Chimpanzees flexibly update working memory contents and show susceptibility to distraction in the self-ordered search task.

**Authors:** Christoph J. Völter\textsuperscript{a,b,c,*}, Roger Mundry\textsuperscript{d}, Josep Call\textsuperscript{a,b}, Amanda M. Seed\textsuperscript{a}

\textsuperscript{a} University of St Andrews, School of Psychology & Neuroscience, St Andrews, UK

\textsuperscript{b} Max Planck Institute for Evolutionary Anthropology, Department of Developmental and Comparative Psychology, Leipzig, Germany

\textsuperscript{c} Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna, Vienna, Austria

\textsuperscript{d} Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

\*Corresponding author: Christoph Völter, Email: christoph.voelter@vetmeduni.ac.at
Abstract

Working memory (WM) is a core executive function that allows individuals to hold, process, and manipulate information. WM capacity has been repeatedly nominated as a key factor in human cognitive evolution; nevertheless, little is known about the WM abilities of our closest primate relatives. In this study, we examined signatures of WM ability in chimpanzees (Pan troglodytes). Standard WM tasks for humans (Homo sapiens) often require participants to continuously update their WM. In Experiment 1, we implemented this updating requirement in a foraging situation: zoo-housed chimpanzees (N=13) searched for food in an array of containers. To avoid redundant searches, they needed to continuously update which containers they had already visited (similar to WM paradigms for human children) with 15-s retention intervals in between each choice. We examined chimpanzees’ WM capacity and to what extent they used spatial cues and object features to memorize their previous choices. In Experiment 2, we investigated how susceptible their WM was to attentional interference, an important signature, setting WM in humans apart from long-term memory. We found large individual differences with some individuals remembering at least their last four choices. Chimpanzees used a combination of spatial cues and object features to remember which boxes they had chosen already. Moreover, their performance decreased specifically when competing memory information was introduced. Finally, we found that individual differences in task performance were highly reliable over time. Together, these findings show remarkable similarities between human and chimpanzee WM abilities despite evolutionary and life history differences.

Keywords: short-term memory, working memory, chimpanzees, executive functions, primate cognition, cognitive control
1. Introduction

Working memory (WM) is thought to be one of the key ingredients of human intellect, allowing individuals to hold, process, and manipulate information in support of problem solving, perhaps requiring or giving rise to conscious experience [1]. It has been suggested that changes in working memory might have been an important factor in human evolution [2-4], resulting in qualitative differences between humans and nonhuman apes, for example, enabling complex stone tool manufacture [4, 5].

However, the extent to which human and nonhuman WM differs requires further investigation [6]. Most evidence to date for an increase in working memory capacity in human evolution comes from analyses of hominin stone artefacts [2-4]. Even though this line of research provides important hints, trying to specify which cognitive difference might explain differences in the material end-products of a behavioral sequence is problematic [7]. A more direct evaluation of the hypothesis is needed, for example by comparing humans to our closest living relatives.

What are the key characteristics of human WM, and how do they compare to findings about nonhuman primates? Of course, humans can rely on language to rehearse information, which might be the most straight-forward difference between human and non-human WM [8, 9]. Nevertheless, there is evidence that non-human primates engage in the active maintenance of information over a delay, from research looking at how performing a secondary task disrupts memory performance. WM is closely related to executive control of attention [10], and so susceptibility to attentional interference is an important signature of WM engagement. WM measures that involve resisting interference are highly correlated with general intelligence, while short term memory tasks without this component are not [11]. Evidence for susceptibility to attentional interference was found in studies with rhesus macaques (Macaca mulatta) performing delayed matching tasks on a computer. When simpler strategies such as relative
familiarity could not be used (for example because all of the stimuli involved in the memory task were equally familiar) performance was disrupted, but performance on versions of the task that could be solved through simpler memory processes was not (for example if novel images were used as stimuli, so that the target stimulus 'stood out' as the match by virtue of its relative familiarity) [12, 13].

Human WM is also characterized by a capacity limit; though the source of this limit is controversial [14]. WM capacity in adult humans corresponds to 3 – 5 pieces of information when processing-related strategies such as chunking or rehearsal are prevented [15]. Studies comparing humans and rhesus monkeys on the same task found a decline in performance with increasing memory load in both species but a larger WM capacity in humans than in rhesus monkeys [16, 17].

