Maugard, A., Marzouki, Y. & Fagot, J. (2013). Contribution of Working Memory Processes to Relational Matching-to-Sample Performance in Baboons (Papio papio). Journal of Comparative Psychology, 127, 4, 370

Contribution of Working Memory Processes to Relational Matching-to-Sample Performance in Baboons (*Papio papio*)

Anaïs Maugard, Yousri Marsouki & Joël Fagot*

Laboratory of Cognitive Psychology,
CNRS, Aix-Marseille University, Marseille, France

* Corresponding author : J. Fagot, Laboratory of Cognitive Psychology, CNRS, Aix-Marseille University, Fédération de recherche 3C, 3 place Victor Hugo, Bât B, Case D, 13331 Marseille Cedex, France. Email: joel.fagot@univ-amu.fr

Abstract

The scientific literature suggests that analogical thinking is permitted by the acquisition of linguistic skills in humans, but this hypothesis is challenged by the recent demonstrations that language naïve monkeys can solve relational-matching-to-sample (RMTS) problems. However, doubts remain about the real cognitive strategies adopted by monkeys to complete this task. In the present study, 10 baboons were tested in the RMTS task under three conditions of memory load. The introduction of either a delay or a dual task between the sample and comparison pairs hampered their performance, and more so when the sample instantiated a relation of difference. These results suggest memory storage of detailed information on the sample pair to be used for relational comparison, rather than the storage of an overall estimate of the perceptual variability of the stimulus pairs. These results are in accordance with the idea that analogical thinking is possible in absence of linguistic skills.

Keywords: monkeys, analogy, working memory, dual task

Contribution of working memory processes to relational matching-to-sample performance in baboons (*Papio papio*)

Analogical reasoning is a fundamental and ubiquitous aspect of human thought. It is at the core of a variety of cognitive processes of considerable importance, such as categorization (Ramscar & Pain, 1996), inductive inferencing (Holland, Holyoak, Nisbett, & Thagard, 1986) and more generally the ability to develop a flexible body of knowledge (Brown, Kane, & Echols, 1986). Technically, analogical reasoning implies judgments of relations between relations. When confronted to analogy problems, the participant must represent a first relation between the items of a source domain, for instance between a rocket and a moped, in order to subsequently identify that same relation in a target domain, such as between a rabbit and a turtle in our example. Reasoning by analogy therefore implies a sense of sameness, and the coding of the relational similarity between the source and target domains.

There is now a large consensus that analogical reasoning interacts with language acquisition. For instance, 4- and 6-year old children learned relations faster and demonstrated greater transfer to new relational pairs when they have associated verbal labels to these relations (e.g., Gentner, Anggoro, & Klibanoff, 2011). Thus far, two animal studies converge to indicate that language (or symbol) training is a prerequisite for analogical reasoning. Thus, Gillan, Premack, and Woodruff (1981) showed that a female chimpanzee named “Sarah” could construct analogical relationships instantiated by geometric figures and objects. Because Sarah had received prior language training, this study advocates for the role of language training in the development of analogical thinking. Thompson, Oden, and Boysen (1997) more recently showed that language-naïve chimpanzees only trained to label *Same* and *Different* relations can also solve an analogy task with the same level of performance as Sarah, whereas the control chimpanzee who did not received that training failed in the task.

However, this theory on the contribution of symbolic labeling to analogical reasoning is currently challenged by the recent discoveries that baboons (Fagot & Thompson, 2011) and a capuchin monkey (Truppa, Mortari, Garfoli, Privera & Visalberghi, 2011) can solve two-by-two relational-matching-to-sample task (RMTS) in absence of previous language or symbol training. In these studies, the monkeys firstly perceived two shapes that could either be identical (*Same* relation) or different (*Different* relation). After an exploration period, they saw two other stimulus pairs, one showing the same (*Same* or *Different*) relation as the sample and the other the alternative relation. Because the items in the comparison pairs were all novel, matching could only be made by considering relational cues. Six baboons (Fagot & Thompson, 2011) and one capuchin (Truppa et al., 2011) demonstrated above chance matching performance in this task, albeit after thousands of training trials. After training, the baboons also demonstrated reliable transfer to novel stimulus items never seen before, and could furthermore perform above chance on trials involving a shared item between the sample and foil pairs. Success in the RMTS tasks suggests that the monkeys matched relations with relations, and this ability is arguably a cognitive foundation for analogical reasoning.

