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

Authors: Christoph J. Völter[a,b,c]*, Roger Mundry[d], Josep Call[a,b], Amanda M. Seed[a]

[a] University of St Andrews, School of Psychology & Neuroscience, St Andrews, UK

[b] Max Planck Institute for Evolutionary Anthropology, Department of Developmental and

Comparative Psychology, Leipzig, Germany

[c] Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of

Vienna, University of Vienna, Vienna, Austria

[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

24

25   Working memory (WM) is thought to be one of the key ingredients of human intellect, allowing

26   individuals to hold, process, and manipulate information in support of problem solving, perhaps

27   requiring or giving rise to conscious experience [1]. It has been suggested that changes in

28   working memory might have been an important factor in human evolution [2-4], resulting in

29   qualitative differences between humans and nonhuman apes, for example, enabling complex

30   stone tool manufacture [4, 5].

31   However, the extent to which human and nonhuman WM differs requires further investigation

32   [6]. Most evidence to date for an increase in working memory capacity in human evolution

33   comes from analyses of hominin stone artefacts [2-4]. Even though this line of research

34   provides important hints, trying to specify which cognitive difference might explain differences in

35   the material end-products of a behavioral sequence is problematic [7]. A more direct evaluation

36   of the hypothesis is needed, for example by comparing humans to our closest living relatives.

37   What are the key characteristics of human WM, and how do they compare to findings about

38   nonhuman primates? Of course, humans can rely on language to rehearse information, which

39   might be the most straight-forward difference between human and non-human WM [8, 9].

40   Nevertheless, there is evidence that non-human primates engage in the active maintenance of

41   information over a delay, from research looking at how performing a secondary task disrupts

42   memory performance. WM is closely related to executive control of attention [10], and so

43   susceptibility to attentional interference is an important signature of WM engagement. WM

44   measures that involve resisting interference are highly correlated with general intelligence, while

45   short term memory tasks without this component are not [11]. Evidence for susceptibility to

46   attentional interference was found in studies with rhesus macaques (*Macaca mulatta*)

47   performing delayed matching tasks on a computer. When simpler strategies such as relative

48   familiarity could not be used (for example because all of the stimuli involved in the memory task

49   were equally familiar) performance was disrupted, but performance on versions of the task that

50   could be solved through simpler memory processes was not (for example if novel images were

51   used as stimuli, so that the target stimulus 'stood out' as the match by virtue of its relative

52   familiarity) [12, 13].

53   Human WM is also characterized by a capacity limit; though the source of this limit is

54   controversial [14]. WM capacity in adult humans corresponds to 3 – 5 pieces of information

55   when processing-related strategies such as chunking or rehearsal are prevented [15]. Studies

56   comparing humans and rhesus monkeys on the same task found a decline in performance with

57   increasing memory load in both species but a larger WM capacity in humans than in rhesus

58   monkeys [16, 17].

59   WM in non-human primates is therefore capacity-limited and susceptible to interference, much

60   like human working memory. But what about the ability to update working memory contents?

61   The requirement for attentional control in WM tasks is intensified in tasks that demand not only

62   maintenance but continuous updating (i.e., addition and deletion) of WM contents [e.g. n-back

63   tasks; 18]. This ability is central to the definition of WM as a memory system that allows for the

64   processing and manipulation of information and not just its maintenance. A promising task for a

65   comparative analysis of updating ability is the self-ordered-search paradigm, in which

66   participants are asked to search an array of stimuli until they have visited every stimulus just

67   once [19]. Efficient search requires the continuous updating of memory contents (i.e., which

68   stimuli have already been visited or which unvisited stimuli remain). The continuous addition

69   and deletion of memory contents within the same trial distinguishes this task from other

70   paradigms that are used to examine WM abilities such as delayed matching tasks.

71   The self-ordered-search paradigm has been used to examine the development of WM young

72   children. Diamond and colleagues [20] presented 3.5 to 7-year-old children with six visually

73   distinct boxes as hiding places for rewards. During the 10-s retention interval between each

74   search, the boxes either remained stationary or were scrambled. In both conditions, older

75   children made more correct choices before their first mistake than younger ones; a finding that

76   has been confirmed by others versions of this paradigm, providing evidence for developmental

77   improvement in WM - indeed adult levels of performance are not evident until the age of 16

78   years [21, 22].

79   A basic version of the self-ordered-search task (with 3 boxes) has been used with macaques to

80   examine the neural basis of working memory, though these studies did not examine limits and

81   signatures of WM updating at the behavioral level, as we aim to do in this study. Interestingly,

82   self-ordered-search performance in both adult humans and rhesus macaques is affected by

83   lesions in the frontal lobes [23], especially in the lateral prefrontal cortex [19, 24], while

84   performance on a spatial short-term memory task with little planning and updating demands was

85   not [23]. Neonatal hippocampal lesions also lead to an impairment in self-ordered-search

86   performance in adult rhesus macaques possibly by affecting the maturation of the dorsolateral

87   prefrontal cortex [25]. These areas in the prefrontal cortex have been identified across different

88   WM paradigms to be important for spatial and non-spatial working memory in humans and

89   nonhuman primates [26, 27]. Research focusing on individual differences provides further

90   evidence that self-ordered-search tasks provide a valid measure of WM updating, requiring

91   executive control of attention: performance is correlated with planning tasks (such as the Tower

92   of Hanoi task) and other working memory measures [28, 29].

93   In the current study, we adapted the self-ordered-search paradigm for chimpanzees (*Pan*

94   *troglodytes*) with the aim to examine signatures and limits of WM updating abilities in one of

95   humans' closest living relatives. The task is particularly attractive for a comparative framework

96 because of its ethological validity: efficient search and avoidance of previously depleted foraging

97 opportunities is relevant for many animal species that forage on patchily distributed food

98 resources. Consequently, and in contrast to previous studies, the self-ordered-search paradigm

99 requires minimal training. Akin to a version of the task for preschool children [20], chimpanzees

100 searched for hidden rewards in an array of boxes while they needed to avoid repeated choices

101 of the same box to retrieve all of the rewards, with 15-s retention intervals in between each

102 choice without visual access to the search array. In this experiment, we examined the cues

103 used by chimpanzees and their tendency to employ a search strategy. In a second experiment,

104 we then introduced a secondary task to examine the extent to which chimpanzees' performance

105 was susceptible to interference.

106

107

108 ## 2. Experiment 1

109 In Experiment 1, we adapted a self-ordered-search paradigm for use with chimpanzees, in

110 which they had to look for food rewards in an array of baited boxes while avoiding redundant

111 search. Importantly, these boxes were re-used over multiple trials, so that automatic processes

112 such as novelty detection could not be utilized [13], but rather, the chimpanzees had to keep

113 updating which box they had or had not visited in a given trial. We explored capacity limits by

114 increasing the number of boxes in the array over trials. We explored what kind of information

115 they used to remember their previous choices (spatial information vs. object features, see Fig. 1

116 and S1), and we assessed the test-retest reliability of this measure. We expected that

117 chimpanzees would readily engage with this task, and remember their previous searches

118 without training, but we did not know how many boxes they would be able to keep in mind.