WM in non-human primates is therefore capacity-limited and susceptible to interference, much like human working memory. But what about the ability to update working memory contents? The requirement for attentional control in WM tasks is intensified in tasks that demand not only maintenance but continuous updating (i.e., addition and deletion) of WM contents [e.g. n-back tasks; 18]. This ability is central to the definition of WM as a memory system that allows for the processing and manipulation of information and not just its maintenance. A promising task for a comparative analysis of updating ability is the self-ordered-search paradigm, in which participants are asked to search an array of stimuli until they have visited every stimulus just once [19]. Efficient search requires the continuous updating of memory contents (i.e., which stimuli have already been visited or which unvisited stimuli remain). The continuous addition and deletion of memory contents within the same trial distinguishes this task from other paradigms that are used to examine WM abilities such as delayed matching tasks.
The self-ordered-search paradigm has been used to examine the development of WM young children. Diamond and colleagues [20] presented 3.5 to 7-year-old children with six visually distinct boxes as hiding places for rewards. During the 10-s retention interval between each search, the boxes either remained stationary or were scrambled. In both conditions, older children made more correct choices before their first mistake than younger ones; a finding that has been confirmed by others versions of this paradigm, providing evidence for developmental improvement in WM - indeed adult levels of performance are not evident until the age of 16 years [21, 22].

A basic version of the self-ordered-search task (with 3 boxes) has been used with macaques to examine the neural basis of working memory, though these studies did not examine limits and signatures of WM updating at the behavioral level, as we aim to do in this study. Interestingly, self-ordered-search performance in both adult humans and rhesus macaques is affected by lesions in the frontal lobes [23], especially in the lateral prefrontal cortex [19, 24], while performance on a spatial short-term memory task with little planning and updating demands was not [23]. Neonatal hippocampal lesions also lead to an impairment in self-ordered-search performance in adult rhesus macaques possibly by affecting the maturation of the dorsolateral prefrontal cortex [25]. These areas in the prefrontal cortex have been identified across different WM paradigms to be important for spatial and non-spatial working memory in humans and nonhuman primates [26, 27]. Research focusing on individual differences provides further evidence that self-ordered-search tasks provide a valid measure of WM updating, requiring executive control of attention: performance is correlated with planning tasks (such as the Tower of Hanoi task) and other working memory measures [28, 29].

In the current study, we adapted the self-ordered-search paradigm for chimpanzees (Pan troglodytes) with the aim to examine signatures and limits of WM updating abilities in one of humans’ closest living relatives. The task is particularly attractive for a comparative framework
because of its ethological validity: efficient search and avoidance of previously depleted foraging opportunities is relevant for many animal species that forage on patchily distributed food resources. Consequently, and in contrast to previous studies, the self-ordered-search paradigm requires minimal training. Akin to a version of the task for preschool children [20], chimpanzees searched for hidden rewards in an array of boxes while they needed to avoid repeated choices of the same box to retrieve all of the rewards, with 15-s retention intervals in between each choice without visual access to the search array. In this experiment, we examined the cues used by chimpanzees and their tendency to employ a search strategy. In a second experiment, we then introduced a secondary task to examine the extent to which chimpanzees’ performance was susceptible to interference.

2. Experiment 1

In Experiment 1, we adapted a self-ordered-search paradigm for use with chimpanzees, in which they had to look for food rewards in an array of baited boxes while avoiding redundant search. Importantly, these boxes were re-used over multiple trials, so that automatic processes such as novelty detection could not be utilized [13], but rather, the chimpanzees had to keep updating which box they had or had not visited in a given trial. We explored capacity limits by increasing the number of boxes in the array over trials. We explored what kind of information they used to remember their previous choices (spatial information vs. object features, see Fig. 1 and S1), and we assessed the test-retest reliability of this measure. We expected that chimpanzees would readily engage with this task, and remember their previous searches without training, but we did not know how many boxes they would be able to keep in mind. Based on previous research on short-term memory abilities in chimpanzees [30] and the
sensitivity of the self-ordered-search paradigm to normal ageing in humans [22, 31], we expected young adults to outperform older. We also expected chimpanzees to rely more on spatial cues than object features [32].

2.1. Methods

2.1.1. Subjects

Nine chimpanzees (*Pan troglodytes*; 4f, 5m; aged 11-40 years, $M_{age}$ 26.2 years) of an initial sample of 13 completed initial training and participated in Experiment 1 (see supplementary materials for more detailed methods).