Although the above RMTS task has all the appearances of a relational task, a debate remains regarding the exact nature of the cognitive process involved in this task. Thus, Penn, Holyoak, and Povinelli (2008) recently questioned the validity of the RMTS task to study analogical reasoning. According to these authors, the animals would only pay attention to the perceptual variability of the overall pairs without considering its constituent elements as entities, in order to select the choice display with the same level of perceptual variability as the sample pair. Their mode of processing would thus transform the RMTS task into a conditional discrimination implying the match of perceptual cues rather than relations. Penn et al.'s (2008) hypothesis was inspired from a large set of data showing that the entropy (i.e., an information theoric measure of the perceptual variability) of the stimuli controlled the behavior of pigeons and nonhuman primates in both same/different (Cook, Katz, & Cavoto, 1997; Wasserman, Hugart, & Kirkpatrick-Steger, 1995; Young & Wasserman, 1997) and relational matching tasks (e.g., Fagot, Young, & Wasserman, 2001; Truppa et al., 2011).

Although the above analysis by Penn et al. (2008) appears a valid one, it calls for two remarks. First, entropy controlled behaviors were demonstrated in tasks in which the *Same-Different* relations were instantiated by displays made of multiple icons. Because the range of possible entropy values is large for such displays, these displays may have promoted entropy-based strategies at the expanse of more relational-strategies. Different strategies may therefore be used by animals when the saliency of the entropy variable is reduced, as done by Fagot & Thompson (2011) who used pairs of shapes rather than arrays of icons as stimuli. Second, the fact that there are alternative non-relational strategies in the RMTS tasks does imply that the nonhuman primates really adopted these alternative strategies. Noticeably, the RMTS task revealed the same developmental trends in children (for instance regarding the effect of word labels on the "relational shift", Gentner & Christie, 2007) as other tasks of analogy for which the entropy cannot be a variable (Loewenstein & Gentner, 2005; Ratterman & Gentner, 1998). These convergences suggest that children adopted relational strategies in the RMTS task although entropy-based strategies might work for them too.

Entropy-based strategies only imply to maintain actively in the working memory (WM) an analog estimate of the perceptual variability of the displays to solve the task. By contrast, relational (analogy) strategies require the processing of multiple items in parallel in WM to derive first- and second-order relational structures (e.g., Halford, Wilson, & Phillips, 1998; Mulholland, 1980). Entropy-based strategies have thus the advantage to reduce the

WM load of the RMTS task, in comparison to true relational matching. Our study capitalized on this difference in terms of memory load. It assessed the WM load of the RMTS task to better understand how baboons solve it, and if they adopt an entropy-based strategy or a more relational one. Numerous studies have indeed shown that analogical thinking recruits important WM capacities in humans, suggesting that the use of the WM in tasks of analogical reasoning should be an indication of the strategy really used the participants in these tasks. Thus (1) performance of humans in analogical tasks is reduced when dual tasks are introduced to load their WM (Morrison, Holyoak, & Truong, 2001; Waltz, Lau, Grewal, & Holyoak, 2000); (2) WM capacity and analogical reasoning ability develop simultaneously during childhood (Gathercole, Pickering, Ambridge, & Wearing, 2004; Scholnick, 2008); (3) the prefrontal cortex involved in WM (Glahn et al, 2002; Goldman-Rakic, 1987) is also implicated during the processing of analogical tasks (Waltz et al., 1999; Wharton et al., 2000), and finally (4) computational models of analogical mapping, such as the STAR model of Halford et al. (1994) and the LISA model of Hummel and Holyoak (1997), also postulate that there are inherent limitations in analogy making due to WM limits.

The current study investigated the WM information processing demands of RMTS task in baboons. This is to our knowledge the first animal study focusing its attention on the WM component of this task. Three test conditions were proposed to baboons for that purpose. The first one consisted in a 0-delay RMTS task (*0-Delay* condition). The second test condition (*Delay* condition) introduced a delay between the presentation of the sample and comparison pairs in the RMTS task. The third one (*Dual* condition) replaced that delay by a task aimed at loading the WM resources of the participants in the RMTS task. Comparison of these three conditions will provide insight on the processing the RMTS task in monkeys, and will further document the possibility that these animals can achieve relational thinking in absence of language.

Method

Participants

Ten Guinea baboons (*Papio papio*) were tested, including 6 males and 4 females from 3 to 7 year old. These baboons had approximately three years of experimental history during which they have been tested in a variety of computerized tasks, including relational matching-to-sample (RMTS; see Fagot & Thompson, 2011). They lived within a large social group of 30 individuals maintained within a 700m2 outdoor enclosure adjacent to the test booths. The baboons had a microchip implanted in each forearm for automatic identification. They were never food nor water deprived, but received their regular food ratio everyday at 5 p.m.

