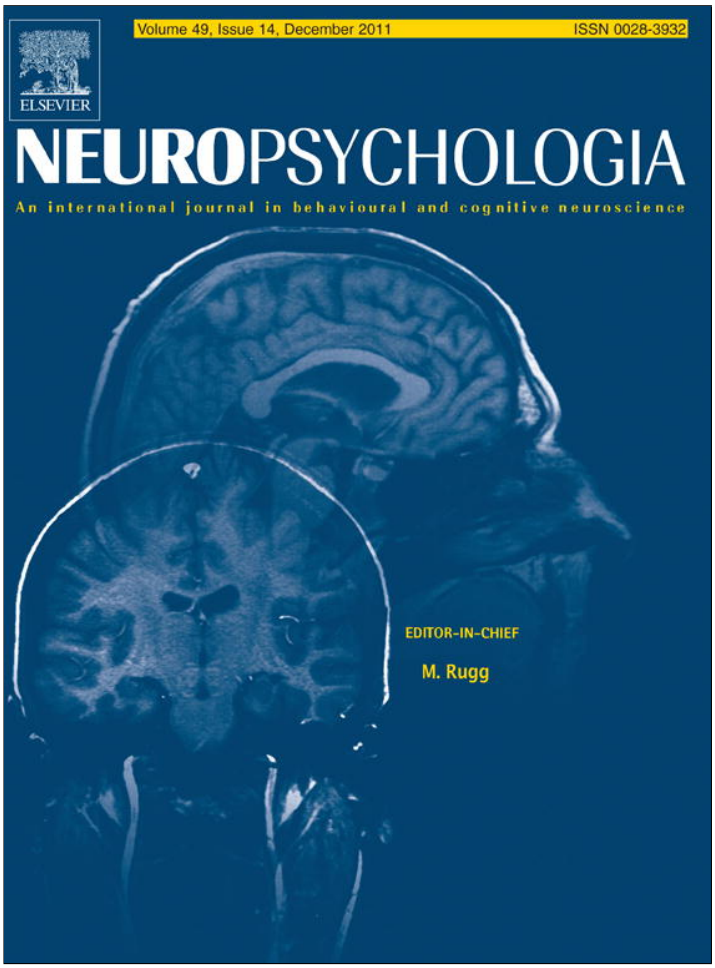

Neuropsychologia 49 (2011) 3870–3880

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

A comparative study of working memory: Immediate serial spatial recall in baboons (*Papio papio*) and humans

Joël Fagot[a], Carlo De Lillo[b,*]

[a] Université de Provence, Laboratory of Cognitive Psychology, 3 Place Victor Hugo, 13331 Marseille Cedex 1, France
[b] University of Leicester, School of Psychology, Lancaster Road, Leicester LE1 9HN, UK

ARTICLE INFO

Article history:
Received 20 January 2011
Received in revised form 3 October 2011
Accepted 5 October 2011
Available online 12 October 2011

Keywords:
Working memory
Primate
Baboon
Corsi test
Serial recall
Spatial cognition

ABSTRACT

Two experiments assessed if non-human primates can be meaningfully compared to humans in a non-verbal test of serial recall. A procedure was used that was derived from variations of the Corsi test, designed to test the effects of sequence structure and movement path length in humans. Two baboons were tested in Experiment 1. The monkeys showed several attributes of human serial recall. These included an easier recall of sequences with a shorter number of items and of sequences characterized by a shorter path length when the number of items was kept constant. However, the accuracy and speed of processing did not indicate that the monkeys were able to benefit from the spatiotemporal structure of sequences. Humans tested in Experiment 2 showed a quantitatively longer memory span, and, in contrast with monkeys, benefitted from sequence structure. The results are discussed in relation to differences in how human and non-human primates segment complex visual patterns.

1. Introduction

Neuropsychological research on working memory (WM) in non-human primates has a long history (see Jacobsen, 1936 for an early account). It has generated extremely important findings, including the demonstration of the involvement of the lateral prefrontal cortex in WM (see Goldman-Rakic, 1992; Petrides, 1998, for reviews). Two main goals may be achieved by studying WM in non-human primates. Firstly, studies with non-human primates allow the localisation of high cognitive functions (which can then allegedly be generalized to humans) in a much more controlled way than human clinical neuropsychological studies (Petrides, 1998). Secondly, considering the close relationship between WM and high level human skills, including general intelligence (Baddeley, 2001, 2003, for a review), reasoning (Kyllonen & Stephens, 1990) and scholastic achievement (Gathercole, Pickering, Knight, & Stegmann, 2004; St Clair-Thompson & Gathercole, 2006), it is important to assess the similarity of WM processes of humans and those of other primates. Ideally, the same tasks which have been used to characterize WM in humans should also be used with other species, for comparative purposes.

One of the most important non-verbal tests in human neuropsychological research (Berch, Krikorian, & Huha, 1998) is the Corsi Tapping Test (CBTT) (Corsi, 1972; Milner, 1971). This is a test of immediate serial spatial recall which is widely used for the assessment of what is typically referred to as "Spatial span". The CBTT has been considered (see Baddeley, 2000) as the purest assessment of the capacity of the visuo-spatial sketch-pad within the influential WM model proposed by Baddeley and Hitch (1974). It is included in many neuropsychological test batteries, including the widely used automated neuropsychological battery test CANTAB (Robbins, James, Owen, & Sahakian, 1994). In its modern computerized version, the CBTT simply requires participants to observe a set of identical icons placed on different locations on a touch screen (De Lillo, 2004; De Lillo & Lesk, 2010; Vandierendonck, Kemps, Fastane, & Szmalec, 2004). The icons will then flash, or change colour, in turn describing a sequence of locations that the participant has to reproduce by touching them on the screen, in the appropriate serial order.

The CBTT has traditionally been used under the assumption that any sequence through the icons is of equivalent memory demand. However, novel versions of the test have been devised to distinguish between sequences which afford specific forms of memory organization and sequences which do not. Studies that have used these variants of the CBTT showed that human spatial WM benefits from particular forms of spatio-temporal organization of the to-be-reported items. For example, within arrays arranged as clusters of locations defined by their spatial proximity, sequences

* Corresponding author. Tel.: +44 0116 229 7193; fax: +44 0116 229 7196.
E-mail addresses: Joel.fagot@univ-provence.fr (J. Fagot), CDL2@LE.AC.UK (C. De Lillo).

0028-3932/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2011.10.003

that require the selection of all the icons within each of the spatial clusters before shifting to a different cluster are reproduced more accurately (De Lillo, 2004). This benefit of spatial–temporal clustering is accounted for by the properties of WM rather than just the motor components of sequence reproduction (De Lillo & Lesk, 2010). In another study, sequences with consecutive items within the same row, column or diagonal, here referred to as "structured", were reproduced at a higher level of accuracy than unstructured sequences that violated these organizational constrains (Bor, Duncan, Wiseman, & Owen, 2003). Importantly, f-MRI data collected by Bor et al. (2003) indicated a selective activation of the Dorso-Lateral-Pre-Frontal-Cortex (DLPC) for structured sequences (Bor et al., 2003). According to Bor and co-workers (2003), this effect suggests that the DLPC is specifically involved in the strategic use of gestalt principles during the encoding of structured sequences.

Another commonly used variation of the Corsi task is the dot task involving the sequential presentation of a number of dots in semi-random locations on a PC monitor. The dots are then displayed simultaneously and the participant is required to select them in the required order using a PC mouse (Parmentier, Elford, & Mayberry, 2005). Results obtained with this task indicate that path-length affects spatial serial recall, even in the relatively small scale display of a computer screen. Another critical factor is the number of crossings present in the path (Parmentier et al., 2005).

It would be of extreme importance to assess the validity of non-human primate models of human WM using similar tasks. However, studies of immediate serial recall in non-human primates have been extremely rare so far, possibly because of the difficulty of training animals to perform this kind of task.

The paradigms most typically used to study working memory in primates are centred on delayed response (Goldman-Rakic, 1992; Gross & Weiskrants, 1962Gross and Weiskrantz, 1962; Jacobsen, 1936), delayed alternation (Butters & Pandya, 1969; Mishkin, Vest, Waxler, & Rosvold, 1969; Mishkin, 1957) and matching to sample tasks (Tavares & Tomaz, 2002). These tasks have been very useful in detecting functional dichotomies such as the short term retention of spatial and non-spatial information (e.g., Goldman-Rakic, 1992; Mishkin & Manning, 1978). However, they do not typically require the temporary monitoring of a series of responses by the subject. Yet, such ability is an important requirement of WM (Petrides, 1998).