2.1.2. Procedure and Design

Subjects were tested individually in purpose-built testing rooms. The experimenter (E; same person throughout the study) sat behind a sliding platform facing the subject. Small boxes served as hiding locations for food rewards. E baited each box in full view of the subject. E slid the platform towards the chimpanzees and they could then make a choice by touching a box. If the box was (still) baited, E gave the reward to the subject and showed them that the box was empty. If the subject revisited a box, E showed the subject it was empty. Boxes were returned to their previous position. Between choices E slid the platform back and occluded the platform for 15 seconds to avoid visual tracking of the containers. E’s hands were visible to the subject during the entire delay to emphasize that the status of boxes remained unchanged. During the retention interval, E looked down to a stopwatch located centrally on the ground. After the retention interval, E slid the platform toward the subject, which was the signal for subjects to make their choice. E looked up straight ahead once the platform had reached the position closest to the subject to record the subject’s choice. This procedure was repeated until the number of choices was equal to the number of boxes on the platform (which increased over trials). After the last choice in a
trial, E opened all boxes on the platform and discarded the remaining food items in a food bucket underneath the platform in full view of the subject.

There were three conditions in which the cues chimpanzees could use to remember their previous choices differed:

- **Feature+Space** (Fig. 1a): The boxes differed in shape and color, and each box remained at the same spatial position within and across trials.

- **Space-Only** (Fig 1b): All boxes looked identical (grey cylinders). Therefore, spatial cues but not feature cues were available in this condition.

- **Feature-Only** (Fig S1): The boxes differed in shape and color, and were transferred to an adjacent platform after each choice in a different order, making feature cues but not spatial cues available.

Difficulty was increased within each condition incrementally, by increasing the number of boxes from 2 to 6. Subjects received one session a day in which they received the number of trials that would yield a maximum of 24 food rewards (if they made no redundant searches), and they had two sessions to reach criterion on a given number of boxes. The passing criterion was therefore different for the different levels, which reflected the differing probabilities to find all of the rewards without redundant search on a given trial: for 2 boxes, it was 5 consecutive trials with no redundant searches within a maximum of 24 trials; for 3 boxes, 3 consecutive trials within a maximum of 16 trials; for 4 boxes, 2 consecutive trials within a maximum of 12 trials; for 5 boxes, 2 consecutive trials within a maximum of 10 trials; and for 6 boxes, 2 consecutive trials within a maximum of 8 trials. When an individual did not reach the predetermined criterion within two sessions we ended the current information type condition with this individual and continued with the next condition. All individuals started with Feature+Space. Subjects then received the Space and Feature-Only conditions in counterbalanced order across individuals.
In the retest phase, we replicated the Feature+Space condition 9-10 months after the initial assessment with the same sample of chimpanzees to assess test-retest reliability. In the re-test we used a more stringent criterion for all trial types (three consecutive trials without redundant choice) and increased the number of boxes up to 10 depending on their performance (see supplementary material for more details).

2.1.3. Scoring and analysis

We scored the number of redundant searches, the order of visited boxes, when in the sequence and where on the platform mistakes were made, the time interval subjects were absent from the platform (beyond arm’s reach) during the retention interval, and the response latencies (interval starting when the platform was pushed forward until subjects’ choice). A second coder naïve to the hypotheses and theoretical background of the study scored 20% of all sessions with regard to the box chosen to assess interobserver reliability which was excellent (K=1, N=161, p<0.001).

We used a Monte-Carlo (MC) simulation to determine whether the number of trials apes needed to reach the predetermined criterion deviated significantly from what could be expected by chance. A memory size of 0 simulated random sampling (with replacement) of the cups on the platform and counted how many unique cups were chosen. Simulating a memory size of X was realized by removing the last X choices from the pool of possible choices the model could sample from. Then we repeated this simulation until the criterion (e.g., 2 consecutive trials without redundant search) was reached. This procedure was repeated 10,000 times. We calculated p-values as the proportion of simulations that resulted in trial numbers to criterion less than or equal to the number of trials to criterion required by the individual chimpanzees. We used Fisher’s combined probability test to test whether chimpanzees in the 6-box condition performed better than assumed by the different MC simulations.
We examined variables that predicted whether subjects committed an error or not in the Feature+Space condition by coding every opportunity for committing an error separately. That is, for every choice within a trial, we coded every empty (i.e., previously visited) box separately and scored whether or not apes chose the empty boxes again. To illustrate the coding, consider a trial with three boxes on the platform. In the first choice, there is no opportunity for making a mistake (all boxes are baited). In the second choice, we would code whether the box that was selected before was chosen again or not. In the third choice within the same trial, we would code for each empty box whether it was selected again or not. Depending on the number of boxes on the platform and subjects’ choices, this re-coding procedure yielded between 1 and 5 data points for every possible choice within a trial (except for the first choice). We used a Generalized Linear Mixed Model (GLMM 01) with binomial error structure and logit link function to analyse these data.