Apparatus

This research used new test devices, named Automated Learning Device for Monkeys (ALDM, see Fagot & Bonté, 2010; Fagot & Paleressompoulle, 2009), allowing the test of the monkeys on a voluntary basis. In our laboratory, the monkeys can freely quit their social group and enclosure to enter one in two experimental rooms, each containing five ALDM experimental test systems freely accessible on a 24 hour basis. Each test system comprises a test chamber (70 cm × 70 cm × 80 cm) which is accessible through an open back entrance, and is fitted in its innermost front side with a view port (7 cm × 7 cm) and two hand ports (8 cm × 5 cm each). Looking through the view port allows vision of a 19-inch LCD touch monitor installed at eye level 25 cm from the port. Introducing one hand through one of the hand ports allows interactions with the touch screen. Two antennas fixed around each hand port automatically read the microchip on the forearm of the baboon when the animal introduced its hand through a hand port. Numeric identification signals from the arm tags served to trigger the computer controlled presentation of the stimulus and to assign behavioral measures (stimulus choices and response times) to each subject. Correct responses were rewarded by a drop of grains of dry wheat which was delivered inside the test booth by a home-made dispenser. The experiment was controlled by a customized test program developed by the JF with E-Prime (Version 1.2, Psychology Software Tools, Pittsburgh). With this program, the appropriate stimulus presentation for a given subject could be administered, irrespective of the order in which the baboons spontaneously entered the test booth, and the test booth it decided to use. Stimulus displays had a 1024 x 768 pixel definition.

Stimuli

The stimulus set comprised 50 white geometrical shapes (100 × 100 pixels maximum) drawn on a blue background. Two additional 100 x 100 pixel color stimuli were also used for the dual task (see below). They were a yellow geometrical shape and a pink one, also drawn on a blue background.

General test procedure

A schematic presentation of the testing procedure is provided in Figure 1. Briefly stated, the trials started when the baboon introduced one hand through a hand port for self identification. This action triggered the presentation of the test trial assigned to that subject. A sample pair made of either two identical (*Same* relation) or different (*Different* relation) shapes appeared in the middle of the screen on a blue background immediately after identification. These two shapes were selected randomly from the set of 50 stimulus shapes, and were displayed adjacent to one another, with four pixels separating their inner borders. When the baboon touched the sample pair, it disappeared from the screen and was replaced by two comparison pairs on the horizontal median axis of the screen, one the left and the other one on the right hemi-screen. One comparison pair instantiated the same relation as the sample (i.e., either the *Same* or *Different* relation, depending on the trial), the other illustrated the alternative relation. The stimulus shapes of the comparison pairs were all different from those of the sample pair, and moreover differed between the positive and negative comparison pairs. The task required to touch the comparison pair illustrating the same (*Same* or *Different*) relation as the sample pair. Touching the comparison pair showing the same relation as the sample cleared the screen and delivered a food reward. Touching the alternative (incorrect) relation also cleared the screen, but triggered a 3-s time-out indicated by a green screen. The *Same* or *Different* relation shown by the sample varied randomly from one trial to the next, as varied randomly the left/right location of the positive comparison pair on the screen. An inter-trial interval of three seconds minimum separated two continuous trials, but this delay could be longer as the participants initiated the trials by themselves. A maximum of 10 seconds was allowed to respond to all display.

Insert Figure 1 about here

The experiment consisted in three consecutive test phases, hereafter referred to as the *0-Delay*, *Delayed* and *Dual* Conditions. They are presented in details below.

0-Delay condition. That condition used the same procedure as described in the General test procedure section. Its main feature was that there was no delay between the offset of the sample pair and the onset of the two comparison pairs. It was therefore a 0-delay sequential RMTS task. This condition was aimed at establishing baseline performance to be compared with the performance obtained in the other two test conditions of memory load.

Each baboon performed three consecutive sessions of 100 randomized trials (50 *Same* trials randomly intermixed 50 *Different* trials). The accuracy of the response (correct or incorrect) and the response time (RT) were recorded in each trial. The RT was defined as the time elapsed between the onset of the comparison pairs and the hand contact on one of these pairs.

Dual task condition. That condition introduced a dual-task between the sample comparison pairs (see Figure 1). In that condition, the baboons perceived a yellow and a pink shape on a blue background immediately after they have touched the sample pair, and were requested to sequentially touch the yellow and pink stimuli in that order, irrespective of their randomized left-right location. Correctly touching these two stimuli in the specified order triggered the immediate display the comparison pairs of the RMTS task. Touching them in an incorrect order aborted the trials and induced a 3-s time-out indicated by a green screen. Each baboon completed four 100 trial sessions in that condition. Sessions were organized as for the *0-Delay* condition and therefore included 50 *Same* trials mixed with 50 *Different* trials. The dependent variables were the accuracy of the response, the RT to complete the dual task, and the RT to touch the correct comparison pair measured from its onset.

Delayed Condition. That condition used the same RMTS procedure as in the *Dual* condition, except that the time interval between the offset of the sample pair and the onset of the comparison pairs was now only filled with a blue background during which no action was required from the baboon. The duration of that delay was controlled independently for each participant, and determined as being equal to the median RTs required by the subject to complete the dual task. It therefore varied among subjects. Each participant completed four 100 trial test sessions which had the same design as the test sessions of the *Delay* condition.

Dependent variables were the accuracy of the response (correct or incorrect) and the RT to select the correct comparison pair measured from its onset.