Limited information about the ability of primates to temporary monitor multiple responses can be obtained from search tasks that require the exploration of series of baited locations in an artificial foraging situation (De Lillo, 1996; De Lillo, Aversano, Tuci, & Visalberghi, 1998; De Lillo, Visalberghi, & Aversano, 1997). Studies using these tasks have shown that the structure of the search environment can affect WM in primates (for a reviews, see De Lillo, 1996, in press). However, in these tasks the order in which the different locations are explored is freely chosen by the subjects and, as such, they cannot be used to test the effects of the spatio-temporal organization of specific sequences to be reported.

Another common task for assessing the short-term retention of serial order in non-human primates is Serial Probe Recognition (SPR). A set of images (see Matzke & Castro, 1998; Sands & Wright, 1980; Wright, 1999a; Wright, Santiago, & Sands, 1984) or sounds (see Wright, 1999b; Wright & Rivera, 1997) is presented in sequence and, after a variable delay, the animal has to judge whether a probe item was present in the original set or not. SPR makes it possible to determine the number of recognition errors which are made for items at different serial positions during the presentation. By enabling an analysis of serial position curves, SPR has allowed the detection in monkeys of primacy and recency effects (Matzke & Castro, 1998; Sands & Wright, 1980) as observed in human recall. Moreover, SPR tasks have allowed researchers to determine how and when the primacy and recency effects emerge or disappear, as a function of inter-item intervals and recall delay in different monkey species and pigeons (Wright, 1999a, 1999b; Wright & Rivera, 1997; Wright et al., 1984; Wright, Santiago, Sands, Kendrick, & Cook, 1985). Nevertheless, SPR does not require serial recall. In other words, in order to obtain a reward in a SPR task the animal does not need to report the items in a specific serial order. As such, SPR has important differences with the CBTT, which has this requirement.

A task which could be more related to the CBTT is the "limited-hold" memory test used by Inoue and Matsuzawa (2007). This study used two chimpanzees trained in previous studies to touch numerals from 1 to 9 on a touch screen, following the sequence 1–2–3–4–5–6–7–8–9 (Kawai & Matsuzawa, 2000; Matsuzawa, 1985). These nine numerals were presented for a limited period of time before being replaced by white squares. The animals were then required to touch the square in the order previously specified by the numerals. Both chimpanzees reported an impressive percentage of sequences correctly, and one of them outperformed humans tested in the same task. One limitation of the "limited-hold" memory test for our purpose is that it features different visual stimuli as items (i.e., numerals) during the presentation period. Thus, remembering a static display of symbols may suffice to indicate the order in which they have to be reported. In fact, this task may tap "eidetic imagery" (Inoue & Matsuzawa, 2007; Jaensch, 1930). If a static image of the display can be retained, even if very briefly, then it would be the Long Term Memory (LTM) representation of order in which the numeral have to be reported to guide the subjects' responses.

Thus, a component of the "limited-hold" memory test may rely on LTM representations of serial order for non-spatial visual patterns (i.e., the numerals). Convincing evidence of the capacity of non-human primates to represent serial order in LTM has been also provided by serial learning studies (Chen, Swartz, & Terrace, 1997; D'Amato & Colombo, 1988, 1989; Swartz & Terrace, 1991; Terrace, 1993; Terrace & McGonigle, 1994). Nevertheless, neither the "limited hold" nor serial learning tasks require the ability to attend a serial display, encode the order of presentation of the items and then report it, which are all critical components of the CBTT.

To our knowledge, only one study (Botvinick et al., 2009) has assessed WM in non-human primates using a CBTT-type task. In this study, one macaque (*Macaca mulatta*) was presented with items arranged as a square matrix. A red frame appeared in turn around each of the items pertaining to random sequences of three and four items that the animal was required to recall. Several aspects of human immediate serial recall were observed in the error patterns and serial position curves of the monkey. However, this study only used sequences of three and four items and did not experimentally manipulate the structure of the displays. Thus the ability of monkeys to cope with sequences with a higher number of items and to benefit from the organizational principles demonstrated in humans by Bor et al. (2003) remains unknown. This latter issue is particularly critical because there is evidence that monkeys may not benefit from gestalt organizational principles in the same way as people do (Fagot & Deruelle, 1997; Spinozzi, De Lillo, & Castelli, 2004; Spinozzi, De Lillo, Truppa, & Castorina, 2009).

In fact, in contrast with humans, monkeys' ability to group local elements in order to derive the shape of their configuration is particularly vulnerable to the increase of their distance, suggesting a relative difficulty with grouping by proximity (Fagot & Deruelle, 1997). Moreover, in embedded figure tasks monkeys are not sensitive to stimulus properties such as closure in the same way as people are (Spinozzi et al., 2004). Finally, a study specifically designed to compare humans and monkeys on their ability to

segregate figures from their background, by using specific grouping cues, suggested possible differences in the preferential use of grouping cues such as orientation and proximity in monkeys and people (Spinozzi et al., 2009). The presence of differences between monkeys and people in the use of organizational principles in serial recall has not been assessed before and we considered it to be of interest to do it here as it has been claimed that benefits in spatial WM may derive from chunking based on the use of gestalt grouping principles (Bor et al., 2003).

The main goal of the present study was to implement with monkeys a CBTT task that could be used to further evaluate similarities and differences in the WM of humans and of non-human primates. Our task was designed to test if memory performance in monkeys is affected by: (1) the presence of spatio-temporal organization of the sequences to be recalled; and (2) the path-length of the movement required to reproduce them. Spatiotemporal organization, or sequence structure, is operationally defined here as the constraint of having consecutive items within the same row column or diagonal of a virtual display consisting of a 4×4 square matrix of locations. The absence of spatiotemporal organization or structure refers to the violation of these constraints within a sequence. The path-length of the sequence refers to the sum of the distance required to move a finger between item locations in any given sequence. Thus, different path-lengths could be featured by different sequences with the same number of items.

Our study was conducted on Guinea baboons (*Papio papio*). Baboons are among the species tested in the first studies on WM in primates (Jacobsen, 1936), where evidence for a selective involvement of frontal regions of the brain in delayed response tasks was provided. Baboons have a remarkable capacity of remembering thousands of stimulus–response associations in LTM (Cook & Fagot, 2009; Fagot & Cook, 2006).

The test protocol used in this study involved a new technique, where animals housed in social groups can spontaneously engage in self-testing without any need for the subjects to be removed from their group (Fagot & Paleressompoulle, 2009). This protocol allowed us to present the animals with a very large number of trials, as they could engage in self-testing at any time of day or night. The collection of an affluent database was considered critical for the parametric manipulation of several variables in our task, such as the structure of the sequences. Moreover, this protocol ensured that subjects were tested when they were highly motivated to engage with the task (since they would otherwise decide not to do so). We conjectured that a high motivation in the task would have beneficial effects on the ability of the subjects to sustain their attention during the presentation of the sequences. In a second experiment, we tested humans for direct comparison of the monkey and human data.

We recorded recall accuracy, in terms of both number of sequences correctly reported and of number of items correctly reported within sequences that contained some errors. These measures of accuracy were used to determine the memory span of the subjects and the presence of serial position effects. We also recorded RT to provide an additional measure of the relative ease of processing of a given type of sequence. With the use of these measures we aimed to verify whether or not the number of items, structure and path length played a role in the spatial serial recall of baboons and people.

2. Experiment 1: serial spatial recall in monkeys

2.1. Methods

2.1.1. Subjects

The subjects were two 22-year-old male Guinea baboons (*P. papio*) named B11 and B15. They were housed in a 3 m × 3 m indoor enclosure connected to an outdoor enclosure of the same size. The outdoor enclosure contained climbing structures and the automatic test system described below. All the baboons had a long experimental history as they had been previously tested in a variety of computerized tasks in which they had to respond using either a joysticks (Fagot & Cook, 2006; Fagot & Deruelle, 1997) or a touch screen (e.g., Fagot & Parron, 2010). They were however naive with regard to the CBTT task. The baboons were not food deprived, but obtained their normal ration of food (fruits, vegetables, and monkey chows) at approximately 5PM. Each subject has a micro-chip implanted in each forearm that is used to automatically identify the subject during the experiment.

2.1.2. Apparatus

Our research used a new automatic operant conditioning test device, named ALDM (Automated Learning Device for Monkeys). It allows testing the monkeys on a voluntary basis, while they are maintained in their social group, without any need to capture them. More details of this system can be found in Fagot and Paleressompoulle (2009); see also Fagot and Bonté (2010)). Briefly described, this system allows the identification of the subjects automatically when they enter a test chamber, thereby allowing them to self-test on a voluntary basis at any time of the day or night.