119 Based on previous research on short-term memory abilities in chimpanzees [30] and the

120   sensitivity of the self-ordered-search paradigm to normal ageing in humans [22, 31], we

121   expected young adults to outperform older. We also expected chimpanzees to rely more on

122   spatial cues than object features [32].

123   **2.1.   Methods**

124   2.1.1. Subjects

125   Nine chimpanzees (*Pan troglodytes*; 4f, 5m; aged 11-40 years, M_{age} 26.2 years) of an initial

126   sample of 13 completed initial training and participated in Experiment 1 (see supplementary

127   materials for more detailed methods).

128   2.1.2. Procedure and Design

129   Subjects were tested individually in purpose-built testing rooms. The experimenter (E; same

130   person throughout the study) sat behind a sliding platform facing the subject. Small boxes served

131   as hiding locations for food rewards. E baited each box in full view of the subject. E slid the

132   platform towards the chimpanzees and they could then make a choice by touching a box. If the

133   box was (still) baited, E gave the reward to the subject and showed them that the box was empty.

134   If the subject revisited a box, E showed the subject it was empty. Boxes were returned to their

135   previous position. Between choices E slid the platform back and occluded the platform for 15

136   seconds to avoid visual tracking of the containers. E's hands were visible to the subject during

137   the entire delay to emphasize that the status of boxes remained unchanged. During the retention

138   interval, E looked down to a stopwatch located centrally on the ground. After the retention interval,

139   E slid the platform toward the subject, which was the signal for subjects to make their choice. E

140   looked up straight ahead once the platform had reached the position closest to the subject to

141   record the subject's choice. This procedure was repeated until the number of choices was equal

142   to the number of boxes on the platform (which increased over trials). After the last choice in a

143   trial, E opened all boxes on the platform and discarded the remaining food items in a food bucket

144   underneath the platform in full view of the subject.

145   There were three conditions in which the cues chimpanzees could use to remember their

146   previous choices differed:

147   *Feature+Space* (Fig. 1*a*): The boxes differed in shape and color, and each box remained at

148   the same spatial position within and across trials.

149   *Space-Only* (Fig 1b): All boxes looked identical (grey cylinders). Therefore, spatial cues but

150   not feature cues were available in this condition

151   *Feature-Only* (Fig S1): The boxes differed in shape and color, and were transferred to an

152   adjacent platform after each choice in a different order, making feature cues but not spatial cues

153   available.

154   Difficulty was increased within each condition incrementally, by increasing the number of

155   boxes from 2 to 6. Subjects received one session a day in which they received the number of

156   trials that would yield a maximum of 24 food rewards (if they made no redundant searches), and

157   they had two sessions to reach criterion on a given number of boxes. The passing criterion was

158   therefore different for the different levels, which reflected the differing probabilities to find all of

159   the rewards without redundant search on a given trial: for 2 boxes, it was 5 consecutive trials with

160   no redundant searches within a maximum of 24 trials; for 3 boxes, 3 consecutive trials within a

161   maximum of 16 trials; for 4 boxes, 2 consecutive trials within a maximum of 12 trials; for 5 boxes,

162   2 consecutive trials within a maximum of 10 trials; and for 6 boxes, 2 consecutive trials within a

163   maximum of 8 trials. When an individual did not reach the predetermined criterion within two

164   sessions we ended the current information type condition with this individual and continued with

165   the next condition. All individuals started with Feature+Space. Subjects then received the Space

166   and Feature-Only conditions in counterbalanced order across individuals.

167   In the retest phase, we replicated the Feature+Space condition 9-10 months after the

168   initial assessment with the same sample of chimpanzees to assess test-retest reliability. In the

169   re-test we used a more stringent criterion for all trial types (three consecutive trials without

170   redundant choice) and increased the number of boxes up to 10 depending on their performance

171   (see supplementary material for more details).

172   2.1.3. Scoring and analysis

173   We scored the number of redundant searches, the order of visited boxes, when in the

174   sequence and where on the platform mistakes were made, the time interval subjects were absent

175   from the platform (beyond arm's reach) during the retention interval, and the response latencies

176   (interval starting when the platform was pushed forward until subjects' choice). A second coder

177   naïve to the hypotheses and theoretical background of the study scored 20% of all sessions with

178   regard to the box chosen to assess interobserver reliability which was excellent (K=1, N=161,

179   p<0.001).

180   We used a Monte-Carlo (MC) simulation to determine whether the number of trials apes

181   needed to reach the predetermined criterion deviated significantly from what could be expected

182   by chance. A memory size of 0 simulated random sampling (with replacement) of the cups on

183   the platform and counted how many unique cups were chosen. Simulating a memory size of X

184   was realized by removing the last X choices from the pool of possible choices the model could

185   sample from. Then we repeated this simulation until the criterion (e.g., 2 consecutive trials

186   without redundant search) was reached. This procedure was repeated 10,000 times. We

187   calculated p-values as the proportion of simulations that resulted in trial numbers to criterion

188   less than or equal to the number of trials to criterion required by the individual chimpanzees. We

189   used Fisher's combined probability test to test whether chimpanzees in the 6-box condition

190   performed better than assumed by the different MC simulations.

191   We examined variables that predicted whether subjects committed an error or not in the

192 Feature+Space condition by coding every *opportunity* for committing an error separately. That is,

193 for every choice within a trial, we coded every empty (i.e., previously visited) box separately and

194 scored whether or not apes chose the empty boxes again. To illustrate the coding, consider a trial

195 with three boxes on the platform. In the first choice, there is no opportunity for making a mistake

196 (all boxes are baited). In the second choice, we would code whether the box that was selected

197 before was chosen again or not. In the third choice within the same trial, we would code for each

198 empty box whether it was selected again or not. Depending on the number of boxes on the

199 platform and subjects' choices, this re-coding procedure yielded between 1 and 5 data points for

200 every possible choice within a trial (except for the first choice). We used a Generalized Linear

201 Mixed Model (GLMM 01) with binomial error structure and logit link function to analyse these data.

202   We examined the effect of condition on whether or not chimpanzees emptied all boxes in a

203 trial in GLMM 02. More details regarding the model fitting and assumption checks can be found

204 in the supplementary material.

205 **2.2. Results and discussion**

206 In the Feature+Space condition of the initial test phase, five individuals reached the test criterion

207 with 6 boxes, one individual with 5 boxes, two with 4 boxes, and one with 3 boxes. As the

208 probability to reach the test criterion increased with increasing trial number, we used a Monte-

209 Carlo (MC) simulation to determine whether the number of trials apes needed to reach the

210 predetermined criterion deviated significantly from what could be expected by chance. Rather

211 than just testing against a completely random model, we simulated different memory sizes.

212 Individuals who reached the criterion of two consecutive trials without redundant choices with

213 six boxes in the Feature+Space condition (N=5; see Fig. 1a) performed significantly better than

214 a Monte-Carlo simulation of a memory size (MS) of 3 items (Fisher's combined probability test:

215   X^2 (10)=18.31, *p*=0.026; see also Fig. 2 and supplementary material for individual level results).

216   <Overall, we found no evidence that chimpanzees used a simple search strategy (such as linear

217   search or serial ordering strategies, see supplementary material).