Test order and pre-training

Because all participants were already familiar with the RMTS (see Fagot & Thompson, 2011), they only received pre-training to the task prior to the 0-*Delay* condition. Pre-training continued until the baboons achieved 80% in 40-trial training sessions. It required from 2 to 8 sessions to reach that criterion. The *0-Delay*, *Dual* and *Delay* conditions were presented in that order. However, an additional training was required to have the baboons learning the dual task. After completion of the *0-Delay* condition, the baboons were thus presented with displays only containing the yellow and pink stimuli of the dual task, and were requested to touch them in the correct order to obtain a reward, regardless of their left-right spatial location. Correct trials were rewarded. Incorrect trials were followed by a 3-s time-out indicated by a green screen. Training sessions of 100 trials were continually repeated until each baboon reached an accuracy rate of 80% or better in three consecutive sessions. We proceeded to the *Dual* condition immediately after they reached that criterion.

Results

Statistical analyses

Repeated-measure ANOVAs require the dependent variables to follow a normal distribution. Arcsine transformations were thus applied on accuracy data to correct for normality. For RTs data, we computed the median correct reaction times of each baboon for each condition, as strongly recommended by Ratcliff (1993). Our analyses considered the Trial type (*Same*, *Different*) and the Test condition (*0-Delay*, *Dual* and *Delayed*) as factors in a 2 x 3 full factorial design. Planned contrasts were used to test the decrease in performance over time as a function of the Test Condition. In addition *a posteriori* comparisons were conducted using Tukey *HSD* tests. Trials for which the RT to respond to the sample, dual task or comparison displays were greater than 5 seconds were discarded from the analyses. This rejection procedure only removed 0.5% of the total trials.

Accuracy

The Trial Type (*Same, Different*) by Test Condition (*0-Delay, Dual, Delayed)* repeated measure ANOVA on accuracy data indicated significant main effects of Trial Type, $F(1, 9) = 9.2$, $p < .05$, $\eta_p^2 = .50$, and Test Condition, $F(2, 18) = 84.5$, $p < .001$, $\eta_p^2 = .90$. The significant main effect of Trial Type was reflected by a higher accuracy for the *Same* (M = 77.85; SD = 10.19) than for the *Different* trial types (M = 68.84; SD = 16.43). The effect of Test Condition was analyzed by planned (contrast analyses) comparisons. Accuracy was significantly lower in the *Dual* condition (M = 62.31, SD = 5.97) than in the *Delayed* condition (M = 72.96, SD = 8.01, $p < .05$), $F(1, 9) = 27.6$, $p < .005$, $\eta_p^2 = .75$. It was also significantly lower in the average of these two latter conditions relative to the *0-Delay* condition (M = 84.80, SD = 5.35), $F(1, 9) =171.7$, $p < .001$, $\eta_p^2 = .95$.

The ANOVA further indicated a significant Trial Type by Test Condition interaction, $F(2, 18) = 4.2$, $p < .05$, $\eta_p^2 = .32$, see Figure 2. Post hoc analyses (Tukey *HSD* tests, $p<.05$) of the *Same* trials showed that the *Dual* condition induced a lower performance (M = 71.85, SD = 9.0) than the *0-Delay* condition (M = 85.87, SD = 5.95, $p < .05$). By contrast, there were no reliable difference for the *Same* trials between the *Delayed* condition (M = 75.84, SD = 10.12) and the other two conditions (all *p*s > .05). Results were different for the *Different* trials, as the following pattern of results was obtained for these trials: *Dual* (M = 52.71, SD = 13.16) < *Delayed* (M = 70.09, SD = 10.4) < *0-Delay* condition (M = 83.73, SD = 7.17).

Insert Figure 2 about here

To complement this analysis, two tailed t-tests finally verified in which of our six conditions (2 Trial-Type by 3 Test-Condition) the baboons expressed an average performance greater than 50% correct (i.e., chance level). The group performed reliably above chance in all conditions (all *p*s < .05), except in the *Same* trials of the *Dual* condition, $t(9) = .266$, $p > .05$.

Response times

The extremely low performance of the baboons in the *Different* trials of the *Dual* condition (i.e., 52.71% correct) rendered meaningless the computation of the 2 by 3 (Trial Type by Test Condition) full factorial design ANOVA for RTs. RT data were analyzed with several independent one-way ANOVAs. Thus, the first one-way ANOVA compared the average RTs obtained in the *0-Delay* and *Delay* conditions in which the subjects systematically performed reliably above chance. It showed that RTs were shorter in the *0-Delay* (*M* = 537 ms, *SD* = 71 ms) condition than in the *Delay* condition (*M* = 728 ms, *SD* = 102 ms, F(1,9) = 41.97, $p < .001$, $\eta_p^2 = .82$). Two subsequent one-way ANOVAs were computed to investigate the effect of Test Condition on RTs separately for the *Same* and *Different* trials. The first ANOVA performed on *Same* trials considered all the three test conditions, as the baboons performed above chance in these conditions (t-tests comparing average performance to 50% correct, all *p*s < .05). It revealed a reliable effect of test condition, F(2,18) = 27.3, $p < .001$, $\eta_p^2 = .75$. A post-hoc analysis of this effect showed faster RTs on average in the *0-Delay* condition (*M* = 575 ms, *SD* = 84 ms) than in the two other conditions (Tukey *HSD* test, *p*s < .05). There were by contrast no difference between the *Delay* (*M* = 781 ms, *SD* = 105 ms) and *Dual* conditions (*M* = 742 ms, *SD* = 65 ms; Tukey