The apparatus comprised the test chamber (70 cm × 70 cm × 80 cm) which was freely accessible by the animals through an open back entrance. The test chamber was fitted in its innermost front side with a view port (7 cm × 7 cm) and two hand ports (8 cm × 5 cm). Looking through the view port allowed vision of a 17-in LCD touch monitor installed at eye level 25 cm from the port. Introducing one hand through one of the hand ports allowed 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 grains of dry wheat which were delivered inside the test booth by a home-made dispenser. The experiment was controlled by a customized test program developed by the first author with EPrime (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.

2.1.3. Stimuli

The stimuli were white squares 100 × 100 pixels in size, displayed on a black background. Their number and arrangement varied according to the procedure detailed below. Each square served as an item in the CBTT task. Thus, 3-, 4-, 5- and 6-item sequences implied the display of 3–6 squares on the screen, respectively. Location of the items on the screen was determined considering a virtual 4×4 virtual matrix of 16 possible locations. One hundred and sixty pixels separated the centre of two vertically or horizontally adjacent cells in the matrix. Considering the viewing distance, each item had a visual size of approximately 6 degrees of visual angle, and the visual size of the 4×4 matrix was 57.5 degrees.

2.1.4. Test procedure

A schematic presentation of the testing procedure is provided in Fig. 1. The trial started when the baboons introduced one hand through one of the two hand ports to trigger the fixation display. The first display consisted in a yellow cross (100 × 100 pixels), which was used as a "start button". It appeared on a black background in the central bottom part of the screen. Touching the start button triggered a series of 3 to 6 consecutive displays of 250 ms each, again on a black background. Each display during that observation period added an item on the screen, therefore defining 3- to 6-item recall sequences. The screen turned black during 250 ms once the full sequence of items had been shown on the screen. The, all the items defining the sequence reappeared simultaneously on the screen during the response period. The task for the subjects was to touch the items on the screen, one by one, in their order of appearance during the observation period. The baboon was food rewarded immediately after it had correctly reproduced the full sequence. It was penalized with a 3-s time-out indicated by a green screen in case of an incorrect response. That time-out was delivered after the subjects had sampled a number of items on the screen corresponding to the sequence length. An inter-trial interval of 3 s minimum separated two consecutive trials.

This was the interval corresponding to the period of time during which the self-identification procedure remained inactive following the completion of the previous trials. After these 3 s, the animal was able to spontaneously initiate the trial. The delay between two trials could thus be longer than 3 s. The testing procedure involved series of 92 trials blocks, which were continuously presented in random order during the time period allocated to this study. Three independent variables were used in the test design. The first was the number of items in the sequence. During testing, the sequences included 3, 4, 5 or 6 items. The second independent variable was the spatio-temporal organization of the sequence, with two levels: structured and unstructured. Structured sequences always involved the presentation of two consecutive sequence items within the same horizontal, vertical or diagonal axis of the matrix of 4×4 virtual cells. Unstructured sequences never had two consecutive items in the same horizontal, vertical or diagonal axis. The last independent variable is the length of the path. The range of the possible path lengths is provided in Table 1 for sequences of any given number of items. These ranges were

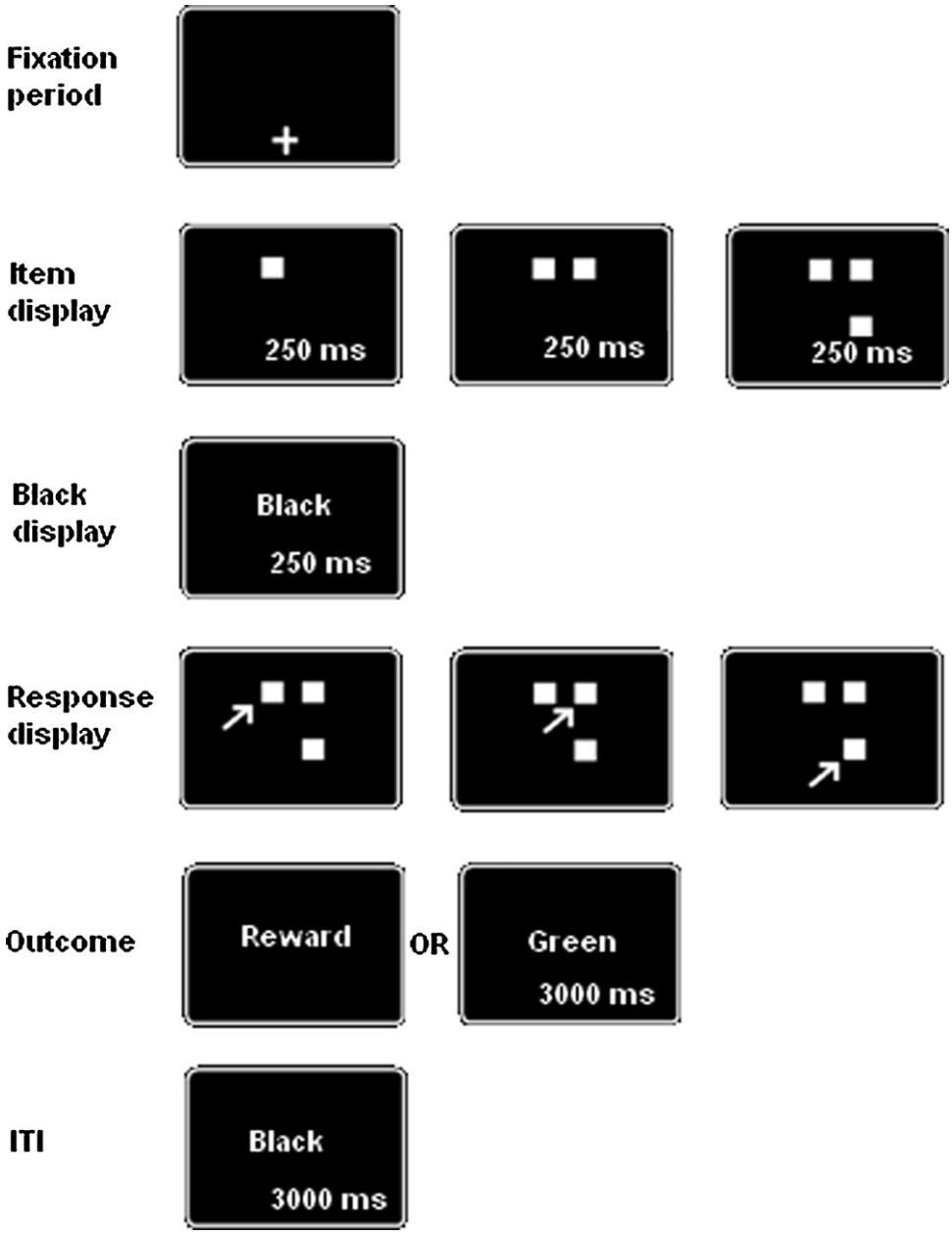

Fig. 1. Testing procedure.

determined so that possible overlaps between the short and long sequences were avoided, while allowing the production of a large number of different sequences in each condition. Examples of sequences for the different conditions are presented in Fig. 2.

Each test session comprised a total of 80 trials with 3-item sequences, randomly inter-mixed with four trials for each other possible number of items (i.e., 4, 5 and 6). For each possible number of items, there was an even number of trials per structure (structured, unstructured) and path length (short, long) conditions (see Fig. 2). Test sessions were continuously presented to the monkeys during approximately one month of self-testing. The use of the self-testing procedure allowed the collection of a very large number of test trials for the two monkeys. During testing, B15 made a total of 37,465 trials and B11 26,685 trials.

Table 1
Minimum and maximum distance (in pixels) of the trajectory connecting the different items in the sequence in the prescribed order.

Number of items	Min long	Max long	Min short	Max short
3	933	1154	706	901
4	1478	1736	1074	1451
5	1866	2290	1430	1857
6	2149	2478	1789	2122

2.1.5. Training procedure

The training procedure aimed at making the baboons learn to attend to sequences of 3 items and to recall them by touching the items on the screen in the required order. It involved several training phases of increasing complexities. At the beginning of the training, the baboons were requested to touch 1-item as it appeared on the screen and then to reproduce sequences of 2- and 3 items. The first, second and third item of the sequence were presented in a green, blue and yellow colour, respectively, during the initial phase of testing. Once performance was above 80% correct in three 80-trial sessions, a fading procedure was introduced to progressively remove the colour of the items until they were all displayed in plain white. Fading trials were presented until the baboons reached a learning criterion (80% correct over 3 consecutive sessions) with the white items. Removing the colour forced the subject to process the order of appearance of the items on the screen, and ruled out the use of non-spatial information in the identification of the items. In a final training phase, the baboons received series of 92-trial blocks similar to the testing procedure described above. In these trials, however, the stimulus display was removed from the screen as soon as the subject selected an item in the wrong order in the sequence. The training procedure lasted approximately $2^{1/2}$ months and involved approximately 46,500 training trials for B11 and 84,200 trials for B15.