218   To identify predictor variables associated with mistakes (revisits), we fitted GLMM 01 comprising

219   the test predictors distance between revisits (i.e., number of visits between revisits within a trial),

220   whether subjects had made any mistake within this trial before, the spatial position of the boxes

221   on the platform (inner boxes vs outer boxes), the number of boxes on the platform (3-6, the two-

222   box trials were excluded from this analysis because there were no inner boxes), the interval

223   subjects were absent from the platform before each choice, the response latency, sex, and age

224   and the trial number as control predictor. We also included an offset term to control for varying

225   probabilities for mistakenly choosing a particular empty box (log(1/number of empty boxes)).

226   This model fitted the data significantly better than a null model (full-null model comparison:

227    $\chi^2(8)$=58.86, *p*<0.001; see Table S1 and Fig. S2). We found that apes were more likely to revisit

228   an empty box as the distance between revisits increased, as would be expected if memory

229   capacity limits were increasing the likelihood of an error ($\chi^2(1)$=11.95, *p* = 0.001). Moreover, the

230   probability of mistakes decreased with increasing number of boxes on the platform ($\chi^2(1)$=13.34,

231   *p*<0.001), which indicates improvement over trials, perhaps as the apes became more used to

232   the overall set-up. Apes were less likely to revisit outer boxes compared to inner boxes in line

233   with previous research [33] ($\chi^2(1)$=17.86, *p*<0.001) and younger apes made fewer mistakes than

234   older ones ($\chi^2(1)$=7.99, *p*=0.005). When chimpanzees left the platform within a retention

235   interval, the probability to revisit a box in the next choice was higher. ($\chi^2(1)$=6.61, *p*=0.010). It is

236   possible that chimpanzees left the platform particularly when they were increasingly uncertain

237   about the location of the remaining food items or when they got distracted during the retention

238   interval (e.g., due to vocalisations outside the testing room). Finally, whether apes had made a

239   mistake within the same trial before or not did not significantly affect the probability to make a

240   mistake in the current choice ($\chi^2(1)=0.97$, $p=0.326$). The response latency ($\chi^2(1)=0.26$,

241    $p=0.611$), sex ($\chi^2(1)=0.17$, $p=0.684$) and trial number ($\chi^2(1)=3.79$, $p=0.051$) did not have

242   obvious effects on error rates either. We replicated these findings in a re-test phase except for

243   the effect of the number of boxes (which might be an artefact of apes' increasing experience

244   with the task-relevant contingencies during the initial assessment) on the error probabilities (see

245   supplementary material).

246   To examine the types of information apes used to remember previous choices, we fitted GLMM

247   02 with whether apes emptied all boxes in a trial without any redundant search or not as the

248   dependent variable. The model comprised the test predictors condition, number of boxes, and

249   their interaction as well as sex, age, and the control predictors trial and session (full-null model

250   comparison: $\chi^2(7)=48.70$, $p<0.001$, see Table S2). The proportion of correct trials declined with

251   increasing number of boxes; however, the extent of this decline varied across conditions

252   (interaction: $\chi^2(2)=7.52$, $p=0.023$; see Fig. 3). In the Space-Only condition, adding boxes to the

253   search array led to a steeper decline in performance compared to the Feature+Space condition

254   (interaction: Space-Only condition x number of boxes: z=-2.84, $p=0.004$). There was no

255   significant difference between the Feature-Only and Feature+Space condition regarding the

256   decline in performance with increasing number of boxes (interaction: Feature-Only condition x

257   number of boxes: z=-1.05, $p=0.295$). Post-hoc tests showed that in the initial two-box stages of

258   each condition subjects performed better in the Feature+Space (z=2.73, $p=0.015$) and Space-

259   Only condition (z=2.56, $p=0.025$) compared to the Feature-Only condition. There was no

260   significant difference between the Feature+Space and Space-Only condition (z=-1.66, $p=0.203$).

261   Besides, younger apes performed better than older ones ($\chi^2(1)=5.22$, $p=0.022$) whereas sex did

262   not significantly affect their performance ($\chi^2(1)=3.00$, $p=0.083$). Finally, performance improved

263   across sessions ($\chi^2(1)=4.11$, $p=0.043$) whereas trial number within a session did not show a

264   significant effect on performance ($\chi^2(1)=0.62$, $p=0.430$).

265 Finally, we conducted a retest of the Feature+Space condition 9-10 months after the initial

266 assessment to examine the reliability of subjects' task performance. We found that individual

267 performance was highly stable across time: performance was significantly correlated across the

268 test and re-test phase both in terms of the average proportion of correct trials with 4 and 5

269 boxes (the highest number of boxes that all subjects completed in the retest; r_S=0.832, N=9,

270 p=0.008) and the rank order of individuals (r_S=0.714, N=9, p=0.036).

271   In summary, nine out of 13 chimpanzees avoided revisiting previously chosen boxes from

272 their first session on, which shows that this test fulfilled our aim of being an intuitive task that did

273 not require extensive training. Chimpanzees' memory performance declined within trials: the

274 probability that apes would revisit a box increased with the number of intervening searches

275 since they last visited it, which is what we would expect if they were relying on working memory

276 to solve the task. In general, there were large individual differences with regard to memory

277 abilities, but younger individuals tended to perform better than older ones, consistent with

278 previous findings on executive functions in great apes [34, 35], including a short-term memory

279 task [30]. The best-performing individuals remembered the last four visited boxes as shown by

280 the comparison with MC simulations of different memory sizes. One individual performed even

281 better than a simulation of a memory size of seven items in the re-test phase. Individual

282 differences in chimpanzees' WM performance were stable even with multiple months in

283 between the two assessments. Finally, subjects seemed to rely at least initially more on spatial

284 cues than on feature cues, in line with previous research using a similar set-up [32]. But there

285 were again large individual differences with five individuals performing better in the Space-Only

286 condition and two individuals performing better in the Feature-Only condition

287

288   **3. Experiment 2**

289   One important criterion of WM tasks compared to short-term memory tasks is the rehearsal of

290   information in the face of potential interference [8, 36]. As described in the introduction, active

291   WM has been dissociated from passive recognition in monkeys by manipulating the familiarity of

292   the stimuli, and examining the effect of introducing interference on performance. In a delayed

293   match-to-sample task, the performance of rhesus macaques was only sensitive to interference

294   (a secondary task within the retention interval) when monkeys could not rely on familiarity as a

295   mnemonic cue [12, 13]. The current task also uses highly familiar stimuli (the boxes), and so we

296   would expect that it would be susceptible to interference. Consistent with this notion,

297   chimpanzees made more mistakes in Experiment 1 when they left the platform during the

298   retention interval. In Experiment 2, we tested this by adding a potentially interfering secondary

299   task within the retention interval of the self-ordered-search task. We predicted that the more

300   similar the secondary task was to the primary task, the more difficult it would be for the apes to

301   keep the two tasks separate in working memory. To test this hypothesis we presented

302   chimpanzees with two adjacent sets of four boxes each and let the apes search alternatingly in

303   one of these two sets (see Fig. 4). We predicted that chimpanzees' performance would decline

304   when they were presented with two identical sets of boxes compared to two visually distinct sets

305   of boxes due to interference from competing memory contents in the identical boxes condition.