HSD tests, $p > .05$). The one-way ANOVA on *Different* trials only considered the two test conditions for which the performance was above chance (t-tests, *p*s < .05), namely the *0-Delay* and *Delay* conditions. These two conditions differed reliably in terms of RTs, $F(1,9) = 39.7$, $p < .001$, $\eta_p^2=.82$. They were shorter on average in the *0-Delay* ($M = 506$ ms, $SD = 67$ ms) than in the *Delay* condition ($M = 686$ ms, $SD = 113$ ms).

Discussion

With very few exceptions (Gillan et al., 1981; Haun & Call, 2009; Hribar, Haun, & Call, 2011), the literature on analogical reasoning by animals has so far exclusively used the RMTS task (e.g., Cook & Wasserman, 2007; Fagot & Thompson, 2011; Flemming, Beran, & Washburn, 2007; Thompson et al., 1997; Truppa et al., 2011), but the cognitive strategy adopted to solve this task remains largely unknown. Because early studies in humans demonstrated that analogy making recruits important WM resources (Morrison et al., 2001; Waltz et al., 2000), the current research focused on the demand of the RMTS task in terms of WM. We thus explored the baboons' RMTS performance in three test conditions implying different WM loadings. The first condition (i.e., *0-Delay*) minimized the WM demand of the task as there was no delay between the sample and comparison displays. The second one (i.e., *Delay*) imposed a delay between these two displays during which the animal had to actively remember the sample display to later use it when the analogy judgment has to be made. The last one (i.e., *Dual*) introduced a dual task between the sample and choice display, the baboons having to actively maintain information on the sample pairs during the dual task, in order to subsequently identify its relational match within the comparisons pairs.

Our manipulations of the memory loads modified baboons' RMTS performance. Thus, the best performance (mean = 84.8%) was obtained in the *0-Delay* condition which minimized the memory load, and this performance was reliably greater than in the two other conditions of memory load. More importantly, we also found that the performance was lower in the *Dual* condition (mean=62.31%) maximizing the memory load, than in the *Delay* condition (mean=72.96). The full-design analysis of RTs was impossible due to a poor performance in the *Different* trials of in the *Dual* condition. Nevertheless, we found that the RTs were reliably longer in the *Delay*- than in the *0-Delay* condition, therefore confirming that this latter condition was more difficult that the former one.

The lower performance in *Delay* as compared to the *0-Delay* condition can arguably be explained by a progressive lost of the memory trace of the sample pair due to the delay. However, the memory decay does not explain why performance in the dual task was lower than in the *Delay* condition, because the delays had equal durations in these two conditions. The dual task implied WM executive resources to inhibit the response to the incorrect item that changed from the first to the second item in the sequence. Because the dual and RMTS tasks conflicted in their use of the WM resources, results demonstrate that the RMTS task imposes strong constraints on the WM. Complementary, our results further demonstrated a reduced performance in *Different* than in the *Same* trials. Performance difference between these two types of trials remained small in amplitude (2.14%) in the *0-Delay* condition, but increased in the *Delay* condition (5.75%) and even more drastically in the *Dual* condition (19.14%). Findings therefore indicate that the processing of *Different* sample pairs was much more demanding in terms of WM resources than the processing of the *Same* pairs.

In their review, Penn et al. (2008) claimed that nonhuman primates succeed in the RMTS task because they reduce the *Same*/*Different* binary relations into a unitary continuous dimension (presumed to be the perceptual variability), and compare the value characterizing each pair along the relevant dimension to infer the correct answer. In doing so, they would neglect information about the items and their relations because they are irrelevant for them. One interesting prediction from this hypothesis is that the storage of the codes characterizing the *Same* and *Different* relations should roughly occupy the same space in WM. Consider for instance the encoding of the pairs along the *high*/*low* variability dimension suggested by Penn et al. (2008), the storage of *high* code should not be more demanding than the storage of the *low* code, and one bit of information should suffice to encode each pair as being either a *high* or *low* pair. Penn et al.'s theory (2008) therefore cannot explain why the memory load was greater for the *Different* than the *Same* trials in our task. We believe that this result discounts their hypothesis of an encoding of the stimulus pairs along a single unitary perceptual dimension.