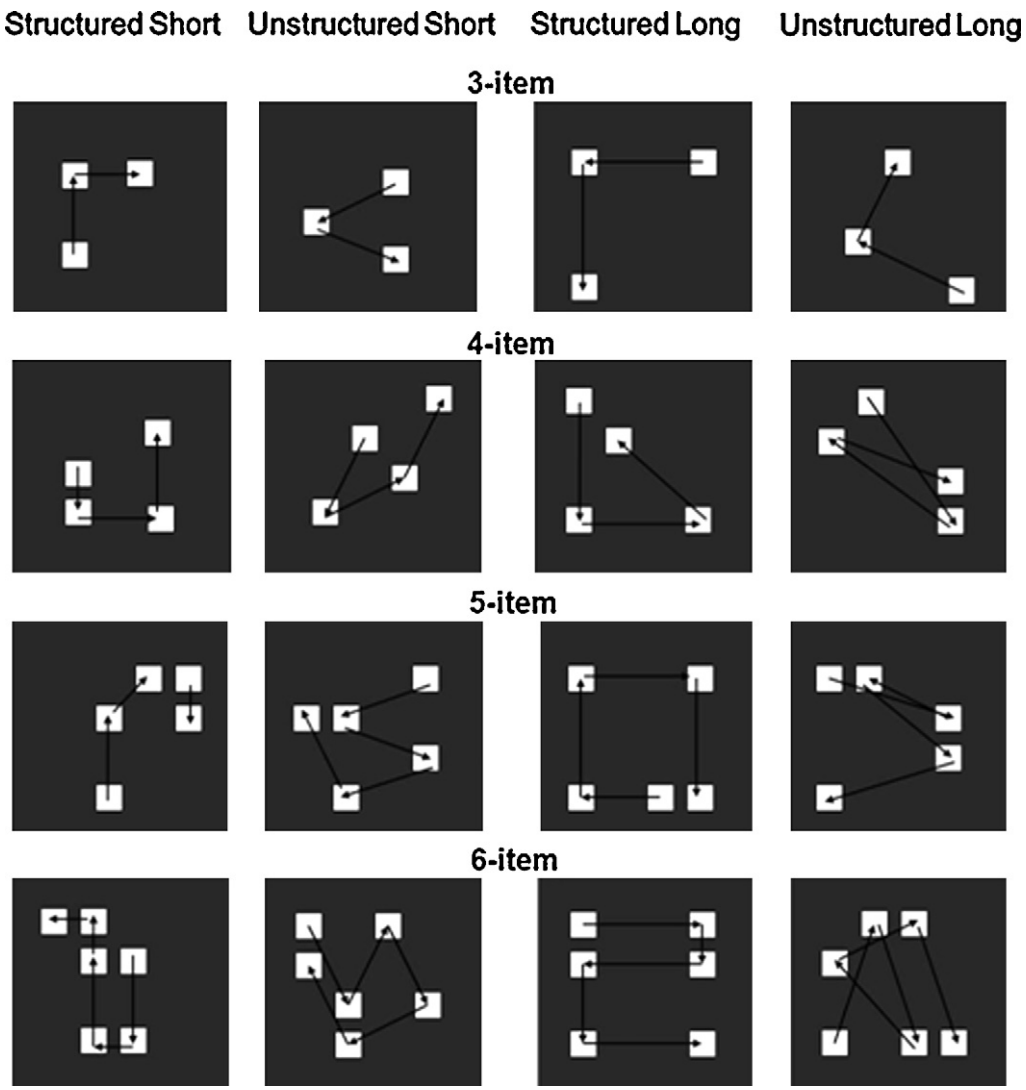

Fig. 2. Examples of sequences used in the different conditions of Experiment 1 and 2.

2.2. Results

2.2.1. Proportion of correct sequences

The proportion of correct sequences reported by each of the two subjects is presented in Table 2, along with the probability of reproducing correct sequences of a particular number of items by chance. This theoretical probability is extremely low, particularly for sequences with a higher number of items (Table 2).

Binomial tests were carried out to compare the frequency of sequences correctly reproduced by each subject to chance level. The following levels of expected probabilities were used for sequences of different number of items: 3-item = .1667, 4-item = .04167; 5-item = .0083, 6-item = .00138. They were obtained considering the probability of producing a correct sequence on the basis of all the possible sequences which could be generated for sequences of a given number of items. The analyses revealed that subject B11 was above chance with the sequences of 3 ($p < .001$) and 4 items ($p < .001$). Subject B15 was above chance with sequences of 3, 4, 5 items (all $ps < .001$).

In order to compare the different conditions, the database for each subject was divided in 10 equal blocks of trials. The proportion of sequences correctly observed in each block was then subjected to a 4 (number of items) × 2 (structure) × 2 (path length) repeated measures ANOVA. There was a significant effect of number of items for both B15 [$F(3, 27) = 580.42$, $p < .001$, $\eta_p^2 = .985$] and B11 [$F(3, 27) = 820.28$, $p < .001$, $\eta_p^2 = .989$]. None of the other main effects or interactions was significant. The proportion of correct sequences reproduced by the two monkeys is presented in Fig. 3 (left panel). Given the absence of significant effects, structured and unstructured sequences have been combined in the figure.

Table 2
Proportion of sequences correctly reported by the two baboons (Experiment 1) and by humans in Experiment 2.

N items	Chance proportion correct	B11	B15	Humans (Experiment 2)
3 items	.1667	.7287*	.6856*	.9724* (SE = .0056)
4 items	.0417	.2084*	.3063*	.9240* (SE = .0010)
5 items	.0083	.0094	.0297*	.7833* (SE = .03809)
6 items	.0014	.0000	.0030	.8126* (SE = .0190)

* Indicates $p < .001$.

Fig. 3. Correct sequences. Proportion of sequences correctly recalled by baboons ($N = 2$) in Experiment 1 (left hand panel) and by humans ($N = 24$) in Experiment 2 (right hand panel). Error bars = 1 SE. Black and white bars refer to long and short path length sequences in baboons and to structured and unstructured sequences in humans. This reflects the finding that the factor path length proved significant in baboons whereas the factor structure proved significant in humans (see text for details).

2.2.2. Serial position analysis

The frequency of items correctly reported at each serial position for sequences of different number of items are reported in Fig. 4 for the two monkeys B11 and B15.

Chi square tests with 5 degrees of freedom were carried out to compare the frequency of items correctly reported by each subject at each serial position of sequences different number of items. The results are reported in Table 3. In both subjects the Chi square tests revealed a significant overall difference between the serial positions of the items, in all sequences.

The presence of primacy and recency effects were assessed by 1 degree of freedom Chi square tests used to compare the frequency of correct responses recorded for consecutive items in each sequence. They revealed a clear primacy effect in both subjects. In fact, the first item, and occasionally the second item, was recalled more accurately than other items in the sequence. However, there was no evidence of a recency effect in either subject since the last items in the sequence were never recalled more accurately than previous items. The results of these tests are reported in the paragraph below for each subject. For brevity the following notations have been used. The sign ">" indicates a significantly ($p < .05$) higher frequency of correct responses and the sign "=" indicates the absence of a significant difference.

B 11. 3-item sequences: item 1 > item 2 = item 3; 4-item sequences: item 1 > item 2 > item 3 = item 4; 5-item: item 1 > item 2 = item 4 = item 5; 6-item sequences: item 1 > item 2 = item 3 = item 4 = item 5 = item 6. B15: 3-item sequences, item 1 > item 2 = item 3; 4 item sequences, item 1 > item 2 > item 3 = item 4; 5-item sequences, item 1 > item 2 > item 4 = item 5. A slightly different pattern of results emerged for sequences of 6 items: item 1 = item 2 > item 3 item > 4-item = item 5 = item 6. Although a flexion in the curve for the 6-item sequences that can be observed from Fig. 4 (left hand panel) suggests the presence of a recency effect for B15, the difference between the frequency correct for items 5 and 6 only approached significance add (Chi, df = 1, $CHI^2 = 3.78$, $p = .052$).

