when a gap was introduced in between the two elemental elements composing the sample stimulus.

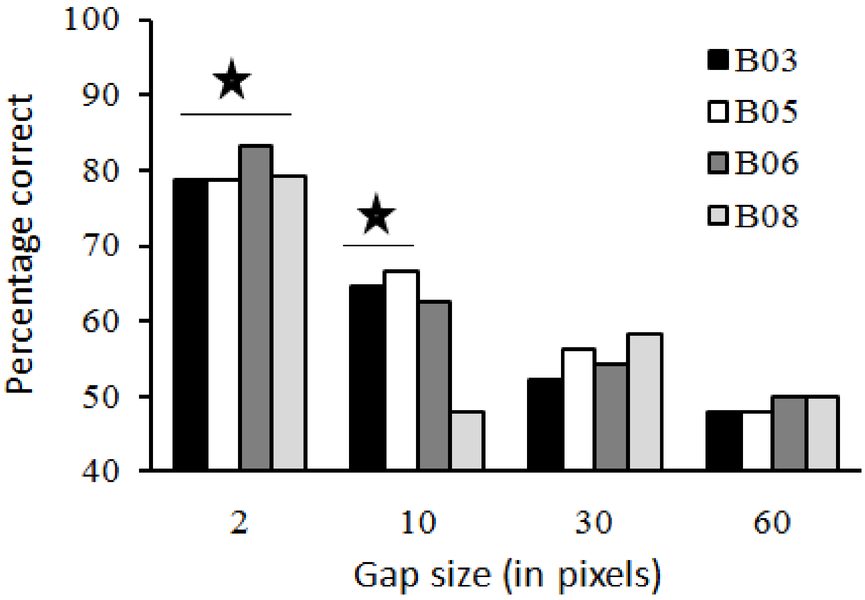

Figure 7. Performance in the probe trials of Experiment 3, as a function of gap size (in pixels). Stars indicate an above chance performance (one-tailed binomial tests, p<.05).

Experiment 4

Experiment 4 verified if the drastic effect of gap size in Experiment 3 could disappear with training. In Experiment 4, we progressively increased the size of the gap separating the elemental stimuli of both the sample and comparison forms. Findings showed that baboons could finally master the 2 X 2 RMTS task with high performance, in absence of token training, and although SAME and DIFFERENT pairs were now composed of spatially distinct elements.

Methods

Participants and test apparatus. B05, B06, B07 and B09 participated to the task on a voluntary basis. No data could be collected for B03 who declined testing.

Stimuli. They consisted in pairs of 100 X 200 pixels color patches made up with the same sets of colors as in Experiment 1.

Training and test procedures. There were two training phases. These two phases served to increase the gap size of the sample (training phase 1) and comparison pairs (training phase 2) in a stepwise manner. Training phases 1 and 2 followed the same training procedure as in the training phase of Experiment 1. However, sample gap size in phase 1 was incremented by 2 pixels when the subject reached a reliable (binomial test, $p<.05$) performance of 75% correct or more in the previous sessions. In that phase, the gap for the comparison pairs remained at a fixed value of 2 pixels. Sample gap size was initially set at 2 pixels in the first training session, and training phase 1 ended when baboons could demonstrate 75% correct minimum with a sample gap size of 32 pixels. Note that gaps of 30 pixels or more gave rise to chance performance in Experiment 3. Training phase 2 followed the same incremental procedure as in phase 1, but now expanded the size of the gap for the comparison pairs. The sample gap size remained at a fixed at 30 pixels in that phase. The task corresponded to the traditional 2 X 2 RMTS test by the end of training phase 2, all pairs being now composed of spatially separated elements.

Immediately after training, the baboons received one 100-trial transfer test session with the same set of colors as already used in the transfer test of Experiment 1. The main goal of these final tests was to verify that matching performance was not strongly tied to the use of color set 1, with which the baboons had been over trained during Experiments 1-4.