How then can we account for our findings? To answer this question, it should be noted that *Different* stimulus pairs contain *de facto* more information than the *Same* pairs, because the shapes composing them are different. A direct consequence of this structural difference is that the storage of the *Different* pair should recruit more memory resources than the storage of the *Same* pairs, and this is exactly what we found. The fact that the dual task has affected more drastically the processing of the *Different* pairs in our study demonstrates that the monkeys attended to, and maintained actively in their WM, detailed information on sample pairs during the RMTS tasks, and later use that input to compare the (stored) sample pair and (perceived) comparison pair.

Empirical studies with humans have shown that analogical reasoning places great demands in terms of WM resources (Morrison et al., 2001; Waltz et al., 2000). This memory load is explained by the need to process multiple items in parallel (Halford et al., 1998), in order to infer the relations between items belonging to the same (source or target) domain, or between these domains. Two of our most recent findings suggest that baboons too may have sufficient memory resources to process analogies, although their WM capacity is admittedly more reduced in nonhuman primates than in humans (Elmore et al., 2011). Thus, Fagot & De Lillo (2011) tested the baboons in an analog of the Corsi test, in which the participants had to reproduce on the screen a sequence of items in an appropriate serial order. Monkeys showed a memory span of 4 to 5 items in the task, depending on the subjects (for similar findings on macaques, see Botvinick et al., 2009; Wright, 2007). Although smaller than the memory span of humans tested in the same study (Fagot & De Lillo, 2011), a memory span of 4 items appears large enough to process two sets of two items within the WM, and therefore to process analogies. Another constraint of analogy making is that the subjects much must be able to remember the characteristics of the first set of items (source domain), when processing the second (target domain) set, in order to later compare the relations illustrated by these two domains. This ability also appears in the scope of baboons. Thus, Rey, Perruchet and Fagot (2012) recently showed that the baboons can embed the processing of two stimulus pairs in an ABBA structure. They can thus process a first item of a pair, and then process the second pair, to finally resume the processing of second item of the initial pair. In sum, their WM appears to have all the properties required to solve analogy problems.

Admittedly, one limit of the current study is that it did not definitively demonstrate that monkeys use and compare relational information to solve the RMTS task. However, this study demonstrates for the first time a reliance on WM which appears highly compatible with the monkeys used of analogical reasoning strategies to solve the RMTS task. It also shows that the hypothesis of variability encoding cannot account for the findings, further stressing the possibility that the monkeys can solve analogy problems in absence of language.

References

Botvinick, M. M., Wang, J., Cowan, E., Roy, S., Bastianen, C., Mayo, J. P., et al. (2009). An analysis of immediate serial recall performance in a macaque. *Animal cognition, 12,* 671-678. doi:10.1007/s10071-009-0226-z

Brown, A. L., Kane, M. J., & Echols, C. H. (1986). Young children's mental models determine analogical transfer across problems with a common goal structure. *Cognitive Development, 1*, 103-121. doi:10.1016/s0885-2014(86)80014-4

Christie, S., & Gentner, D. (2007). Relational similarity in identity relation: The role of language. In S. Vosniadou & D. Kayser (Eds.), *Proceedings of the Second European Cognitive Science Conference* (pp. 601–666). London: Taylor & Francis.

Cook, R. G., Katz, J. S., & Cavoto, B. R. (1997). Pigeon Same-Different concept learning with multiple stimulus classes. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 417-433. doi:10.1037//0097-7403.23.4.417

Cook, R. G., & Wasserman, E. A. (2007). Learning and transfer of relational matching-to-sample by pigeons. *Psychonomic Bulletin & Review, 14,* 1107-1114. doi:10.3758/bf03193099

Elmore, L. C., Ma, W. J., Magnotti, J. F., Leising, K. J., Passaro, A. D., Katz, J. S., et al. (2011). Visual short-term memory compared in rhesus monkeys and humans. *Current Biology, 21,* 975-979. doi:10.1016/j.cub.2011.04.031

Fagot, J., & Bonte, E. (2010). Automated testing of cognitive performance in monkeys: Use of a battery of computerized test systems by a troop of semi-free-ranging baboons (*Papio papio*). *Behavior Research Methods, 42,* 507-516. doi:10.3758/brm.42.2.507

Fagot, J., & De Lillo, C. (2011). A comparative study of working memory: Immediate serial spatial recall in baboons (*Papio papio*) and humans. *Neuropsychologia, 49*, 3870-3880. doi:10.1016/j.neuropsychologia.2011.10.003

Fagot, J., & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods, 41*, 396-404. doi:10.3758/brm.41.2.396

Fagot, J., & Thompson, R. K. R. (2011). Generalized Relational Matching by Guinea Baboons (Papio papio) in Two-by-Two-Item Analogy Problems. *Psychological Science, 22,* 1304-1309. doi:10.1177/0956797611422916

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. doi:10.1037//0097-7403.27.4.316

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. doi:10.1037/0097-7403.33.1.55

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177-190. doi:10.1037/0012-1649.40.2.177

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure mapping and relational language support children's learning of relational categories. *Child Development, 82,* 1173-1188. doi:10.1111/j.1467-8624.2011.01599.x

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

Glahn, D. C., Kim, J., Cohen, M. S., Poutanen, V. P., Therman, S., Bava, S., et al. (2002). Maintenance and manipulation in spatial working memory: Dissociations in the prefrontal cortex. *NeuroImage, 17*, 201-213. doi:10.1006/nimg.2002.1161

Goldman-Rakic, P. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational knowledge. *Handbook of Physiology, 5*, 373-417.