306   In the food distraction condition there was no additional memory demand on platform 2, which

307   we expected to result in less task interference than either dual task condition.

308

309   **3.1.   Methods**

310   3.1.1. Subjects

311   In Experiment 2, our sample consisted of the same nine individuals who passed the initial

312   training in Experiment 1. One individual was excluded from data analysis because he failed to

313 reach criterion with four boxes in Experiment 1 (we used a 4-box array in Experiment 2). This

314 individual showed a floor effect in Experiment 2 (0 trials without any mistake).

315   3.1.2. Procedure

316 We used the same setup as in the Feature-Only condition of Experiment 1 (with two adjacent

317 platforms) and a new set of boxes that differed in shape and color (see Fig. 4).There were three

318 different within-subject conditions: Different Boxes, Identical Boxes, and Food Distraction. In all

319 three conditions, we used the same set of four boxes that differed in shape and color on

320 platform 1. In Different Boxes and Identical Boxes, we used an additional set of boxes on

321 platform 2, which was either identical to the platform 1 set (Identical Boxes) or consisted of four

322 novel boxes (Different Boxes). As in the Feature+Space condition of Experiment 1, we used the

323 same order of boxes across subjects and throughout the experiment. In the Identical Boxes

324 condition, we used the same order of boxes for the two sets of boxes (see supplementary

325 material, Fig. S5).

326 At the beginning of each trial, E placed four boxes on platform 1 and depending on the condition

327 also on platform 2. E then baited and closed the boxes on platform 1 and 2. In the Food

328 Distraction condition, E also placed four food items (half banana pellets) on platform 2 after

329 baiting the boxes on platform 1 but discarded them into a food bucket underneath the platform

330 right away. E then placed a free-standing occluder on platform 2. Subjects could now choose a

331 box on platform 1. E opened the indicated box, passed the food reward from inside the box to

332 the subject, and closed the box again in the same manner as in the Experiment 1. After the first

333 choice, E moved the occluder from platform 2 to platform 1 and occluded the boxes on platform

334 1. In the Identical Boxes and Different Boxes condition, subjects could now choose from

335 platform 2. In the Food Distraction condition, E placed a reward on top of the platform; whether

336 or not subjects received the reward depended on the performance of a matched individual in the

337   Identical Boxes condition (yoked procedure, see below). If the subject did not receive the

338   reward, E placed it on top of the platform until the subject was sitting in front of the platform 2

339   and looked toward the reward and discarded it then into the food bucket. E then transferred the

340   occluder from platform 1 to platform 2 and subjects could choose again from platform 1. This

341   procedure was repeated four times, i.e., until subjects could have retrieved all food items if they

342   did not revisit boxes.

343   The order of conditions was counterbalanced across subjects. Subjects received six trials per

344   condition over two consecutive sessions of three trials each. With regard to the yoked

345   procedure, we matched trios of individuals for age, sex, and Experiment 1 (Feature+Space)

346   performance as much as possible. Two matched individuals received the Identical Boxes and

347   Food Distraction conditions at the same position within the order of conditions (i.e., as their first,

348   second, or third condition, respectively). The platform 2 performance of one individual in the

349   Identical Boxes condition served as reinforcement schedule for the matched individual in the

350   Food Distraction condition. This procedure ensured that the Food Distraction condition was a

351   suitable control for attentional distraction by the presence of the food rewards and the increased

352   arousal or frustration induced by the temporary loss of food.

353   3.1.3. Scoring and analysis

354   We scored whether apes searched all boxes in a trial without making a mistake and coded their

355   platform 1 and 2 performance separately.

356   **3.2.   Results and discussion**

357   In GLMM 03 (binomial error structure and logit link function), we analyzed whether apes

358   emptied all boxes without revisiting a box on platform 1. We included condition (Different Boxes,

359   Food Distraction, and Identical Boxes), age, and sex as test predictors and trial number (within

360   each condition; 1-6) and the order of conditions (1st to 3rd position) as control predictors. We

361   included these predictors as fixed effects and subject ID as random effect. We included all

362   random slope components of condition (manually dummy coded and then centered), number of

363   boxes, trial number, and session number except for the correlation parameters among random

364   intercepts and random slopes terms. The full-null model comparison was significant

365   ($\chi^2(4)=12.60$, $p=0.013$; see Table S5). We found a significant effect of condition ($\chi^2(2)=9.19$,

366    $p=0.010$; see Fig. 5a). In line with our predictions chimpanzees' memory performance was

367   significantly reduced in the identical box condition compared to the other two conditions

368   (Identical Boxes-Different Boxes: z=3.11, $p=0.005$; Identical Boxes-Food Distraction: z=3.12,

369    $p=.005$). Thus, the two identical sets of boxes seemed to cause interference regarding

370   chimpanzees' ability to remember which boxes they had already visited (or not yet visited). To

371   cope with this interference in the Identical Boxes condition subjects were required to combine

372   two pieces of information: the features of the visited boxes and the identity of the platform.

373   Contrary to our predictions, performance did not differ significantly between Different Boxes and

374   Food Distraction conditions (z=0.003, $p=1$). The Food Distraction condition was emotionally

375   arousing a least for some individuals (e.g., some chimpanzees reacted negatively by spitting or

376   banging against the mesh panel when the experimenter removed the food from the second

377   platform). This emotional reaction might have impeded WM performance (similar detrimental

378   effects of emotional arousal on WM performance have been found in humans [37]) possibly

379   leading to a similar performance in the Food Distraction and Different Boxes condition even

380   though the Food Distraction condition did not impose an additional memory load in the

381   secondary task on platform 2.

382   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

383   their performance. The control predictors order of condition and trial number had no significant

384   effects on performance either (both $p > 0.1$; see Table S5). Similar results were obtained on

385   platform 2 (see supplementary material, Fig. 5b).

386

4. General discussion

388 The current study systematically examined chimpanzees' working memory abilities regarding

389 capacity, types of stored information, sensitivity to attentional interference, and the stability of

390 individual differences over time. We focused on the updating of memory contents, which is a

391 central component of many WM tasks for humans. The best performing chimpanzees

392 remembered more than three items at the group level and one young adult seemed to

393 remember more than seven items. Generally, young adults performed better than older ones.

394 Within trials, their performance declined the longer the distance between revisits. A comparison

395 between different conditions suggests that they used a combination of spatial and feature

396 information to memorize previous choices. Alternating search between two visually identical

397 sets of stimuli was more difficult than between visually distinct sets, indicating that their

398 performance was sensitive to interference from competing memory contents. Finally, individual

399 differences in performance were stable over a period of at least 9 to 10 months. Next, we

400 highlight how the current findings extend previous work on chimpanzees' short-term memory

401 abilities and discuss the similarities and differences between WM updating in chimpanzees and

402 humans.