Results and discussion

Table 1 reports the number of training required to have the baboons achieving 75% correct in each phase, and for each gap size. The four baboons succeeded in all training phases, and were thus able to ultimately match the SAME and DIFFERENT pairs with gap sizes of 30 pixels. Achieving that performance required from 64 (B06) to 86 (B08) 100-trial session**s**, altogether (see Table 1). Results of the transfer tests using set color 2 furthermore revealed an above chance performance in all **baboons (B05:??%** correct; B06: 65%, B07: 71%; B08: 82%, binomial tests, all ps < .05).

GAP	B05		B06		B07		B08	
	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase1	Phase 2
2	1	6	1	2	2	1	1	6
4	5	2	12	3	23	1	5	8
6	1	3	1	2	1	1	1	6
8	1	2	3	1	4	1	1	2
10	6	1	2	1	2	1	6	3
12	6	3	1	5	3	1	6	2
14	1	4	1	5	1	1	1	1
16	1	1	1	1	3	1	1	3
18	1	1	1	1	9	3	1	4
20	1	4	3	1	1	1	1	1
22	9	4	4	1	1	2	9	1
24	1	1	1	1	2	1	1	4
26	2	4	1	1	1	6	2	3
28	3	2	1	1	3	2	3	1
30	1	6	2	3	1	1	1	1
Total	40	44	35	29	57	24	40	46

Table 1: Number of 100-training trials required to reach 75% correct for each baboon, test phase and gap size (from 2 to 30 pixels).

General Discussion

Our study assessed the ability of baboons to reason by analogy. Because earlier studies showed that baboons are highly sensitive to inter-element gaps during the perception of compound stimuli (Fagot & Deruelle, 1997), we have adopted a novel 2 X 2 item RMTS procedure using pairs of adjacent color patches as SAME/DIFFERENT stimuli. Inspection of the training performance in Experiment 1 revealed that the baboons eventually learned the task, after a number of training trials ranging from 1900 to 4700 trials. This apparently long training should be compared to the numbers of trials (range 12700-23520) to learn to match 2 or SAME or DIFFERENT icons with arrays of 16 icons (Fagot et al., 2001). Differences in training durations suggest that the use of continuous stimuli promoted leaning in our task. For two reasons, it is unlikely that performance achieved in training reflects the learning of specific configurations of colors. First, use of a first set of 10 different colors allowed creation of very large number of stimulus configurations (20160 distinct SAME and 241920 distinct DIFFERENT configurations) which makes rote learning unlikely. Second, training performance transferred successfully for all baboons, when transfer was assessed with a new set of 10 colors. Positive transfer minimally suggests use of open-ended RMTS strategies.

Earlier, we have demonstrated that baboons can solve 2- by 16 item RMTS task with appropriate training procedures (Fagot et al., 2001). According to Flemming et al. (2007), that performance is subject to interpretation as it might best be explained by a control of the behavior by perceptual entropy, rather than purely abstract conceptualization. To address that concern, we note that it is impossible to create SAME or DIFFERENT displays with identical entropies, variations in abstract relationships being consubstantial to entropy variations. Flemming et al.'s (2007) criticism may therefore apply to all RMTS tasks used so far, including Premack's early 2- by 2 item RMTS procedure for which the DIFFERENT pairs had also a higher entropy than the SAME pairs. No such clear distinction can apparently be made between perceptual and cognitive influences in this task, the abstract relations of "sameness" or "differentness" being by essence grounded in perception (Goldstone & Barsalou, 1998). Nevertheless, we accept the idea that early studies employing muti-element arrays illustrated the SAME and DIFFERENT relationships with much greater entropy differences (i.e., 4 in the case of 16-icon arrays) than in the original 2- by 2 item RMTS procedure (i.e., 1 in that case), and that this aspect of the procedure might have helped baboons to discriminate the SAME vs DIFFERENT relationships.