Halford, G. S., Wilson, W. H., Guo, J., Gayler, R. W., Wiles, J., & Stewart, J. E. M. (1994). Connectionist implications for processing capacity limitations in analogies. In K. J. Holyoak & J. A. Barnden (Eds.), *Analogical connections: Advances in connectionist and neural computation theory* (Vol. 2, pp. 363-415). Westport, CT: Ablex Publishing.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803-865.

Haun, D. B. M., & Call, J. (2009). Great apes' capacities to recognize relational similarity. *Cognition, 110*, 147-159. doi:10.1016/j.cognition.2008.10.012

Holland, J. H., Holyoak, K. J., & Nisbett, R. E. (1989). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: Bradford Books/MIT Press.

Hribar, A., Haun, D., & Call, J. (2011). Great apes' strategies to map spatial relations. *Animal cognition, 14*, 511-523. doi:10.1007/s10071-011-0385-6

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*, 427-466. doi:10.1037//0033-295x.104.3.427

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*, 315-353. doi:10.1016/j.cogpsych.2004.09.004

Morrison, R. G., Holyoak, K. J., & Truong, B. (2001). Working memory modularity in analogical reasoning. In *Proceedings of the TwentyThird Annual Conference of the Cognitive Science Society* (pp. 663–668). Mahwah, NJ: Erlbaum.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology, 12*, 252-284. doi:10.1016/0010-0285(80)90011-0

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109-178. doi:10.1017/s0140525x08003543

Ramscar, M., & Pain, H. (1996). Can a real distinction be made between cognitive theories of analogy and categorization? In *Processings of the18th Annual Conference of the Cognitive Science Society* (pp. 346-351): Erlbaum.

Ratcliff, R. (1993). Methods for dealing with reaction-time outliers. *Psychological Bulletin, 114*, 510-532. doi:10.1037/0033-2909.114.3.510

Rattermann, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In K. Holyoak, D. Gentner & B. Kokinov (Eds.), *Advances in analogy research: Integration of theory & data from the cognitive,*

computational, and neural sciences (pp. 274–282). Sophia: New Bulgarian University.

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (*Papio Papio*). *Cognition, 123,* 180-184. doi:10.1016/j.cognition.2011.12.005

Scholnick, E. K. (2008). New directions in Piagetian theory and practice. *Journal of Applied Developmental Psychology, 29*, 345-347. doi:10.1016/j.appdev.2008.04.001

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes, 23*, 31-43. doi:10.1037/0097-7403.23.1.31

Truppa, V., Mortari, E. P., Garofoli, D., Privitera, S., & Visalberghi, E. (2011). Same/Different concept learning by capuchin monkeys in matching-to-sample tasks. *Plos One, 6*. doi:10.1371/journal.pone.0023809

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., Santos, M. D., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119-125. doi:10.1111/1467-9280.00118

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28*, 1205-1212. doi:10.3758/bf03211821

Wasserman, E. A., Hugart, J. A., & Kirkpatricksteger, K. (1995). Pigeons show Same-Different conceptualization after training with complex visual-stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 21*, 248-252. doi:10.1037/0097-7403.21.3.248

Wharton, C. M., Grafman, J., Flitman, S. S., Hansen, E. K., Brauner, J., Marks, A., et al. (2000). Toward neuroanatomical models of analogy: A positron emission tomography study of analogical mapping. *Cognitive Psychology, 40*, 173-197. doi:10.1006/cogp.1999.0726

Wright, A. A. (2007). An experimental analysis of memory processing. *Journal of the experimental analysis of behavior, 88*, 405-433. doi:10.1901/jeab.2007.88-405

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. doi:10.1037/0097-7403.23.2.157

Acknowledgements

JF and AM were funded by Grant ANR-2010-BLANC-1908-01 (project Anafonex). JF was further supported by a research grant “Projet Exploratoire” from the Provence Regional Council, and the EuprimNet-II grant from the European Council. This research was conducted at the Rousset-sur-Arc primate center (CNRS-UPS846). The authors thank J.C Marin, J. Di Grandi and D. Paleressompoulle for technical assistance, as well as Timothy Flemming for stimulating discussions on preliminary data.

Figures captions

Figure 1. Illustration of the trials in each test condition.

Figure 2. Percentage of correct responses for *Same* and *Different* trials under the three test conditions. The three test conditions are organized on the X axis as a function of their memory load. The error bars represent the SD of the distribution.