403 Previous studies focusing on WM performance in chimpanzees used serial learning paradigms

404 in which participants learn to touch an array of stimuli in order (e.g., Arabic numerals) through

405 extensive, step-wise training. Following this training, the array was masked once the subjects

406 touched the first numeral in the sequence [38, 39]. One juvenile chimpanzee mastered this

407 masking condition with 9 numerals on the screen. When chimpanzees were exposed to the

408 stimuli for a predefined time interval before each trial [210–650 ms; 39], the shortest exposure

409 time of 210 ms did not impair the performance of the best performing juvenile (with 5 numerals

410   on the screen). The authors suggested that apes' performance in this study might be based on

411   an eidetic memory strategy [40], the ability to recall an image after only a brief moment of

412   exposure. Similarly, Carruthers [6] argued that chimpanzees performance in this case might

413   reflect a form of sensory short-term memory. Important task demands that distinguish short-term

414   memory tasks from WM tasks, such as resistance to attentional interference (induced, for

415   example, by a secondary task) or the continuous updating of memory contents, were not part of

416   the research design. However, anecdotal evidence suggested that chimpanzees' performance

417   in this task was not susceptible to attentional interference [38, 39].

418   In contrast, the current paradigm was administered with minimal prior experience with the task

419   contingencies and required individuals to continuously keep the memory contents with a 15-sec

420   retention interval in between each memory update. Importantly, we explored the effect of

421   interference from competing memory contents, which had a negative effect on chimpanzees'

422   memory performance – consistent with research on human WM and attention [e.g., 41, 42].

423   In line with previous findings with the serial learning paradigm [30], younger individuals in the

424   current study performed better than older ones. This difference across the age range in our

425   study (11 to 40 years) suggests that there are developmental changes in WM updating abilities

426   during later stages of life. In the human literature, self-ordered-search tasks provide evidence

427   for late developmental changes and adult levels of performance are not evident until the age of

428   16 years [21, 22]. Young adults perform better than older adults [50–64 years; 22], and there is

429   a further decline in performance between 55 and 79 years of age [31]. The sensitivity of this test

430   to cognitive ageing has been related to a frontal lobe dysfunction in elderly participants.

431   Our findings also provide evidence regarding chimpanzees' WM capacity limits. Cowan [43]

432   differentiated between processing-related vs. storage-related capacity limits. The difference

433   between these two limits is whether processing strategies are prevented that might allow

434   individuals to improve their performance. Such processing strategies include chunking of

435   multiple items (grouping of objects or spatial locations) and memory rehearsal. As in the

436   versions of the self-ordered search paradigm employed with humans [e.g., 23], we examined

437   the extent to which individuals used a search strategy. We also purposefully hampered the use

438   of such strategies by scrambling the stimuli in-between choices in the Feature-Only condition.

439   While there was some evidence for a linear search strategy in the Space-Only condition, there

440   was no consistent pattern in the other conditions, and individual differences in the variability of

441   their search patterns did not predict accuracy either (see supplementary material). The best

442   performing individual, however, seemed to engage in a chunking strategy in that he tended to

443   end his search with the outer stimuli (see Table S4). It is unclear why chimpanzees did not

444   consistently use search strategies that would reduce the memory demands. Future research

445   might further examine the conditions under which chimpanzees might engage in such search

446   strategies. In humans, search strategies seem to undergo late developmental changes [21, 44].

447   The most similar paradigm compared to our current approach has been administered by

448   Diamond and colleagues [20]. They presented 3.5 to 7-year-old human children also with six

449   visually distinct boxes as hiding places for rewards. In their Feature+Space condition, seven-

450   year olds made on average 5.3 unique choices before their first mistake; in their Feature-Only

451   condition, children reached an average score of 4.0 unique boxes. The five chimpanzees who

452   passed the 6-box condition in the Feature+Space condition of the current study chose on

453   average 5.62 (range: 5.3 to 6) boxes before their first mistake. The single chimpanzee who

454   reached the 6-box stage of the Feature-Only condition selected on average 5.7 unique boxes

455   before his first mistake suggesting remarkable similarities in the performance of chimpanzees

456   and human children.

457 In conclusion, the current study provides evidence for remarkable WM updating abilities in

458 chimpanzees. Working memory capacity has been repeatedly invoked as one of the key

459 aspects that separates humans from our closest living relatives [3, 4]. We tested these

460 assertions in an empirical investigation of WM performance in chimpanzees. Chimpanzees

461 exhibited performance levels comparable with human school-age children in similar self-

462 ordered-search tasks [e.g., 20, 36]. However, this direct comparison might be hampered by the

463 processing strategies that humans typically adopt in these tasks and that can reduce the

464 memory load. The search strategies (or the lack thereof) seem to be a more promising

465 candidate for a dividing line between humans and chimpanzees than memory capacity per se.

466 Future work might further explore how such processing strategies develop and to what extent

467 they can also be found in nonhuman animals.

468

474

475 **References**

476 1. Baars B.J., Franklin S. 2003 How conscious experience and working memory interact.

477 *Trends in cognitive sciences* **7**(4), 166-172.

478 2. Engle R.W. 2010 Role of working-memory capacity in cognitive control. *Curr Anthropol*

479 **51**(S1), S17-S26.

480   3.   Coolidge F.L., Wynn T. 2005 Working memory, its executive functions, and the

481   emergence of modern thinking. *Camb Archaeol J* **15**(1), 5-26.

482   4.   Read D.W. 2008 Working memory: a cognitive limit to non-human primate recursive

483   thinking prior to hominid evolution. *Evol Psychol* **6**(4), 147470490800600413.

484   5.   Stout D. 2010 The evolution of cognitive control. *Top Cogn Sci* **2**(4), 614-630.

485   6.   Carruthers P. 2013 Evolution of working memory. *Proc Natl Acad Sci USA* **110**(2),

486   10371-10378.

487   7.   Tennie C., Premo L.S., Braun D.R., McPherron S.P. 2017 Early Stone Tools and

488   Cultural Transmission: Resetting the Null Hypothesis. *Curr Anthropol* (5), 652-672.

489   8.   Baddeley A. 2003 Working memory: looking back and looking forward. *Nat Rev Neurosci*

490   **4**(10), 829.

491   9.   Wright A.A., Cook R.G., Rivera J.J., Shyan M.R., Neiworth J.J., Jitsumori M. 1990

492   Naming, rehearsal, and interstimulus interval effects in memory processing. *J Exp Psychol*

493   *Learn Mem Cogn* **16**(6), 1043.

494   10.   Engle R.W. 2002 Working memory capacity as executive attention. *Curr Dir Psychol Sci*

495   **11**(1), 19-23.

496   11.   Engle R.W., Tuholski S.W., Laughlin J.E., Conway A.R. 1999 Working memory, short-

497   term memory, and general fluid intelligence: a latent-variable approach. *J Exp Psychol Gen*

498   **128**(3), 309.

499   12.   Brady R.J., Hampton R.R. 2018 Nonverbal working memory for novel images in rhesus

500   monkeys. *Curr Biol* **28**(24), 3903-3910. e3903.

501   13.   Basile B.M., Hampton R.R. 2013 Dissociation of active working memory and passive

502   recognition in rhesus monkeys. *Cognition* **126**(3), 391-396.

503   14.   Oberauer K., Kliegl R. 2006 A formal model of capacity limits in working memory. *J Mem*

504   *Lang* **55**(4), 601-626.