By construction, the 2-item pairs of stimuli can only convey two values of entropy, i.e., a value of 0 in the case of the SAME pairs and a value of 1 in the case of the DIFFERENT pairs. In the current research, baboons could successfully solve the 2-items RMTS task although the SAME and DIFFERENT configurations maximally differed by an entropy of 1. To our knowledge, that performance is the best achieved so far by a non-ape species tested in a RMTS task. From the standpoint of entropy differences, the performance of our baboons matches perfectly that of language (Premack, 1983) or token train apes (Thompson et al., 1997; Vonk, 2003) which have already proved their ability to solve similar 2- by 2 item RMTS tasks also characterized by a maximal entropy difference of 1.

True abstract conceptualization should transcend (visual) appearance, and its demonstration implies a performance relatively independent of specific stimulus configurations, suggesting cognitive flexibility. Baboon's performance in the 2- by 2-item RMTS task met that criterion in Experiment 2. Although weaker than in Experiment 1, performance was above chance in the probe trials of Experiment 2, when the sample and comparisons configurations had different sizes. For two baboons at least, positive transfers rule out the idea that the subjects solved the task by considering the total surface area occupied by the two identical color patches in SAME trials, or by the elemental color patches in the DIFFERENT trials.

Experiment 3 has shown that the performance of the baboons by the end of Experiment 1-2 required was controlled by gap size. To account for this finding, the reviewers of this paper suggested that our subjects may have neglected the two pixel line separating the elements of the training pairs, and considered the SAME pairs as made of one single element, and the DIFFERENT pairs as made with two elements. Doing so, they would solve the task by considering the number of elements, rather than the abstract relations between relations. We believe that this explanation is unlikely for two reasons at least. First, a counting strategy should have induced judgments of all pairs as being made of two elements in Experiment 3, because of the gaps, and should have thus lead to a systematic choice of the comparison pairs also made two elements, namely the DIFFERENT pairs. Inspection of the data shows that there were no such systematic preferences for the DIFFERENT comparison pairs in this task. Thus, the analysis of the gap trials (10, 30 and 60 pixel conditions combined) for the four baboons revealed either a preference for the SAME pattern (for B6 with 75% of the SAME responses), no preference at all (for B5, 51.3% of SAME response), or a preference for the DIFFERENT pairs (B3: 67.4% of DIFFERENT responses, B08: 86.8%). Note moreover that the two baboons which expressed the most which had the most clear-cut findings in Experiment 3 (i.e., B06 and B08) showed opposite biases, suggesting idiosyncratic strategies in transfer trials (note 2). Second, use of a counting strategy cannot explain why performance declined with gap size in Experiment 3. If the baboons did count the number of elements, then their choices should be affected by the presence/absence a gap, but should be independent of gap size.

Experiment 4 revealed that four baboons could finally match the SAME and DIFFERENT stimulus pairs, when the sample of comparison pairs were made of elements separated by 30 pixels. This finding further rules out the counting hypothesis, because the use of a large gap made enhances the discriminability of the bars, and therefore alleviates any doubt that a pair could be perceived as being made of one, rather than two elements. Because we can imagine no other strategies that may account for this performance, we are inclined to conclude that the baboons used the abstract relations of sameness or differentness as cues, in absence of major entropy variations.

It has been proposed that consideration of the relational structure of the stimuli in a RMTS task requires language or token training (Premack, 1983; Thompson et al., 1997), or is restricted to apes (Flemming et al., 2007; Thompson & Oden, 2000). The current research does not support these hypotheses, and suggests instead that even naïve monkeys may possess at least rudimentary abilities for relational thinking. Of course, that conclusion does not rule out the idea that relational thinking might be best developed in apes than in monkeys. It also does not discount the possibility that language or token trainings may promote relational thinking. The current research only implies that symbols or language are not mandatory for relational thinking. Noticeably, solving the 2 X 2 RMTS tasks with gaps required a large number of trials for our baboons. As suggested by Thompson et al. (2001), one of the contributions symbols (or language) might be provide direct access to relational knowledge, in particular in the 2 X 2 RMTS task. Access to that knowledge was possible in our task, but required an extensive and progressive training for our language/token naïve subjects.