Figure 1.

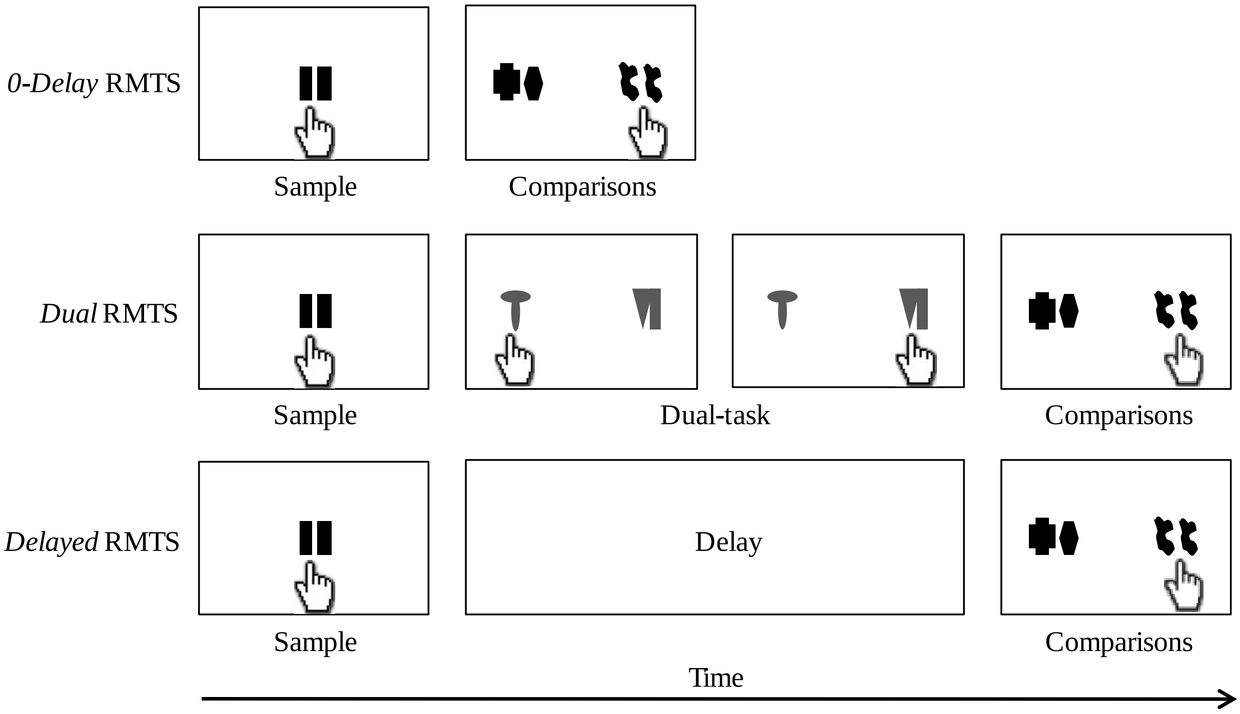
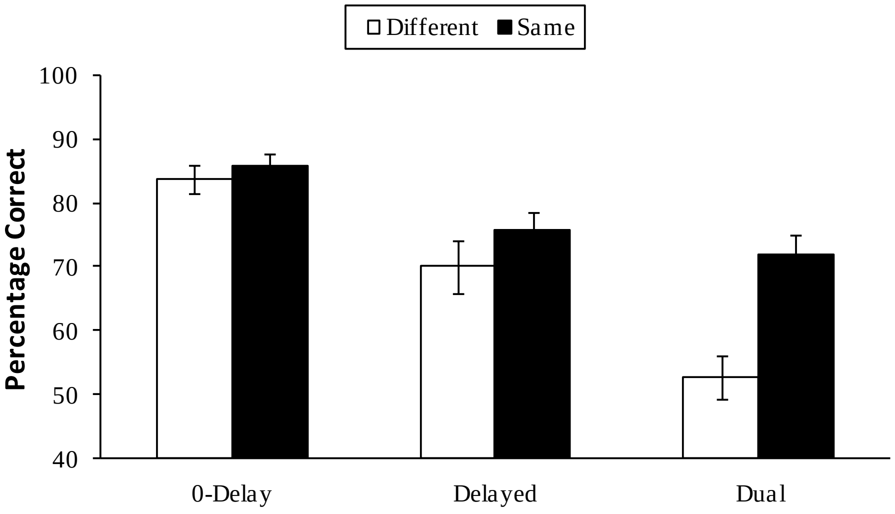

0-Delay RMTS
Sample
Comparisons
Dual RMTS
Sample
Dual-task
Comparisons
Delayed RMTS
Delay
Sample
Comparisons
Time

Figure 2.


☐ Different ■ Same
Percentage Correct
100
90
80
70
60
50
40
0-Delay
Delayed
Dual