We examined the effect of condition on whether or not chimpanzees emptied all boxes in a trial in GLMM 02. More details regarding the model fitting and assumption checks can be found in the supplementary material.

2.2. Results and discussion

In the Feature+Space condition of the initial test phase, five individuals reached the test criterion with 6 boxes, one individual with 5 boxes, two with 4 boxes, and one with 3 boxes. As the probability to reach the test criterion increased with increasing trial number, we used a Monte-Carlo (MC) simulation to determine whether the number of trials apes needed to reach the predetermined criterion deviated significantly from what could be expected by chance. Rather than just testing against a completely random model, we simulated different memory sizes. Individuals who reached the criterion of two consecutive trials without redundant choices with six boxes in the Feature+Space condition (N=5; see Fig. 1a) performed significantly better than a Monte-Carlo simulation of a memory size (MS) of 3 items (Fisher’s combined probability test:...
X² (10) = 18.31, p = 0.026; see also Fig. 2 and supplementary material for individual level results.

Overall, we found no evidence that chimpanzees used a simple search strategy (such as linear search or serial ordering strategies, see supplementary material).

To identify predictor variables associated with mistakes (revisits), we fitted GLMM 01 comprising the test predictors distance between revisits (i.e., number of visits between revisits within a trial), whether subjects had made any mistake within this trial before, the spatial position of the boxes on the platform (inner boxes vs outer boxes), the number of boxes on the platform (3-6, the two-box trials were excluded from this analysis because there were no inner boxes), the interval subjects were absent from the platform before each choice, the response latency, sex, and age and the trial number as control predictor. We also included an offset term to control for varying probabilities for mistakenly choosing a particular empty box (log(1/number of empty boxes)).

This model fitted the data significantly better than a null model (full-null model comparison: χ²(8) = 58.86, p < 0.001; see Table S1 and Fig. S2). We found that apes were more likely to revisit an empty box as the distance between revisits increased, as would be expected if memory capacity limits were increasing the likelihood of an error (χ²(1) = 11.95, p = 0.001). Moreover, the probability of mistakes decreased with increasing number of boxes on the platform (χ²(1) = 13.34, p < 0.001), which indicates improvement over trials, perhaps as the apes became more used to the overall set-up. Apes were less likely to revisit outer boxes compared to inner boxes in line with previous research [33] (χ²(1) = 17.86, p < 0.001) and younger apes made fewer mistakes than older ones (χ²(1) = 7.99, p = 0.005). When chimpanzees left the platform within a retention interval, the probability to revisit a box in the next choice was higher (χ²(1) = 6.61, p = 0.010). It is possible that chimpanzees left the platform particularly when they were increasingly uncertain about the location of the remaining food items or when they got distracted during the retention interval (e.g., due to vocalisations outside the testing room). Finally, whether apes had made a mistake within the same trial before or not did not significantly affect the probability to make a
mistake in the current choice ($\chi^2(1)=0.97, p=0.326$). The response latency ($\chi^2(1)=0.26, p=0.611$), sex ($\chi^2(1)=0.17, p=0.684$) and trial number ($\chi^2(1)=3.79, p=0.051$) did not have obvious effects on error rates either. We replicated these findings in a re-test phase except for the effect of the number of boxes (which might be an artefact of apes’ increasing experience with the task-relevant contingencies during the initial assessment) on the error probabilities (see supplementary material).