Given these findings, the question arises of why other attempts to have monkeys solving a 2 X 2 RMTS systematically failed (e.g., Fagot et al., 2001; Flemming et al., 2007), or why did baboons failed to solve the task with gaps greater than 30 pixels in Experiment 3 ? We propose that this is due to the way monkeys spontaneously structure the RMTS displays. Solving the 2-by 2-item RMTS task requires that the six stimuli presented during a trial are grouped in pairs of objects, in order to compare the abstract relations conveyed by the sample and comparison pairs. It thus implies a process of structural organization. There is now strong evidence that at least monkeys (De Lillo et al., 2005; Deruelle & Fagot, 1998; Fagot & Deruelle, 1997; Spinozzi et al., 2003) and birds (Cavoto & Cook, 2001) find it difficult to group forms separated by gaps in a structure of a higher order. Whether chimpanzees have this difficulty for grouping remains more uncertain (Fagot & Tomonaga, 1999). In baboons, a local mode of stimulus processing emerges both with illusory displays (Parron & Fagot, 2007) and in tasks requiring more explicit responses to the global or local structure of the stimuli (Fagot & Deruelle, 1997). Using a visual search task, Deruelle & Fagot (1998) required baboons to search for global or local differences between large shapes made of smaller non-adjacent shapes. Response times increased with display size in global trials, suggesting an attentional treatment of the global structure of the stimuli. There were by contrast no such search slopes in local trials, indicating an automatic detection of the local targets in that condition. Based on this observation, we suggest that the RMTS is a highly attention demanding task for baboons, considering that they have to group objects in pairs. In the case of the 2- by 2 items RMTS with spatially separated stimuli, grouping is required three times in succession, first to retrieve relational information from the sample, and then to retrieve it from the two comparison pairs, prior to the process of relation comparison. Use of adjacent pairs in Experiments 1-2 may have alleviated that difficulty in terms of grouping, to better reveal their abilities for relational thinking. In our research, we suspect that the initial presentation of the elemental features of the pairs in close proximity in Experiments 1-2, and the progressive enlargement of the gap in Experiment 4, have both favored the structural organization of the displays as made up with pairs of elements to be compared, and by extension reference to the SAME and DIFFERENT relations in the RMTS tasks. Incidentally, success in our RMTS task might have also been promoted by our original testing procedure based on voluntary participations to the task. Voluntary participation may have enhanced sustained attention to the task, and thus learning.

Consideration of the grouping factor may explain why baboons and pigeons are better in the RMTS task, when the relations are illustrated by multi-element displays (e.g., Fagot et al., 2001), but are poor performers with displays made up with pairs of objects. Use of the multi-element arrays has the advantage to reduce average inter-element gap size, therefore limiting the need for grouping. This explanation can also explain why language or token naïve chimpanzees showed sensitivity to relation between relations in Thompson and Oden (1995). In this study, implicit perception of the relation between relations was presumably facilitated by use of pairs of objects mounted on a single board, again reducing the need for grouping.

In sum, the present study provides evidence that the baboons possess (at least rudimentary) skills to judge relations between identity-nonidentity relations, and suggests that this ability to judge the relation between relations was hidden in earlier studies by their difficulty to group the objects as pairs. Use of continuous stimuli made of two adjacent SAME or DIFFERENT elements has alleviated the grouping demand in our research, to reveal their ability for relational thinking. Further studies on the evolutionary origins of human relational thinking will indicate if this ability is in the exclusive realm of primates, or is also shared by non primate animals.

Acknowledgements

JF and CP were supported by the European Community SEDSU grant # 012-984, and the European Origin of Man, Language and Languages (OMLL) program. Dany Paleressompoulle is acknowledged for electronical support. Marc Martin is acknowledged for his technical contribution. This work received an agreement from the “Comité regional d’éthique pour l’expérimentation animale”.