505   15.   Cowan N. 2001 Metatheory of storage capacity limits. *Behav Brain Sci* **24**(1), 154-176.

506   16.   Elmore L.C., Ma W.J., Magnotti J.F., Leising K.J., Passaro A.D., Katz J.S., Wright A.A.

507   2011 Visual short-term memory compared in rhesus monkeys and humans. *Curr Biol* **21**(11),

508   975-979.

509   17.   Elmore L.C., Wright A.A. 2015 Monkey visual short-term memory directly compared to

510   humans. *J Exp Psychol Animal Learn Cogn* **41**(1), 32.

511   18.   Schmiedek F., Hildebrandt A., Lövdén M., Wilhelm O., Lindenberger U. 2009 Complex

512   span versus updating tasks of working memory: The gap is not that deep. *J Exp Psychol Learn*

513   *Mem Cogn* **35**(4), 1089.

514   19.   Petrides M., Milner B. 1982 Deficits on subject-ordered tasks after frontal-and temporal-

515   lobe lesions in man. *Neuropsychologia* **20**(3), 249-262.

516   20.   Diamond A., Prevor M.B., Callender G., Druin D.P. 1997 Prefrontal cortex cognitive

517   deficits in children treated early and continuously for PKU. *Monogr Soc Res Child Dev*, i-206.

518   21.   Luciana M., Conklin H.M., Hooper C.J., Yarger R.S. 2005 The development of nonverbal

519   working memory and executive control processes in adolescents. *Child Dev* **76**(3), 697-712.

520   22.   De Luca C.R., Wood S.J., Anderson V., Buchanan J.-A., Proffitt T.M., Mahony K.,

521   Pantelis C. 2003 Normative data from the CANTAB. I: development of executive function over

522   the lifespan. *J Clin Exp Neuropsychol* **25**(2), 242-254.

523   23.   Owen A.M., Downes J.J., Sahakian B.J., Polkey C.E., Robbins T.W. 1990 Planning and

524   spatial working memory following frontal lobe lesions in man. *Neuropsychologia* **28**(10), 1021-

525   1034.

526   24.   Petrides M. 1995 Impairments on nonspatial self-ordered and externally ordered working

527   memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *J*

528   *Neurosci* **15**(1), 359-375.

529   25.   Heuer E., Bachevalier J. 2011 Neonatal hippocampal lesions in rhesus macaques alter

530   the monitoring, but not maintenance, of information in working memory. *Behavioral*

531   *neuroscience* **125**(6), 859.

532   26.   Goldman-Rakic P.S. 1995 Cellular basis of working memory. *Neuron* **14**(3), 477-485.

533   27.   Owen A.M., Evans A.C., Petrides M. 1996 Evidence for a two-stage model of spatial

534   working memory processing within the lateral frontal cortex: a positron emission tomography

535   study. *Cerebral Cortex* **6**(1), 31-38.

536   28.   Bryan J., Luszcz M.A. 2001 Adult age differences in self-ordered pointing task

537   performance: Contributions from working memory, executive function and speed of information

538   processing. *J Clin Exp Neuropsychol* **23**(5), 608-619.

539   29.   Ross T.P., Hanouskova E., Giarla K., Calhoun E., Tucker M. 2007 The reliability and

540   validity of the self-ordered pointing task. *Arch Clin Neuropsychol* **22**(4), 449-458.

541   30.   Inoue S., Matsuzawa T. 2009 Acquisition and memory of sequence order in young and

542   adult chimpanzees (Pan troglodytes). *Anim Cogn* **12**(1), 59-69.

543   31.   Robbins T.W., James M., Owen A.M., Sahakian B.J., Lawrence A.D., McInnes L.,

544   Rabbitt P.M. 1998 A study of performance on tests from the CANTAB battery sensitive to frontal

545   lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive

546   functioning and cognitive aging. *J Int Neuropsychol Soc* **4**(5), 474-490.

547   32.   Haun D.B., Call J., Janzen G., Levinson S.C. 2006 Evolutionary psychology of spatial

548   representations in the hominidae. *Curr Biol* **16**(17), 1736-1740.

549   33.   Hribar A., Call J. 2011 Great apes use landmark cues over spatial relations to find

550   hidden food. *Anim Cogn* **14**(5), 623.

551   34.   Völter C.J., Call J. 2014 Younger apes and human children plan their moves in a maze

552   task. *Cognition* **130**(2), 186-203.

553   35.   Manrique H.M., Call J. 2015 Age-dependent cognitive inflexibility in great apes. *Anim*

554   *Behav* **102**, 1-6.

555   36.   Cragg L., Nation K. 2007 Self-ordered pointing as a test of working memory in typically

556   developing children. *Memory* **15**(5), 526-535.

557   37.   Kensinger E.A., Corkin S. 2003 Effect of negative emotional content on working memory

558   and long-term memory. *Emotion* **3**(4), 378.

559   38.   Kawai N., Matsuzawa T. 2000 Numerical memory span in a chimpanzee. *Nature* **403**,

560   39-40.

561   39.   Inoue S., Matsuzawa T. 2007 Working memory of numerals in chimpanzees. *Curr Biol*

562   **17**(23), R1004-R1005.

563   40.   Jaensch E. 1931 *Eidetic imagery and typological methods of investigation*. Abingdon,

564   UK, Routledge.

565   41.   Oberauer K. 2002 Access to information in working memory: exploring the focus of

566   attention. *J Exp Psychol Learn Mem Cogn* **28**(3), 411.

567   42.   Olivers C.N., Meijer F., Theeuwes J. 2006 Feature-based memory-driven attentional

568   capture: visual working memory content affects visual attention. *J Exp Psychol Hum Percept*

569   *Perform* **32**(5), 1243.

570   43.   Cowan N. 2010 The magical mystery four: How is working memory capacity limited, and

571   why? *Curr Dir Psychol Sci* **19**(1), 51-57.

572   44.   Luciana M., Nelson C.A. 2002 Assessment of neuropsychological function through use

573   of the Cambridge Neuropsychological Testing Automated Battery: performance in 4-to 12-year-