To examine the types of information apes used to remember previous choices, we fitted GLMM 02 with whether apes emptied all boxes in a trial without any redundant search or not as the dependent variable. The model comprised the test predictors condition, number of boxes, and their interaction as well as sex, age, and the control predictors trial and session (full-null model comparison: $\chi^2(7)=48.70, p<0.001$, see Table S2). The proportion of correct trials declined with increasing number of boxes; however, the extent of this decline varied across conditions (interaction: $\chi^2(2)=7.52, p=0.023$; see Fig. 3). In the Space-Only condition, adding boxes to the search array led to a steeper decline in performance compared to the Feature+Space condition (interaction: Space-Only condition x number of boxes: $z=-2.84, p=0.004$). There was no significant difference between the Feature-Only and Feature+Space condition regarding the decline in performance with increasing number of boxes (interaction: Feature-Only condition x number of boxes: $z=-1.05, p=0.295$). Post-hoc tests showed that in the initial two-box stages of each condition subjects performed better in the Feature+Space ($z=2.73, p=0.015$) and Space-Only condition ($z=2.56, p=0.025$) compared to the Feature-Only condition. There was no significant difference between the Feature+Space and Space-Only condition ($z=-1.66, p=0.203$). Besides, younger apes performed better than older ones ($\chi^2(1)=5.22, p=0.022$) whereas sex did not significantly affect their performance ($\chi^2(1)=3.00, p=0.083$). Finally, performance improved across sessions ($\chi^2(1)=4.11, p=0.043$) whereas trial number within a session did not show a significant effect on performance ($\chi^2(1)=0.62, p=0.430$).
Finally, we conducted a retest of the Feature+Space condition 9-10 months after the initial assessment to examine the reliability of subjects' task performance. We found that individual performance was highly stable across time: performance was significantly correlated across the test and re-test phase both in terms of the average proportion of correct trials with 4 and 5 boxes (the highest number of boxes that all subjects completed in the retest; $r_S=0.832$, $N=9$, $p=0.008$) and the rank order of individuals ($r_S=0.714$, $N=9$, $p=0.036$).

In summary, nine out of 13 chimpanzees avoided revisiting previously chosen boxes from their first session on, which shows that this test fulfilled our aim of being an intuitive task that did not require extensive training. Chimpanzees’ memory performance declined within trials: the probability that apes would revisit a box increased with the number of intervening searches since they last visited it, which is what we would expect if they were relying on working memory to solve the task. In general, there were large individual differences with regard to memory abilities, but younger individuals tended to perform better than older ones, consistent with previous findings on executive functions in great apes [34, 35], including a short-term memory task [30]. The best-performing individuals remembered the last four visited boxes as shown by the comparison with MC simulations of different memory sizes. One individual performed even better than a simulation of a memory size of seven items in the re-test phase. Individual differences in chimpanzees’ WM performance were stable even with multiple months in between the two assessments. Finally, subjects seemed to rely at least initially more on spatial cues than on feature cues, in line with previous research using a similar set-up [32]. But there were again large individual differences with five individuals performing better in the Space-Only condition and two individuals performing better in the Feature-Only condition.

3. Experiment 2
One important criterion of WM tasks compared to short-term memory tasks is the rehearsal of information in the face of potential interference [8, 36]. As described in the introduction, active WM has been dissociated from passive recognition in monkeys by manipulating the familiarity of the stimuli, and examining the effect of introducing interference on performance. In a delayed match-to-sample task, the performance of rhesus macaques was only sensitive to interference (a secondary task within the retention interval) when monkeys could not rely on familiarity as a mnemonic cue [12, 13]. The current task also uses highly familiar stimuli (the boxes), and so we would expect that it would be susceptible to interference. Consistent with this notion, chimpanzees made more mistakes in Experiment 1 when they left the platform during the retention interval. In Experiment 2, we tested this by adding a potentially interfering secondary task within the retention interval of the self-ordered-search task. We predicted that the more similar the secondary task was to the primary task, the more difficult it would be for the apes to keep the two tasks separate in working memory. To test this hypothesis we presented chimpanzees with two adjacent sets of four boxes each and let the apes search alternatingly in one of these two sets (see Fig. 4). We predicted that chimpanzees’ performance would decline when they were presented with two identical sets of boxes compared to two visually distinct sets of boxes due to interference from competing memory contents in the identical boxes condition. In the food distraction condition there was no additional memory demand on platform 2, which we expected to result in less task interference than either dual task condition.