In order to assess the extent to which the subjects were above chance when reporting items at each serial position, one tailed binomial tests were carried out to compare the theoretical and the observed proportions of correct responses for each subject and each number of item conditions. In order to take into account the fact that subjects could not make repetition errors by selecting the same item twice, the theoretical proportions correct used for the analysis were .33, .25, .20 and .17 for the first item of sequences of 3, 4, 5 and 6 respectively. For the remaining items of sequences of 3, 4, 5 and 6 items, the proportions tested were .5, .33, .25 and .20. As shown in Fig. 4, B11 was successful at every serial position when reproducing 3-item sequences. It was also above chance at serial positions 1–3 for 4 item sequences, and at serial position 1 for sequences of 5 and 6 items. B15 was even more successful (Fig. 4) by reporting items above chance at all serial positions of sequences of 3 and 4 items. This subject was also above chance when selecting items 1–3 in 5 and 6-item sequences.

Table 3
Results of Chi square tests used to assess serial position effects for sequences of different number of items in each monkey tested in Experiment 1. Degrees of freedom (df), CHI^2 and p levels are reported for sequences of different number of items (3, 4, 5 and 6) for the two monkeys (B11 and B15).

Subject	*N*-items	df	CHI^2	$p <$
B11	3-items	2	259.57	.001
	4-items	3	114.02	.001
	5-items	4	57.28	.001
	6-items	5	21.04	.001
B15	3-items	2	605.35	.001
	4-items	3	326.23	.001
	5-items	4	377.75	.001
	6-items	5	166.50	.001

2.2.3. RT analysis

Only Response Times (RT) corresponding to items correctly recalled were included in the analysis. In order to ensure that the database was affluent enough even for sequences with 5 and 6 items, RT for correct items were included irrespectively of whether these were recorded in sequences that were reported correctly in their entirety or in sequences which contained some errors.

As for the accuracy analyses, we divided the database of each subject into ten blocks, and subjected those to the same 2 (structure) × 2 (path length) × 4 (number of steps in the sequences) ANOVA model. This analysis detected a strong effect of number of items in the sequence in both B11 [$F(3, 27) = 26.33$, $p < .001$, $\eta_p^2 = .75$] and B15 [$F(3, 27) = 14.08$, $p < .01$, $\eta_p^2 = .61$]. Consistently with the accuracy results, which indicated an increased difficulty for sequences with more steps, RT increased in parallel with the increase in the number of steps in the sequence in both subjects (Fig. 5).

None of the other factors or interactions proved significant in subject B11. By contrast, B15 showed a significant effect of path length [$F(1, 9) = 23.30$, $p < .01$, $\eta_p^2 = .72$] with sequences with a shorter path length (mean = 374.07, 95% CI = 355.3–392.80) being reproduced faster than sequences with a longer path length

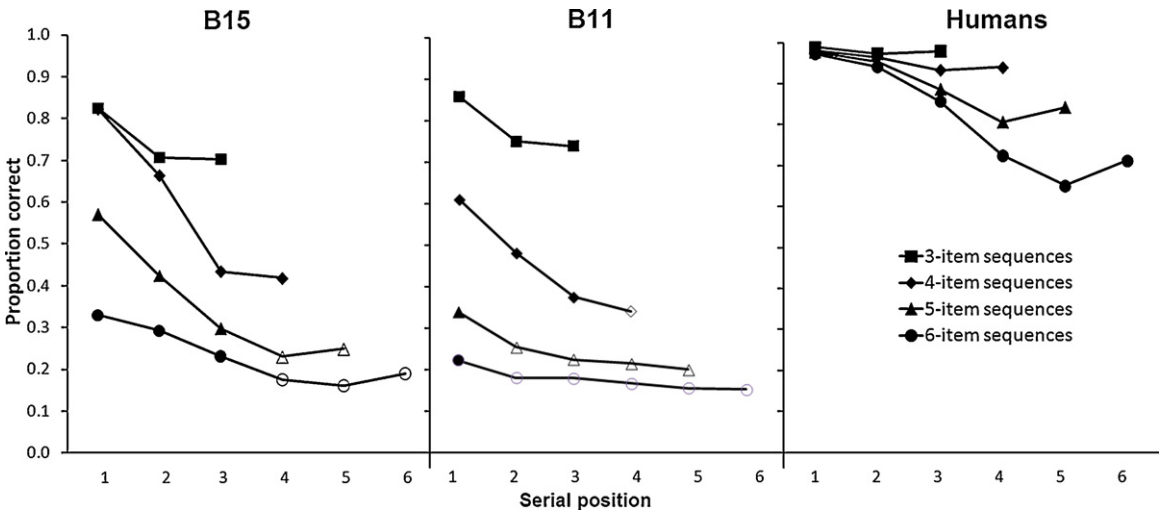

Fig. 4. Serial position curves. Proportion of items correctly reproduced in Experiment 1 at each serial position for sequences of 3, 4, 5 and 6 items by the two baboons (B15, left hand panel; B11, middle panel) in Experiment 1 and by the group of humans in Experiment 2 (right hand panel). Filled symbols indicate an above chance level of accuracy as determined by binomial tests (see text for details). Empty symbols indicate accuracy at either chance or below chance level.

(mean = 397.88, 95% CI = 379.09–416.67). In B15 the interactions structure by path length [$F(1, 9) = 8.19$, $p < .05$, $\eta_p^2 = .48$] and structure by number of steps in the sequence [$F(1, 9) = 9.58$, $p < .001$, $\eta_p^2 = .52$] were significant. Planned comparisons indicated that the interaction structure by path length was explained by the fact that only in sequences with a long path length RTs were faster [$t(9) = 3.18$, $p < .05$) for unstructured (UL mean = 371.29, 95% CI = 353.23–389.35) than for structured (SL mean = 410.70, 95%CI = 389.47–431.93) sequences. By contrast, no difference in RT in relation to structure was observed for sequences characterized by a shorter path length. The interaction between structure and path length can be explained by the fact that there was a significant difference between structured and unstructured sequences, with unstructured sequences always reported faster [$4.25 < t(9) < 5.9$, all $ps < .01$] than structured sequences, for sequences of 3, 4 and 5 steps [Structured sequences means: 3 step = 351.87, SD = 37.8; = 4 step = 394.13, SD = 28.89; 5 step = 420.26, SD = 37.31; Unstructured sequences means: 3 step = 341.04, SD = 41.09; 4 step = 361.09, SD = 22.94; 5 step = 392.8, SD = 29.66], but not for sequences of 6 steps. The lack of statistical significance for the difference between unstructured and structured sequences of 6 steps can be possibly explained by the relatively small number of correct items reported for sequence length 6 and the resulting paucity of the database.

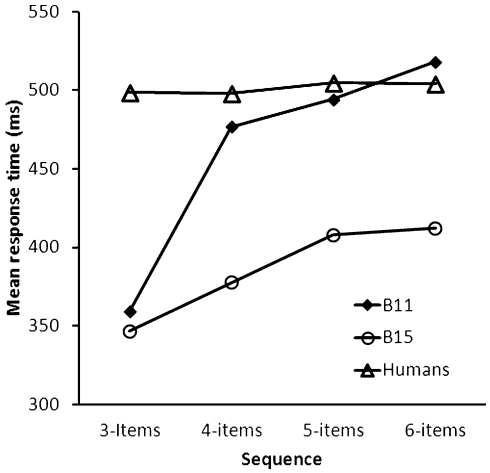

Fig. 5. RT. Mean RT for items correctly recalled by two baboons in Experiment 1 and by the human subjects in Experiment 2. The horizontal axis represents the number of items in the sequence (3–6). The factors path length and structure have been combined in the graph.

2.3. Discussion

Experiment 1 indicates that baboons are able to cope with an immediate serial spatial recall task. One monkey had a span of 4 (B11) and the other of 5 items (B15), as indicated by the above chance frequencies of correct sequences respectively reproduced by the two subjects. Monkeys' recall was affected by the number items, but not by the structure or the path length of the sequence.

The analysis of the RT provided a useful complement to the results obtained for the accuracy measures. The fact that an increase in RT was observed in parallel with the increase in the number of steps in the sequences corroborates the finding that subjects find sequence reproduction more challenging when the sequence contained more items. This result seems to indicate that the monkeys were affected by global properties of the sequence to be reported.

RT data also indicated a clear effect of path length with a slower average RT in sequences with longer paths. A positive effect on recall, deriving from the presence of structure in the sequence however was absent from the RT results, as it was in the accuracy analysis. Given these unexpected results in relation to the structure of the sequence in relation to previous results reported in the human literature (Bor et al., 2003; De Lillo, 2004; De Lillo & Lesk, 2010), it was important to assess how humans behaved in a task as similar as possible to the one used with the monkeys in Experiment 1. Therefore, we ran a second experiment with humans using the same task as that employed with the monkeys in Experiment 1.