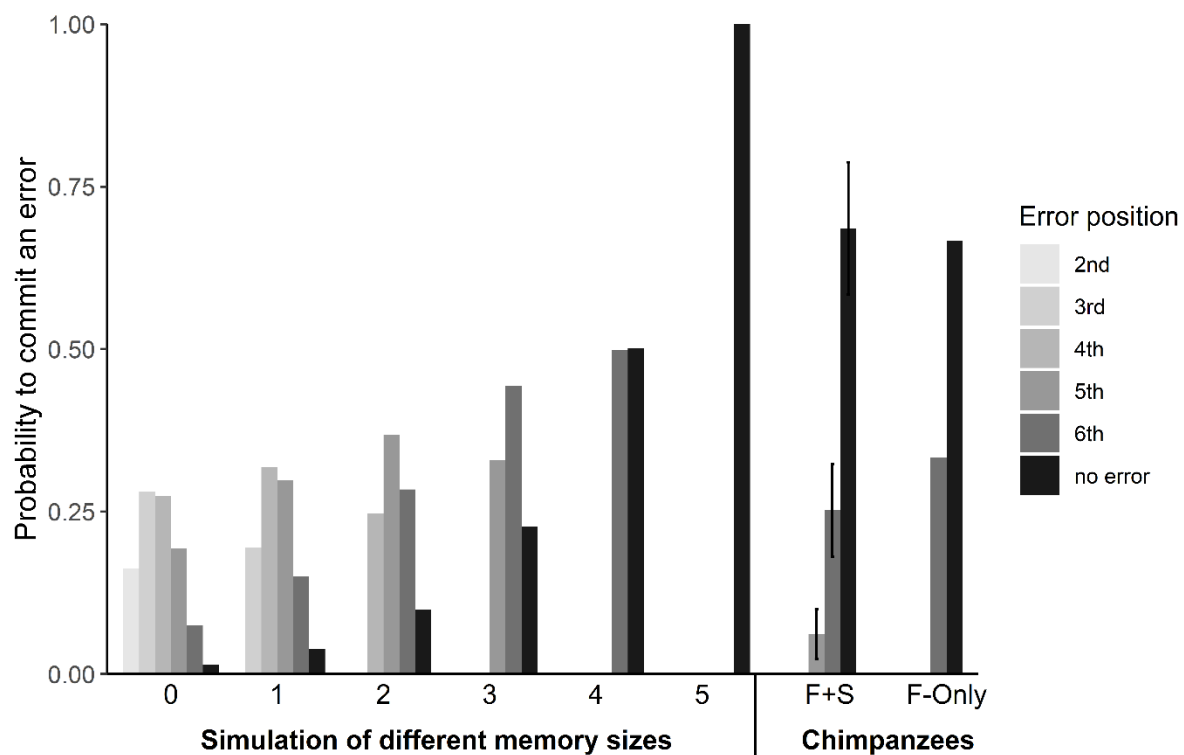574   old children. *Dev Neuropsychol* **22**(3), 595-624.

575

576

577
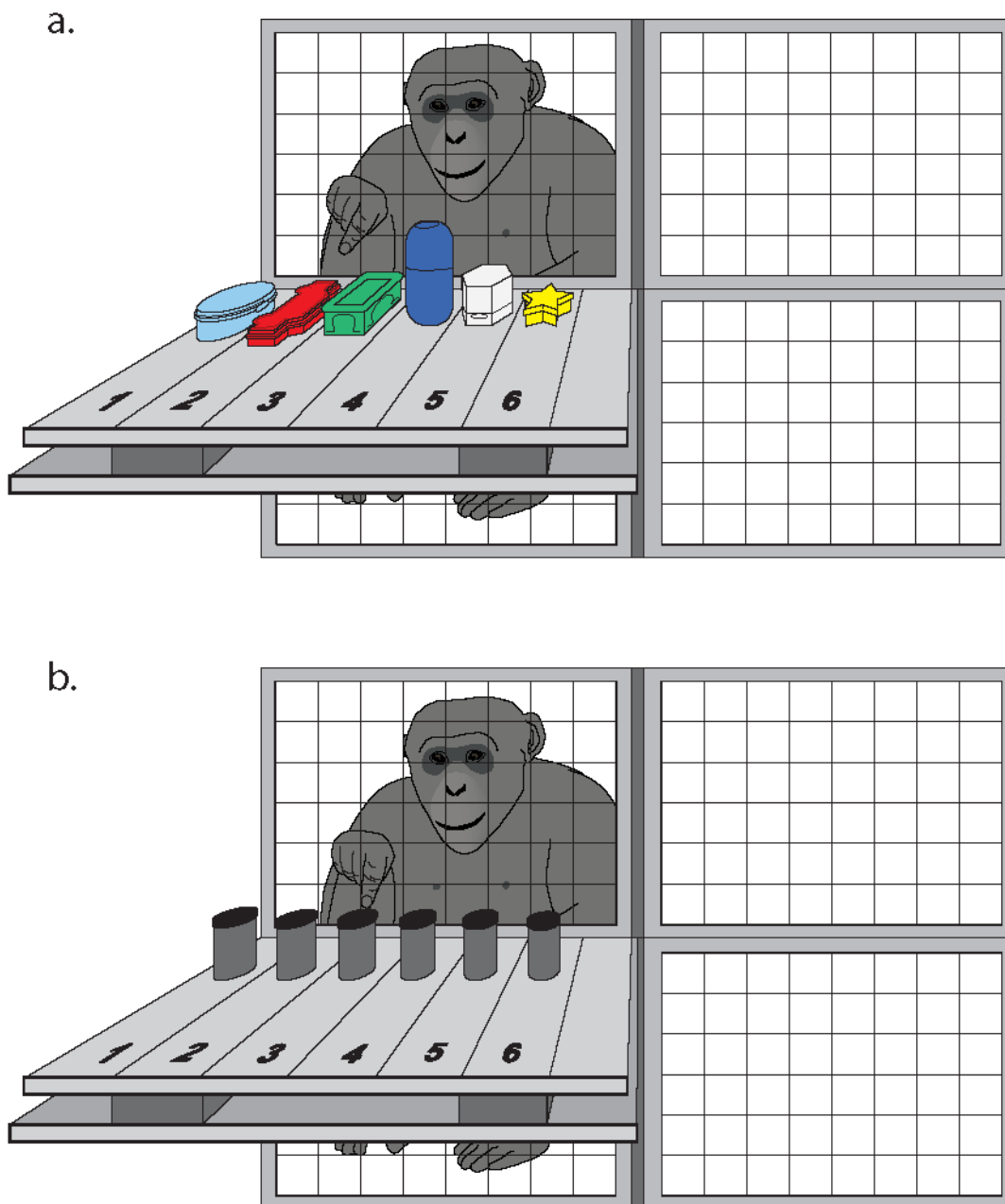
Figure 1 Illustration of the setup in Experiment 1; *a:* Feature+Space condition; *b*: Space-Only
condition. The 6-box stage is shown here.

580

581

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

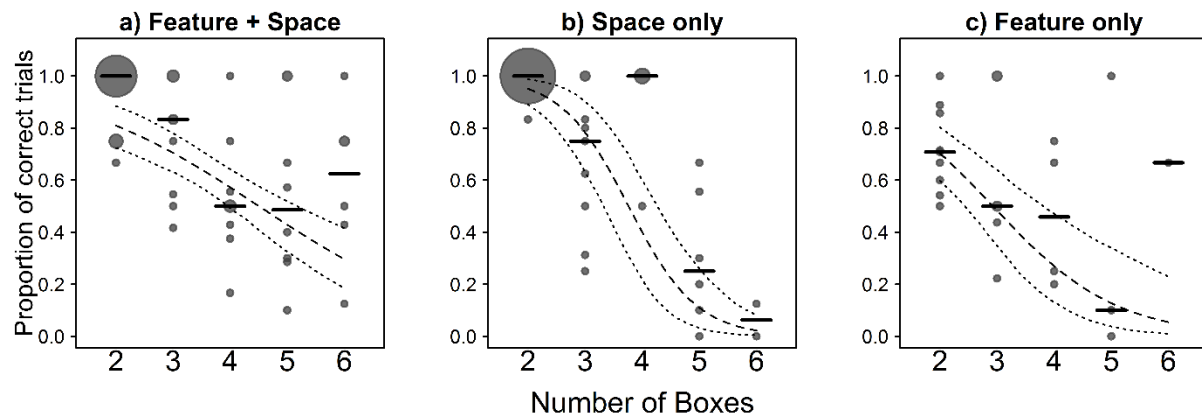simulated by removing the last X choices from the possibilities the simulation can sample from.