### 3.1. Methods

#### 3.1.1. Subjects

In Experiment 2, our sample consisted of the same nine individuals who passed the initial training in Experiment 1. One individual was excluded from data analysis because he failed to
reach criterion with four boxes in Experiment 1 (we used a 4-box array in Experiment 2). This individual showed a floor effect in Experiment 2 (0 trials without any mistake).

3.1.2. Procedure

We used the same setup as in the Feature-Only condition of Experiment 1 (with two adjacent platforms) and a new set of boxes that differed in shape and color (see Fig. 4). There were three different within-subject conditions: Different Boxes, Identical Boxes, and Food Distraction. In all three conditions, we used the same set of four boxes that differed in shape and color on platform 1. In Different Boxes and Identical Boxes, we used an additional set of boxes on platform 2, which was either identical to the platform 1 set (Identical Boxes) or consisted of four novel boxes (Different Boxes). As in the Feature+Space condition of Experiment 1, we used the same order of boxes across subjects and throughout the experiment. In the Identical Boxes condition, we used the same order of boxes for the two sets of boxes (see supplementary material, Fig. S5).

At the beginning of each trial, E placed four boxes on platform 1 and depending on the condition also on platform 2. E then baited and closed the boxes on platform 1 and 2. In the Food Distraction condition, E also placed four food items (half banana pellets) on platform 2 after baiting the boxes on platform 1 but discarded them into a food bucket underneath the platform right away. E then placed a free-standing occluder on platform 2. Subjects could now choose a box on platform 1. E opened the indicated box, passed the food reward from inside the box to the subject, and closed the box again in the same manner as in the Experiment 1. After the first choice, E moved the occluder from platform 2 to platform 1 and occluded the boxes on platform 1. In the Identical Boxes and Different Boxes condition, subjects could now choose from platform 2. In the Food Distraction condition, E placed a reward on top of the platform; whether or not subjects received the reward depended on the performance of a matched individual in the
Identical Boxes condition (yoked procedure, see below). If the subject did not receive the
reward, E placed it on top of the platform until the subject was sitting in front of the platform 2
and looked toward the reward and discarded it then into the food bucket. E then transferred the
occluder from platform 1 to platform 2 and subjects could choose again from platform 1. This
procedure was repeated four times, i.e., until subjects could have retrieved all food items if they
did not revisit boxes.

The order of conditions was counterbalanced across subjects. Subjects received six trials per
condition over two consecutive sessions of three trials each. With regard to the yoked
procedure, we matched trios of individuals for age, sex, and Experiment 1 (Feature+Space)
performance as much as possible. Two matched individuals received the Identical Boxes and
Food Distraction conditions at the same position within the order of conditions (i.e., as their first,
second, or third condition, respectively). The platform 2 performance of one individual in the
Identical Boxes condition served as reinforcement schedule for the matched individual in the
Food Distraction condition. This procedure ensured that the Food Distraction condition was a
suitable control for attentional distraction by the presence of the food rewards and the increased
arousal or frustration induced by the temporary loss of food.

3.1.3. Scoring and analysis
We scored whether apes searched all boxes in a trial without making a mistake and coded their
platform 1 and 2 performance separately.

3.2. Results and discussion
In GLMM 03 (binomial error structure and logit link function), we analyzed whether apes
emptied all boxes without revisiting a box on platform 1. We included condition (Different Boxes,
Food Distraction, and Identical Boxes), age, and sex as test predictors and trial number (within
each condition; 1-6) and the order of conditions (1st to 3rd position) as control predictors. We
included these predictors as fixed effects and subject ID as random effect. We included all
random slope components of condition (manually dummy coded and then centered), number of
boxes, trial number, and session number except for the correlation parameters among random
intercepts and random slopes terms. The full-null model comparison was significant
\(\chi^2(4)=12.60, p=0.013\); see Table S5). We found a significant effect of condition \(\chi^2(2)=9.19,\n p=0.010\); see Fig. 5a). In line with our predictions chimpanzees’ memory performance was
significantly reduced in the identical box condition compared to the other two conditions
(Identical Boxes-Different Boxes: \(z=3.11, p=0.005\); Identical Boxes-Food Distraction: \(z=3.12, p=.005\)). Thus, the two identical sets of boxes seemed to cause interference regarding
chimpanzees’ ability to remember which boxes they had already visited (or not yet visited). To
cope with this interference in the Identical Boxes condition subjects were required to combine
two pieces of information: the features of the visited boxes and the identity of the platform.
Contrary to our predictions, performance did not differ significantly between Different Boxes and
Food Distraction conditions \(z=0.003, p=1\). The Food Distraction condition was emotionally
arousing a least for some individuals (e.g., some chimpanzees reacted negatively by spitting or
banging against the mesh panel when the experimenter removed the food from the second
platform). This emotional reaction might have impeded WM performance (similar detrimental
effects of emotional arousal on WM performance have been found in humans [37]) possibly
leading to a similar performance in the Food Distraction and Different Boxes condition even
though the Food Distraction condition did not impose an additional memory load in the
secondary task on platform 2.
Subjects’ age \(\chi^2(1)=3.35, p=0.067\) or sex \(\chi^2(1)=2.61, p=0.106\) had no significant effect on
their performance. The control predictors order of condition and trial number had no significant
effects on performance either (both \(p > 0.1\); see Table S5). Similar results were obtained on
platform 2 (see supplementary material, Fig. 5b).
4. General discussion