3. Experiment 2: serial spatial recall in humans

3.1. Method

3.1.1. Participants

Twenty four undergraduate (15 females and 9 males) students from the University of Leicester with an average age of 20.08 years (SD = 3.7) took part in the study. Participants had normal or corrected-to-normal vision and received course credits for taking part.

3.1.2. Apparatus, design and procedure

The apparatus was A PC equipped with a touch-screen, similar to that used in Experiment 1. The same E-prime program and conditions of Experiment 1 were used. However, in Experiment 2 the same number of trials was used for each condition.

A total of 320 trials were administered in two testing sessions, separated by a short break that lasted until the participant felt comfortable to start testing again.

3.2. Results

3.2.1. Accuracy: proportion of sequences correctly reproduced

The proportion of sequences correctly reproduced by humans in Experiment 2 is reported in Fig. 3. From the figure, it can be observed that humans showed a higher level of performance than monkeys B11 and B15. However, the same pattern of results, characterized by a relative decline of performance for the sequences with the highest numbers of items, can be observed in both species. A 4 (number of items) × 2 (structure) × 2 (path length) repeated measures ANOVA carried out on the data obtained with humans in this experiment indicated a significant main effect only for the number of items [$F(3, 69) = 114.46$, $p < .001$, $\eta_p^2 = .83$] and structure [$F(1, 23) = 8.88$, $p < .01$, $\eta_p^2 = .28$], with a higher recall for structured sequences. Only the interaction number of items by structure was significant [$F(3, 69) = 3.69$, $p < .05$, $\eta_p^2 = .14$]. Planned comparisons indicated that this interaction was explained by an advantage for structured sequences in the 6-item sequences only [$t(23) = 3.66$, $p < .001$], which suggests that the benefit of structuring emerges when memory load increases.

3.2.2. Serial position analysis

The proportion of items correctly recalled by humans at each serial position is shown in the right hand panel of Fig. 4. The graph suggests a similarity in serial position functions in humans and monkeys. Nevertheless, humans showed a much higher level of performance than that shown by the baboons.

One way repeated measures ANOVAs revealed a serial position effect at each level of sequence length [3-step sequences, $F(2, 46) = 6.55$, $p < .01$, $\eta_p^2 = .222$; 4-step, $F(3, 69) = 17.31$, $p < .001$, $\eta_p^2 = .43$; 5-step, $F(4, 92) = 41.75$, $p < .001$, $\eta_p^2 = .65$; 6-step, $F(5, 115) = 52.59$, $p < .001$, $\eta_p^2 = .70$]. Post-hoc analyses (Bonferroni corrected) for sequences of 3 ordinal steps revealed a primacy effect as item 1 was recalled better than item 2 ($p < .01$). In sequences of 4 ordinal steps, item 1 was recalled better than items 2, 3 and 4, and item 2 was recalled better than item 3 (all $ps < .01$). In sequences of 5 ordinal steps, item 1 was recalled better than items 2, 3, 4 and items 2 was recalled better than items 3, 4 and 5. Item 3 was recalled more accurately than item 4 (all $ps < .01$). At this level of sequence length, a clear recency effect also started to emerge as item 5 was recalled at a higher level of accuracy than item 4 ($p < .01$). Similarly, a recency effect emerged in sequences of 6 ordinal steps. A similar recency effect was also observed in Experiment 1 in baboon B15 (Fig. 4). Also, as observed in Experiment 1 with baboons, humans showed a clear primacy effect. In fact, item 1 was recalled better by humans than any other item in the sequence (all $ps < .001$), item 2 was recalled better than items 3, 4, 5 and 6 (all $ps < .001$), item 3 was recalled better than items 4, 5 and 6 (all $ps < .001$) item 4 was recalled better than item 5 ($p < .001$) but item 6 was recalled at a higher level of accuracy than item 5 ($p < .001$).

One sample t-tests carried out on the group of 24 subjects revealed that humans recalled items at any serial position above chance level [$17.46 < t(23) < 202.38$, all $ps < .001$], regardless of the number of steps in the sequence

3.2.3. RT analysis

A repeated measure 2 (structure) × 2 (path length) × 4 (number of steps in the sequences) ANOVA revealed, a main effect of number of steps in the sequences [$F(3, 69) = 12.01$, $p < .001$, $\eta_p^2 = .34$], which was similar to that observed in monkeys in Experiment 1 but reduced in amplitude. This effect was mostly due to RTs that were slightly longer (by approximately 6 ms) for the two sequences with the highest number of items (see Fig. 5). As shown in Fig. 6a, and similarly to what was observed in one of the baboons (B15), path length affected RTs in humans. In fact, a significant effect of path length [$F(1, 23) = 23.98$, $p < .001$, $\eta_p^2 = .51$] emerged in Experiment 2, with sequences with a shorter path length being reproduced faster (mean = 475.70 ms, CI = 447.77–503.62 ms) than sequences with a long path length (mean = 493.57 ms, CI = 464.66–522.47 ms). In contrast with baboons, however, humans also showed a significant effect of structure [$F(1, 23) = 15.60$, $p < .001$, $\eta_p^2 = .40$). The presence of structure produced faster mean RT (mean structured = 476 ms, CI = 450.50–503.00 ms) than its absence (mean unstructured = 492.52, CI = 462.00–523.04, see Fig. 6b). The only significant interaction observed in humans was between the factors structure and number of items. Planned comparison (Bonferroni corrected) revealed that this was explained by a strong effect of structure in four-step sequences, $t(23) = 4.68$, $p < .001$, which was absent in sequences with a different number of steps.

3.3. Discussion

Experiment 2 demonstrates that our procedure is suitable for the comparison of WM in baboons and humans. Humans showed an overall higher performance than that observed in baboons in Experiment 1. Nevertheless, there were some similarities between the two species. The accuracy results of humans showed an effect of number of items that was also observed in baboons. Overall, the serial position curves were similar in the two species. The effects were sometimes attenuated in humans due to a ceiling effect deriving from their very high level of performance. Score analysis in Experiment 2 showed an advantage for the structured sequences in humans which was confined to 6-item sequences. Again ceiling

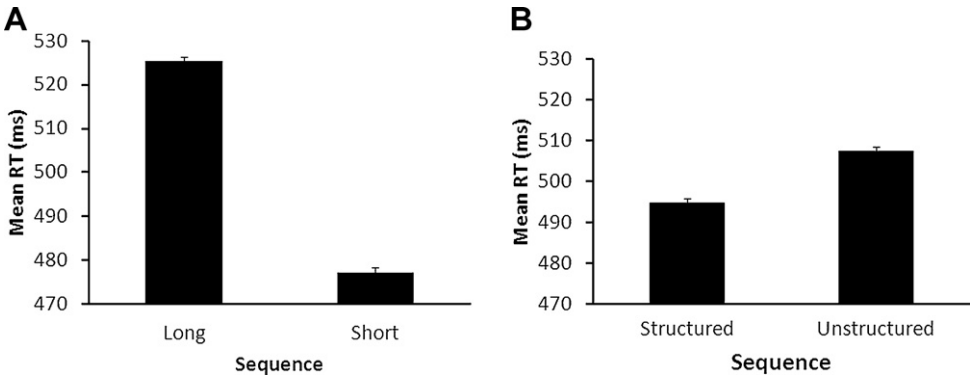

Fig. 6. Humans: RT. Pattern of RT observed in Experiment 2. (A) Average RT recorded for long and short sequences, when the two levels of structure are combined; (B) average RT for structured and unstructured sequences, when the two levels of path length are combined.

effects may have masked the effect of structure in sequences with a smaller number of items. In humans, longer RTs were recorded during the recall of items within sequences with a long path. This is in accordance with results that have indicated such an effect in humans tested with a similar task (Parmentier et al., 2005). One of the baboons also showed this effect, and in the same direction. An effect of structure was also present in the RT data collected on humans. Our participants responded faster on average when reproducing structured sequences. By contrast, none of the monkeys showed a reduction of RT in presence of structure in the sequence. If anything, there was a negative relationship between the presence of structure and speed of responding in one of the baboons (B15). The theoretical relationship between speed of responding and number of items is more difficult to pinpoint as even in humans it emerged mainly with sequences of four items.

4. General discussion

This study shows that the investigation of WM in monkeys can be meaningfully carried out using an immediate serial recall task analogous to the CBTT, widely used in human neuropsychology (Berch et al., 1998). The use of such task with non-human primates is important for a detailed characterization of the memory systems of monkeys, and for comparative studies with humans.