Footnotes

Note 1. This calculation considers the configuration of all three pairs, the two different (left/right) locations of the comparison pairs on the screen, as well as the left/right possible locations of each color bar in the pairs made of two different colors.

Note 2. Idiosyncratic strategies of this type are common in these animals when performance deteriorates (for another example, see Wasserman, Young et al., 2001)

Figure Captions

Figure 1: Stimulus configuration used in Experiment 1. Colors are indicated by numbers for illustrative purposes. The trial illustrated here is a DIFFERENT trial during which the baboon to select the stimulus comparison (S+) showing the same DIFFERENT relation as the compound sample.

Figure 2: Learning curves in Experiment 1.

Figure 3: Individual performance in probe trials. Individual performance was systematically above chance, as revealed by one-tailed binomial tests (p<.05)

Figure 4. Illustration of the test procedure adopted in the first (Figure 4a) and second (Figure 4b) test phase of Experiment 2.

Figure 5: Average transfer performance in Experiment 2 and for each baboon. Stars indicate an above chance performance inferred from a one-tailed binomial test (p<.05).

Figure 6. . Illustration of the test procedure of Experiment 3. Note the gap separating the two color patches composing the sample stimuli. During testing, gap size were of 2 (baseline trials), 10, 30 or 60 pixels.

Figure 7. Performance in the probe trials of Experiment 3, as a function of gap size (in pixels). Stars indicate an above chance performance (one-tailed binomial tests, p<.05).

References

Blaisdell, A. P., & Cook, R. G. (2005). Two-item same-different concept learning in pigeons. *Learning & Behavio*r, 33, 67-77.

Bovet, D., & Vauclair, J. (2001). Judgment of conceptual identity in monkeys. *Psychonomic Bulletin & Review*, 8, 470-475.

Cavoto, K. K., & Cook, R. G. (2001). Cognitive precedence for local information in hierarchical stimulus processing by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 3-16.

Cook, R. G., Kelly, D. M., & Katz, J. S. (2003). Successive two-item same-different discrimination and concept learning by pigeons. *Behavioural Processes,* 62, 125-144.

Cook, R. G., & Wasserman, E. A. (2006). Relational learning in pigeons. In E. A. Wasserman & T. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 307-324). London: Oxford University Press.

Cook, R. G., & Wasserman, E. A. (2007). Discrimination and transfer of higher-order same/different relations by pigeons. *Psychonomic Bulletin & Review*, 14, 1107-1114.

De Lillo, C., Spinozzi, G., Truppa, V., & Naylor, D. M. (2005). A comparative analysis of global and local processing of hierarchical visual stimuli in young children (*Homo sapiens*) and monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 119(2), 155-165.

Deruelle, C., & Fagot, J. (1998). Visual search in global/local stimulus features in humans and baboons. *Psychonomic Bulletin & Review*, 3, 476-481.

Fagot, J. (submitted). Automatic testing of cognitive performance in baboons maintained in social groups.

Fagot, J., & Deruelle, C. (1997). Processing of global and local visual information and hemispherical specialization in humans (Homo sapiens) and baboons (Papio papio). *Journal of Experimental Psychology: Human Perception and Performance*, 23, 429-442.

Fagot, J., & Tomonaga, M. (1999). Global and local processing in humans (*Homo sapiens*) and chimpanzees (Pan troglodyes): Use of a visual search task with compound stimuli. *Journal of Comparative Psychology*, 113, 3-12.

Fagot, J., Wasserman, E. A., & Young, M. E. (2001). Discriminating the relation between relations: The role of entropy in abstract conceptualization by baboons (*Papio papio*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 316-328.

Flemming, T. M., Beran, M. J., Thompson, R. K., Kleider, H. M., & Washburn, D. A. (2008). What meaning means for same and different: Analogical reasoning in humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), and rhesus monkeys (*Macaca mulatta*). *Journal Comparative Psychology*, 122, 176-185.