The current study systematically examined chimpanzees' working memory abilities regarding capacity, types of stored information, sensitivity to attentional interference, and the stability of individual differences over time. We focused on the updating of memory contents, which is a central component of many WM tasks for humans. The best performing chimpanzees remembered more than three items at the group level and one young adult seemed to remember more than seven items. Generally, young adults performed better than older ones. Within trials, their performance declined the longer the distance between revisits. A comparison between different conditions suggests that they used a combination of spatial and feature information to memorize previous choices. Alternating search between two visually identical sets of stimuli was more difficult than between visually distinct sets, indicating that their performance was sensitive to interference from competing memory contents. Finally, individual differences in performance were stable over a period of at least 9 to 10 months. Next, we highlight how the current findings extend previous work on chimpanzees’ short-term memory abilities and discuss the similarities and differences between WM updating in chimpanzees and humans.

Previous studies focusing on WM performance in chimpanzees used serial learning paradigms in which participants learn to touch an array of stimuli in order (e.g., Arabic numerals) through extensive, step-wise training. Following this training, the array was masked once the subjects touched the first numeral in the sequence [38, 39]. One juvenile chimpanzee mastered this masking condition with 9 numerals on the screen. When chimpanzees were exposed to the stimuli for a predefined time interval before each trial [210–650 ms; 39], the shortest exposure time of 210 ms did not impair the performance of the best performing juvenile (with 5 numerals...
on the screen). The authors suggested that apes’ performance in this study might be based on an eidetic memory strategy [40], the ability to recall an image after only a brief moment of exposure. Similarly, Carruthers [6] argued that chimpanzees performance in this case might reflect a form of sensory short-term memory. Important task demands that distinguish short-term memory tasks from WM tasks, such as resistance to attentional interference (induced, for example, by a secondary task) or the continuous updating of memory contents, were not part of the research design. However, anecdotal evidence suggested that chimpanzees’ performance in this task was not susceptible to attentional interference [38, 39].

In contrast, the current paradigm was administered with minimal prior experience with the task contingencies and required individuals to continuously keep the memory contents with a 15-sec retention interval in between each memory update. Importantly, we explored the effect of interference from competing memory contents, which had a negative effect on chimpanzees’ memory performance – consistent with research on human WM and attention [e.g., 41, 42].

In line with previous findings with the serial learning paradigm [30], younger individuals in the current study performed better than older ones. This difference across the age range in our study (11 to 40 years) suggests that there are developmental changes in WM updating abilities during later stages of life. In the human literature, self-ordered-search tasks provide evidence for late developmental changes and adult levels of performance are not evident until the age of 16 years [21, 22]. Young adults perform better than older adults [50–64 years; 22], and there is a further decline in performance between 55 and 79 years of age [31]. The sensitivity of this test to cognitive ageing has been related to a frontal lobe dysfunction in elderly participants.