587

588

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).

595
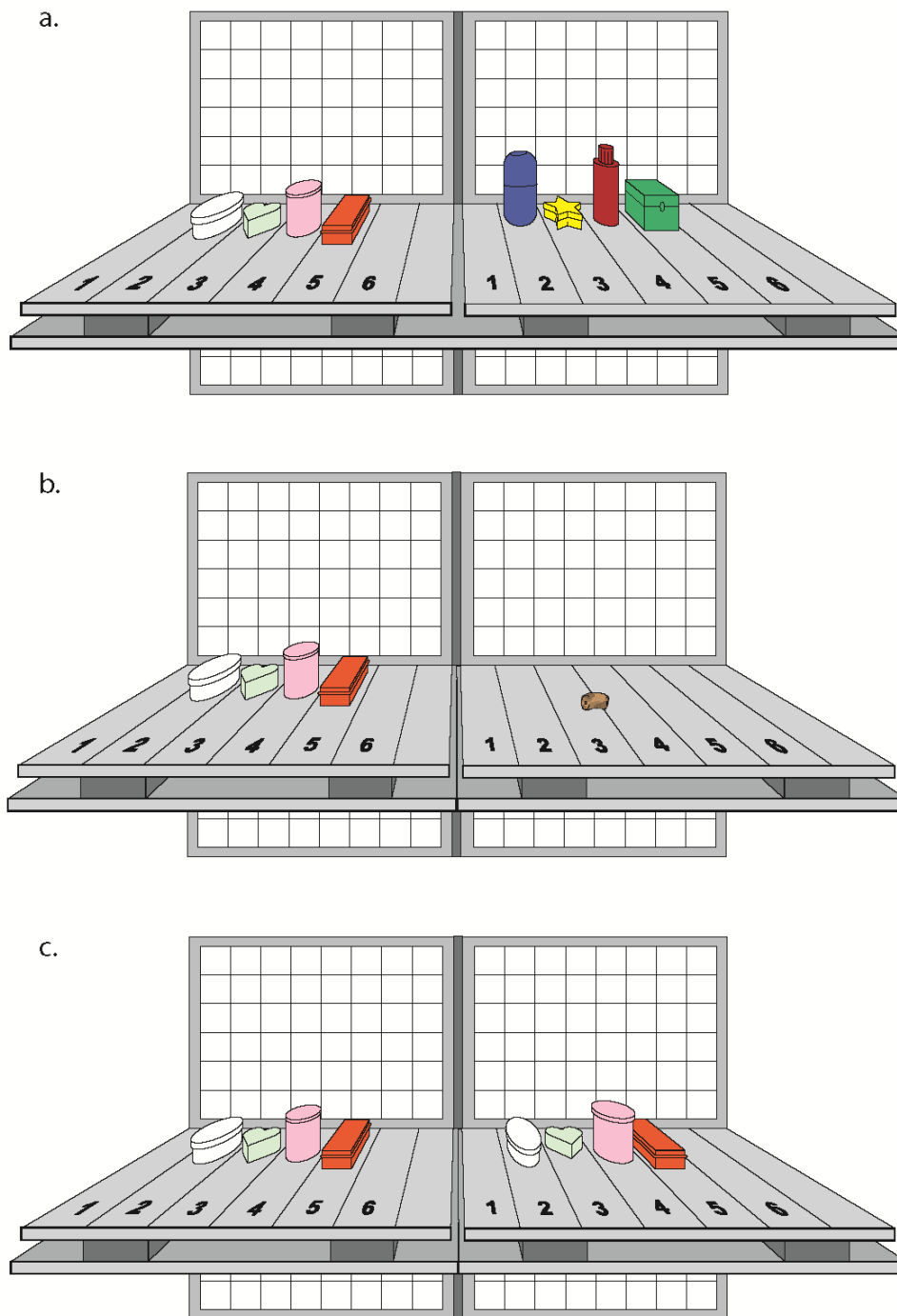
596

597   Figure 4 Illustration of the setup in Experiment 2; a: Different Boxes condition; b: Food

598   Distraction condition; c: Identical Boxes condition.
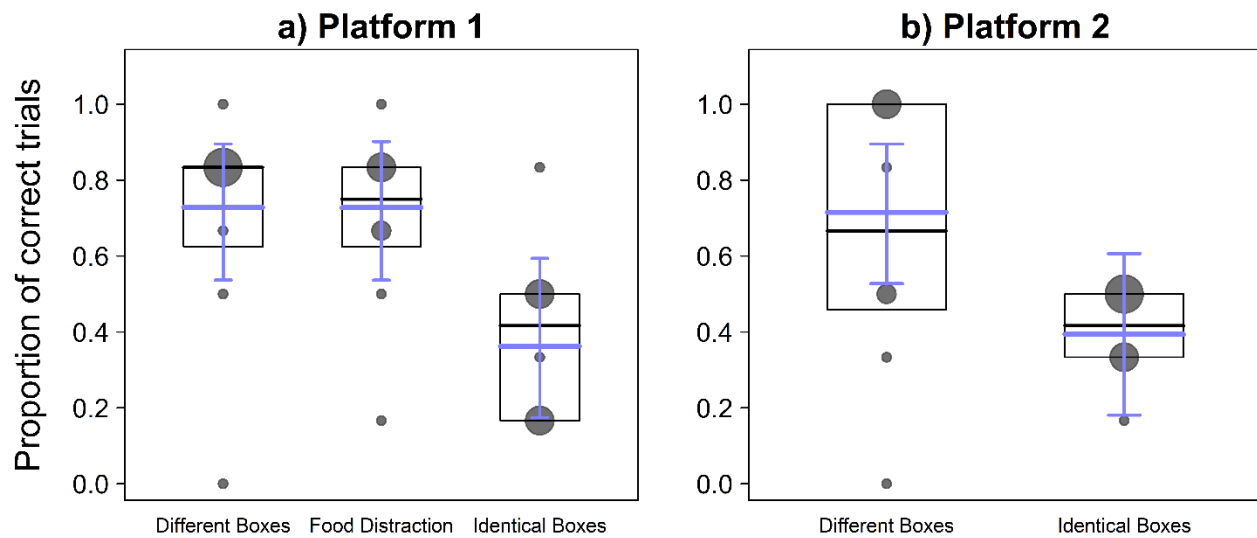
599

600   Figure 5 Experiment 2: Chimpanzees' individual performance (proportion of correct trials) on (a)

601   platform 1 and (b) platform 2 across the different conditions. The boxes indicate the quartiles

602   and the horizontal lines inside the boxes show the median values. The blue vertical lines depict

603   the bootstrapped 95% confidence intervals of the model, the blue horizontal lines depict the

604   model estimates. The area of the dots depicts the number of individuals per condition and mean

605   proportion of correct trials (N=1 to 4).