Flemming, T. M., Beran, M. J., & Washburn, D. A. (2007). Disconnect in concept learning by rhesus monkeys (Macaca mulatta): judgment of relations and relations-between-relations. *Journal of Experimental Psychology: Animal Behavior Processes*, 33, 55-63.

Gentner, D., & Christie, S. (2008). Relational language supports relational cognition in humans and apes. *Behavioural Brain Sciences*, 39, 136-137.

Gillan, D. J., Premack, D., & Woodruff, G. (1981). Reasoning in the chimpanzee: I. Analogical reasoning. *Journal of Experimental Psychology: Animal Behavior Processes*, 7, 1-17.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65, 231-262.

Hopkins, W. D., & Washburn, D. A. (2002). Matching visual stimuli on the basis of global and local features by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Animal Cognition*, 5, 27-31.

Kennedy, E. H., & Fragaszy, D. M. (2008). Analogical reasoning in a capuchin monkey (*Cebus apella*). *Journal of Comparative Psychology*, 122, 167-175.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353-383.

Neiworth, J. J., & Wright, A. A. (1994). Monkeys (*Macaca mulatta*) learn category matching in a nonidentical same/different task. *Journal of Experimental Psychology: Animal Behavior Processes*, 20, 429-435.

Oden, D. L., Thompson, R. K. R., & Premack, D. (1990). Infant chimpanzees (*Pan troglodytes*) spontaneously perceive both concrete and abstract same/different relations. *Child Development*, 61, 621-631.

Parron, C., & Fagot, J. (2007). Comparison of grouping abilities in humans (*Homo sapiens*) and baboons (*Papio papio*) with the Ebbinghaus illusion. *Journal of Comparative Psychology*, 121, 405-411.

Peissig, J. J., Kirkpatrick, K., Young, M. E., Wasserman, E. E., & Biederman, I. (2006). Effects of varying stimulus size on object recognition in pigeons. *Journal of Experimental Psychology: Animal Behaviour Processes*, 32, 419-430.

Premack, D. (1983). The codes of man and beasts. *Behavioral & Brain Sciences*, 6, 125-167.

Sands, S. F., Lincoln, C. E., & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the rhesus monkey. *Journal of Experimental Psychology: General*, 111, 369-389.

Spinozzi, G., De Lillo, C., & Salvi, V. (2006). Local advantage in the visual processing of hierarchical stimuli following manipulations of stimulus size and element numerosity in monkeys (*Cebus apella*). *Behavioural Brain Research*, 166, 45-54.

Spinozzi, G., De Lillo, C., & Truppa, V. (2003). Global and local processing of hierarchical visual stimuli in tufted capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 117, 15-23.

Thompson, R. K. R., & Oden, D. L. (1996). A profound disparity revisited: Perception and judgment of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioral Processes*, 35, 149-161.

Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, 24, 363-396.

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge the relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 31-43.

Vonk, J. (2003). Gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*) understanding of first- and second-order relations. *Animal Cognition*, 6, 77-86.

Wasserman, E. A., Fagot, J., & Young, M. E. (2000). Same-Different conceptualization by baboons (*Papio papio*): The role of entropy. *Journal of Comparative Psychology*, 115, 42-52.

Wasserman, E. A., Fagot, J., & Young, M. E. (2001). Same-different conceptualization by baboons (*Papio papio*): The role of entropy. *Journal of Comparative Psychology*, 115, 42-52.

Wasserman, E. A., Young, M. E., & Cook, R. G. (2004). Variability discrimination in humans and animals: Implications for adaptive action. *American Psychologist*, 59, 879-890.

Wasserman, E. A., Young, M. E., & Fagot, J. (2001). Effects of number of items on the baboon's discrimination of same from different visual displays. *Animal Cognition*, 4, 163-170.

Young, M. E., & Wasserman, E. A. (1997). Entropy detection by pigeons: response to mixed visual displays after same-different discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 157-170.

Young, M. E., Wasserman, E. A., & Garner, K. L. (1997). Effects of number of items on the pigeon's discrimination of same from different visual displays. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 491-501.