Our findings also provide evidence regarding chimpanzees’ WM capacity limits. Cowan [43] differentiated between processing-related vs. storage-related capacity limits. The difference between these two limits is whether processing strategies are prevented that might allow
individuals to improve their performance. Such processing strategies include chunking of multiple items (grouping of objects or spatial locations) and memory rehearsal. As in the versions of the self-ordered search paradigm employed with humans [e.g., 23], we examined the extent to which individuals used a search strategy. We also purposefully hampered the use of such strategies by scrambling the stimuli in-between choices in the Feature-Only condition. While there was some evidence for a linear search strategy in the Space-Only condition, there was no consistent pattern in the other conditions, and individual differences in the variability of their search patterns did not predict accuracy either (see supplementary material). The best performing individual, however, seemed to engage in a chunking strategy in that he tended to end his search with the outer stimuli (see Table S4). It is unclear why chimpanzees did not consistently use search strategies that would reduce the memory demands. Future research might further examine the conditions under which chimpanzees might engage in such search strategies. In humans, search strategies seem to undergo late developmental changes [21, 44]. The most similar paradigm compared to our current approach has been administered by Diamond and colleagues [20]. They presented 3.5 to 7-year-old human children also with six visually distinct boxes as hiding places for rewards. In their Feature+Space condition, seven-year olds made on average 5.3 unique choices before their first mistake; in their Feature-Only condition, children reached an average score of 4.0 unique boxes. The five chimpanzees who passed the 6-box condition in the Feature+Space condition of the current study chose on average 5.62 (range: 5.3 to 6) boxes before their first mistake. The single chimpanzee who reached the 6-box stage of the Feature-Only condition selected on average 5.7 unique boxes before his first mistake suggesting remarkable similarities in the performance of chimpanzees and human children.
In conclusion, the current study provides evidence for remarkable WM updating abilities in chimpanzees. Working memory capacity has been repeatedly invoked as one of the key aspects that separates humans from our closest living relatives [3, 4]. We tested these assertions in an empirical investigation of WM performance in chimpanzees. Chimpanzees exhibited performance levels comparable with human school-age children in similar self-ordered-search tasks [e.g., 20, 36]. However, this direct comparison might be hampered by the processing strategies that humans typically adopt in these tasks and that can reduce the memory load. The search strategies (or the lack thereof) seem to be a more promising candidate for a dividing line between humans and chimpanzees than memory capacity per se. Future work might further explore how such processing strategies develop and to what extent they can also be found in nonhuman animals.

Acknowledgements

This project has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 639072). We thank Brandon Tinklenberg, Hanna Petschauer, and the animal caretakers of the Zoo Leipzig for their support with data collection and Cristina Zickert for creating the illustrations.

References


WORKING MEMORY UPDATING IN CHIMPANZEES

Heuer E., Bachevalier J. 2011 Neonatal hippocampal lesions in rhesus macaques alter the monitoring, but not maintenance, of information in working memory. *Behavioral Neuroscience* 125(6), 859.


Owen A.M., Evans A.C., Petrides M. 1996 Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cerebral Cortex* 6(1), 31-38.


Völter C.J., Call J. 2014 Younger apes and human children plan their moves in a maze task. *Cognition* 130(2), 186-203.


Figure 1 Illustration of the setup in Experiment 1; a: Feature+Space condition; b: Space-Only condition. The 6-box stage is shown here.
Figure 2 Experiment 1: The position of the first error within the search sequence in the 6-box condition. The results of the Monte-Carlo simulations of different memory sizes (ranging from 0 to 5) are shown next to the mean (±SE) performance of the chimpanzees (F+S: Feature+Space: N=5; F-Only: Feature-Only: N=1) who reached criterion with 6 boxes. A memory size of X is simulated by removing the last X choices from the possibilities the simulation can sample from.
Figure 3 Experiment 1: Individual performance (mean proportion of correct trials) as a function of the number of boxes. The horizontal lines indicate the median performance. a: Feature+Space, b: Space-Only, c: Feature-Only condition. The dashed line shows the fitted model (with age at its average and sex manually dummy coded and centered) for each of the conditions and the dotted lines its 95% confidence interval. The area of the dots depicts the number of individuals per number of boxes and proportion of correct trials (N=1 to 7).
Figure 4 Illustration of the setup in Experiment 2; a: Different Boxes condition; b: Food Distraction condition; c: Identical Boxes condition.
Figure 5 Experiment 2: Chimpanzees’ individual performance (proportion of correct trials) on (a) platform 1 and (b) platform 2 across the different conditions. The boxes indicate the quartiles and the horizontal lines inside the boxes show the median values. The blue vertical lines depict the bootstrapped 95% confidence intervals of the model, the blue horizontal lines depict the model estimates. The area of the dots depicts the number of individuals per condition and mean proportion of correct trials (N=1 to 4).