The demonstration that monkeys can be studied with a variation of the CBTT fills a significant gap in the literature, which so far lacks reports of WM skills in non-human primates measured with tasks that require the temporary holding in memory of a series of ordered responses.

To our knowledge only one study with a single macaque has previously used an immediate serial recall task in monkeys similar to the one featured here (Botvinick et al., 2009). That study was however limited, because there were no sequences longer than four items and it did not feature controls for possible effects of structure. In addition, that study did not feature a direct comparison with humans tested in similar conditions.

In our study, we were able to train baboons to attend to serial displays of up to 6 spatial items, and to attempt to recall them. The recall above chance level of sequences of 4 and 5 items for the two monkeys, respectively, suggest that the baboons may have an immediate serial recall span of at least 4 items. Nevertheless, the analysis of the serial position curves for longer sequences of 5 and 6 items, indicates that one of the animals started to perform below chance at serial position 4. The other animal recalled above chance only the item at the first serial position in sequences of 5 and 6 items. This suggests that global properties of the sequence, such as the total number of ordinal steps it features, affect the way in which items are remembered by baboons. The longer the sequence, the more likely it is that the monkeys will start making mistakes early on during their reproduction. This additionally suggests that when baboons were presented with longer sequences, they were trying to remember the whole sequence, rather than coding for the items within memory span and starting rehearsing those as soon as their memory buffer reached its capacity limit.

Our unique procedure featuring the self-testing of the animals (Fagot & Paleressompoulle, 2009) enabled us to collect an extremely large number of trials. Thus, these span values are likely to be the truthful upper limits of the WM capacity in baboons, at least under the task parameters featured here. This is particularly true considering that effects of task practice on visuo-spatial working memory have now been reported in the human literature (Olesen, Westerberg, & Klingberg, 2004). In the case of the baboons tested here, it seems that even after an extensive training their immediate serial spatial span does not match that of humans tested in similar conditions. This gap between the recall performance in the two species could possibly become wider if it were feasible to train humans with the tens of thousands of trials used with monkeys in the present study. There are a number of possible explanations for the overall higher proportion of correct responses shown by humans. One first explanation for the score differences between humans and baboons might be that humans have a larger memory capacity. If so, this difference in memory capacity for unrelated material in humans and monkeys is likely to be confined to short term retention because recent studies of reference memory for arbitrary parings of stimuli and responses have shown that baboons can remember thousands of such associations (Cook & Fagot, 2009; Fagot & Cook, 2006). This capacity is even larger than that reported in one human tested in similar conditions (Vos, 2009). Moreover the limit of human WM capacity for unrelated items in human is now considered to be much smaller than suggested by earlier studies (see Cowan, 2001, 2005 for reviews) and it has been estimated to be in the region of 4. Nevertheless, although not dramatically so, this value is still higher than the span we observed at least in one baboon in the present study. It is nevertheless possible that a larger memory capacity of humans may be the result of top-down influences mediated by the DLPF that recent models suggest could boost memory, and be the basis of inter-individual differences in WM (Edin et al., 2009). Interspecies differences such as those observed here could be mediated by similar mechanisms and, if so, this could explain the increased performance in humans irrespectively of differences that may be related to strategic use of information as discussed below.

A similarity between humans and baboons that emerged in our study is the presence of an effect of path length in humans, effects of path length in CBTT tasks have been put in relation to mental imagery (Parmentier et al., 2005). Longer scan time would be required to rehearse sequences characterized by a longer path length, resulting in reduced refresh rate of the representation and consequent increased memory decay. Our results indicate that this explanation may pertain to monkeys too. In fact baboons also show properties of mental rotations (Vauclair, Fagot, & Hopkins, 1993) suggesting that they may have the ability to form mental images of the array in the present task.

Another important factor which has received considerable attention in WM research is the role of organizational factors (Baddeley, 2000; Bor et al., 2003; De Lillo, 2004; Stuss et al., 1994). Cowan (2001) argued that the magical number 7 ± 2, previously considered the upper limit of human mental capacity, is an expression of both capacity and organizational factors leading to chunking. This number would be reduced to 4 when care is taken to prevent the use of implicit organizational constraints in the material to be remembered. It is possible that humans are able to detect forms of organization, even in the unstructured sequences, and that enables them to reproduce the sequences more accurately than the monkeys. It is also possible that monkeys were more likely than humans to make mistakes when pressing the items due to a less precise reach, apart from memory capacity.

A difference emerged between the two species in the effect of structure. Humans showed some advantage for structured sequences that was not observed in monkeys. Bor et al. (2003) have shown that the ability to use the structure of the sequence as a memory aid is related to the activation of the DLPF. As such, frontal functions appear to be the most likely candidates for the explanation of the differences between humans and non-human primates in this domain. In the present study, only the items pertaining to the sequence to be reported were presented. This was done to avoid contamination between item memory and serial order (see De Lillo & Lesk, 2010; Kemps, 1999 for a discussion of differences between item memory and memory for serial order). We suspect that the presentation of the full matrix of items from the outset (with a subset of them flashing in order) may induce a clearer effect of structure in monkeys. Nevertheless, as it stands, our study seems to indicate

that the ability to pick up structure may be an important difference between humans and monkeys. This would be in accordance to a number of studies carried out in the realm of perceptual organization and global local processing in perception, which indicate that humans are more likely to benefit from grouping parts according to gestalt rules, to find parts of stimuli in embedded figure tasks (Spinozzi et al., 2004) and show a higher propensity than monkeys to extract the global and configural parts of visual stimuli (De Lillo, Spinozzi, Truppa, & Naylor, 2005; Fagot & Deruelle, 1997; Spinozzi, De Lillo, & Truppa, 2003).

A number of studies have clearly shown that human participants code items in the CBTT in a configural way (Avons, 2007; Avons & Trew, 2006; Boduroglu & Shah, 2006). For example, Avons (2007) performed experimental manipulations involving different displacements of Corsi displays during testing. Avons showed that serial recall in humans is affected exclusively by manipulations that distort the relationship between items in the display. Displacements of the items that preserve such relationships do not affect recall. Considering the propensity of monkeys to process visual displays with a local bias (De Lillo et al., 2005; Fagot & Deruelle, 1997; Spinozzi et al., 2003, 2004) it is plausible that whereas humans have a tendency to process the items in relation to each other, monkeys may have a tendency to process each item regardless of its relative position to other items within the display used in this study. By doing so, they would fail to detect some of the structural affordances of the set. Caution would be needed with this interpretation of the results since monkeys do not cope with longer sequences as well as people and this may have also affected their chances of benefitting from the presence of structure in the sequences. Nevertheless, the assessment of the possibility that there are differences in the use of organizational principles in monkeys and people would be an important line of investigation to follow. It could provide important insights into qualitative differences between human and non-human primates' ability to track positions in space. Given that the role of organizational factors, rather than memory span per se, has proved important for the detection of deficits in patients with frontal lobe injuries in relation to controls (Stuss et al., 1994), a detailed evaluation of this possibility would be of great value in experimental and clinical neuropsychology.

Acknowledgments

JF acknowledges the support of grant "Project Ethique, Volet Recherce Exploratoire, 2008" from the Conseil Regional Provence-Alpes-Cotes d'Azur and the ANR-2010-BLANC-1908-01. CDL gratefully acknowledges a semester of study leave granted by the University of Leicester. We are grateful to Georgina Pole for her help with the collection of data on human participants, and Marc Martin for managing the monkey colony.

References

Avons, S. E. (2007). Spatial span under translation: A study of reference frames. *Memory and Cognition*, *35*, 402–417.
Avons, S. E. & Trew, L. (2006). What kind of space is remembered in spatial span? *Cognitive Processing*, *7*(Suppl. 1), S157–S159.
Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
Baddeley, A. D. (2001). Is working memory still working? *American Psychologist*, *56*(11), 849–864.
Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.
Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation*. London: Academic Press.
Berch, D. B., Krikorian, R. & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition*, *38*, 317–338.
Boduroglu, A. & Shah, P. (2006). Configural representations in spatial working memory. *Cognitive Processing*, *7*(Suppl. 1), S160–S162.
Bor, D., Duncan, J., Wiseman, R. J. & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*, *37*, 361–367.
Botvinick, M. M., Wang, J., Cowan, E., Roy, S., Bestianen, C., Mayo, J. P., et al. (2009). An analysis of immediate serial recall performance in a macaque. *Animal Cognition*, *12*, 671–678.
Butters, N. & Pandya, D. N. (1969). Retention of delayed-alternation: Effect of selective lesions of sulcus principalis. *Science*, *165*, 1271–1273.
Chen, S., Swartz, K. B. & Terrace, H. S. (1997). Knowledge of ordinal position of list items in rhesus monkeys. *Psychological Science*, *8*, 80–86.
Cook, R. G. & Fagot, J. (2009). First trial rewards promote 1-trial learning and prolonged memory in pigeon and baboon. *Proceedings of the National Academy of Sciences*, *106*, 9530–9533.
Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International*, *34*(02), 891B (University microfilms No. AA105-77717).
Cowan, N. (2001). The magic number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
Cowan, N. (2005). *Working memory capacity*. Hove: Psychology Press.
D'Amato, M. R. & Colombo, M. (1988). Representation of serial order in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behaviour Processes*, *14*, 131–139.
D'Amato, M. R. & Colombo, M. (1989). Serial learning with wild carditems by monkeys (*Cebus apella*): Implications for knowledge of ordinal position. *Journal of Comparative Psychology*, *103*, 252–261.
De Lillo, C. (1996). The serial organization of behaviour by non-human primates: An evaluation of experimental paradigms. *Behavioural Brain Research*, *81*, 1–17.
De Lillo, C. (2004). Imposing structure on a Corsi-type task: Evidence for hierarchical organisation based on spatial proximity in serial–spatial memory. *Brain and Cognition*, *55*, 415–426.
De Lillo, C. (in press). The use of spatial structure in working memory: A comparative standpoint. In D. McFarland, M. McGonigle, & K. Stenning (Eds.), *The complex mind*. Basingstoke: Palgrave.
De Lillo, C. & Lesk, V. (2010). Spatial clustering and hierarchical coding in immediate serial recall. *European Journal of Cognitive Psychology*, *22*(2), 216–246.
De Lillo, C., Visalberghi, E. & Aversano, M. (1997). The organisation of exhaustive searches in a patchy space by capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *111*(1), 82–90.
De Lillo, C., Aversano, M., Tuci, E. & Visalberghi, E. (1998). Spatial constraints and regulatory functions in monkeys' (*Cebus apella*) search. *Journal of Comparative Psychology*, *112*(4), 353–362.
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 and monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *119*, 155–165.
Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegner, J. & Comte, A. (2009). Mechanisms of top-down control of working memory capacity. *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas/0901894106
Fagot, J. & Bonté, E. (2010). Automated testing of cognitive performance in monkeys: Use of a battery of computerized tes system by a troop of semi-free ranging baboons. *Behavioral Research Methods*, *42*, 507–512.
Fagot, J. & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences*, *103*, 17564–17567.
Fagot, J. & Deruelle, C. (1997). Processing of global and local visual information and hemispheric specialisation in humans (*Homo sapiens*) and baboons (*Papio papio*). *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 429–442.
Fagot, J. & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behavior Research Methods*, *41*, 396–404.
Fagot, J. & Parron, C. (2010). Relational matching in baboons (*Papio papio*) with reduced grouping requirements. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*, 184–193.
Gathercole, S. E., Pickering, S. J., Knight, C. & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16.
Goldman-Rakic, P. S. (1992). Working memory and the mind. *Scientific American*, *267*, 111–117.
Gross, C. G. & Weiskrants, L. (1962). Evidence for dissociation of impairment on auditory discrimination and delayed response following lateral frontal lesions in monkeys. *Experimental Neurology*, *4*, 453–476.
Inoue, S. & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, *17*(23), 1004–1005.
Jacobsen, C. F. (1936). Studies of cerebral function in primates. *Comparative Psychology Monographs*, *13*, 1–68.
Jaensch, E. R. (1930). *Eidetic imagery and typological methods of investigation* (2nd ed.). New York: Harcourt, Brace and Co.
Kawai, N. & Matsuzawa, T. (2000). Numeric memory span in a chimpanzee. *Nature*, *403*, 39–40.
Kemps, E. (1999). Effects of complexity on visual–spatial working memory. *European Journal of Cognitive Psychology*, *11*, 335–356.
Kyllonen, P. C. & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences*, *2*, 129–160.
Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, *315*, 57–59.
Matzke, S. M. & Castro, C. A. (1998). Primacy and recency effects in rhesus monkeys (*Macaca mulatta*) using a serial probe recognition task III. A developmental analysis. *Developmental Psychobiology*, *32*, 215–224.
Milner, B. (1971). Interhemispheric differences in the localisation of psychological processes in man. *Cortex*, *27*, 272–277.

Mishkin, M. (1957). Effects of small frontal lesions of delayed alternation in monkeys. *Journal of Neurophysiology*, *20*, 615–622.
Mishkin, M. & Manning, F. J. (1978). Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Research*, *143*, 313–323.
Mishkin, M., Vest, B., Waxler, M. & Rosvold. (1969). A re-examination of the effects of frontal lesions on object alternation. *Neuropsychologia*, *7*, 357–363.
Olesen, P. J., Westerberg, H. & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*, 75–77.
Parmentier, F. B. R., Elford, G. & Mayberry, M. (2005). Transitional information in spatial serial memory: Path characteristics affect recall performance. *Journal of Experimental psychology: Learning, memory and Cognition*, *31*, 412–427.
Petrides, M. (1998). Specialised systems for the processing of mnemonic information within the primate frontal cortex. In A. C. Roberts, T. W. Robbins, & L. Weiskrantz (Eds.), *The Prefrontal Cortex: Executive and Cognitive Functions* (pp. 103–116). Oxford: Oxford University Press.
Robbins, T. W., James, M., Owen, A. M. & Sahakian, B. J. (1994). Cambridge Neuropsychological Test Automated Batttery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia*, *5*, 266–281.
Sands, S. F. & Wright, A. A. (1980). Serial probe recognition performance by a rhesus monkey and a human with 10- and 20-item lists. *Journal of Experimental Psychology: Animal Behavior Processes*, *6*, 386–396.
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.
Spinozzi, G., De Lillo, C. & Castelli, S. (2004). Detection of "Grouped" and "Ungrouped" parts in visual patterns by tufted capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *118*(3), 297–308.
Spinozzi, G., De Lillo, C., Truppa, V. & Castorina, G. (2009). The relative use of proximity, shape similarity, and orientation as visual perceptual grouping cues in tufted capuchin monkeys (*Cebus apella*) and humans (*Homo sapiens*). *Journal of Comparative Psychology*, *123*, 56–68.
St Clair-Thompson, H. L & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*(4), 745–759.
Stuss, D. T., Alexander, M. P., Palumbo, C. L., Buckle, L., Sayer, L. & Pogue, J. (1994). Organisational strategies of patients with unilateral and bilateral frontal lobe injury in word list learning tasks. *Neuropsychology*, *8*, 355–373.
Swartz, K. B. & Terrace, H. S. (1991). Serila learning by rhesus monkeys: 1. Acquisition and retention of multiple four-item lists. *Journal of Experimental Psychology: Animal behavior Processes*, *17*, 396–410.
Tavares, M. C. H. & Tomaz, C. (2002). Working memory in capuchin monkeys (*Cebus apella*). *Behavioural Brain Research*, *131*, 131–137.
Terrace, H. S. (1993). The phylogeny and ontogeny of serial memory: List learning by pigeons and monkeys. *Psychological Science*, *4*, 162–169.
Terrace, H. S. & McGonigle, B. O. (1994). Memory and representation of serial order by children, monkeys and pigeons. *Current Directions in Psychological Science*, *3*, 180–185.
Vandierendonck, A., Kemps, E., Fastane, M. C. & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, *95*, 57–79.
Vauclair, J., Fagot, J. & Hopkins, W. D. (1993). Rotation of mental images in baboons when the visual input is directed to the left cerebral emisphere. *Psychological Science*, *4*, 99–103.
Vos, J. L. (2009). Long-term associative memory capacity in man. *Psychonomic Bulletin & Review*, *16*, 1076–1081.
Wright, A. A. (1999a). Auditory List Memory and Interference Processes in Monkey. *Journal of Experimental psychology: Animal Behavior Processes*, *25*, 284–296.
Wright, A. A. (1999b). Visual list memory in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *113*, 74–80.
Wright, A. A & Rivera, J. J. (1997). Memory of auditory lists by rhesus monkeys (*Macaco mulatto*). *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 1–9.
Wright, A. A., Santiago, H. C. & Sands, S. F. (1984). Monkey memory: Same/different concept learning, serial probe acquisition, and probe delay effects. *Journal of Experimental Psychology: Animal Behavior Processes*, *10*, 513–529.
Wright, A. A., Santiago, H. C., Sands, S. R., Kendrick, D. R. & Cook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys, and people. *Science*, *229*, 287–289.