Age effect in expert cognitive flexibility in Guinea baboons *(Papio papio)*

Julie Gullstrand [*,1], Nicolas Claidière, Joel Fagot

Laboratory of Cognitive Psychology, CNRS, Aix-Marseille University, and Primatology Station of the CNRS-Celphedia, France

ARTICLE INFO

ABSTRACT

Cognitive flexibility in non-human primates is traditionally measured with the conceptual set shifting task (CSST). In our laboratory, Guinea baboons (N = 24) were continuously tested with a CSST task during approximately 10 years. Our task involved the presentation of three stimuli on a touch screen all made from 3 possible colours and 3 shapes. The subjects had to touch the stimulus containing the stimulus dimension (e.g., green) that was constantly rewarded until the stimulus dimension changed. Analysis of perseveration responses, scores and response times collected during the last two years of testing (approximately 1.6 million trials) indicate (1) that the baboons have developed an "expert" form of cognitive flexibility and (2) that their performance was age-dependent, it was at a developing stage in juveniles, optimal in adults, declining in middle-aged, and strongly impaired in the oldest age group. A direct comparison with the data collected by Bonté , Flemming & Fagot (2011) on some of the same baboons and same task as in the current study indicates that (3) the performance of all age groups has improved after 10 years of training, even for the now old individuals. All these data validate the use of non-human primates as models of human cognitive flexibility and suggest that cognitive flexibility in humans has a long evolutionary history.

1. Introduction

Cognitive flexibility is an executive function defined as the ability to shift attention from one category of stimuli to another [1] which is of major importance to adapt to changes in the physical or social environment. This cognitive function contributes for instance to problem solving requiring adapting the response to the new demands of the situation [2]. Cognitive flexibility is traditionally assessed in humans using the Wisconsin Card Sorting Test (WCST, [3]). In the WCST, subjects must sort cards according to the relevant dimension (number, shape or colour) inferred from the experimenter's feedback, with regular changes of this relevant dimension. Cognitive flexibility is measured from the number of perseverative errors corresponding to the number of repetitions of the initial response after the change of dimension.

The WCST has been amply used in clinical studies, showing for example that lesions of the frontal lobes in human epileptic patients impair cognitive flexibility [4], which supports Luria's [5] claim about the management of executive functions by the frontal lobes. WCST has also been used to assess how flexibility evolves during development and aging in normal populations (e.g., [6]). Results show that 3 years-old children can sort cards on one dimension but fail after the change of the relevant dimension [7]. At 4 years old, children start to succeed and performance improves until the end of adolescence [8]. Performance is optimal in young adults but it later starts declining between 40 and 50 years [9]. These declines in flexibility parallel structural changes in the prefrontal cortex, the volume of which decreases after 50 years with an accentuation of the phenomenon after 65 years [10].

From a comparative perspective, cognitive flexibility has been explored in non-human-primates using an analogue of the WSCT referred to as intradimensional / extradimensional set shifting (ID/ED) task. This task involves the use of bi-dimensional compound stimulus, such as a configuration of black lines shown on a blue-filled shape serving as a background [11,12]. After the participants have learned that the selection of the stimulus containing one cue (i.e., a specific configuration of line or a specific background shape) triggers a food reward, the test involves the learning of a new association between a novel cue from the same dimension (intra-dimensional shift) or the alternative dimension (interdimensional shift). The use of this ID/ED

task with Marmosets (*Callitrix jacchus*) revealed that intra-dimensional set shifting was easier than the extra-dimensional shifting. Performance also improved with successive testing and was altered by lesions of the prefrontal cortex. In rhesus macaques (*Macaca mulatta*), the ID/ED performance depended on the age of the subjects: The young macaques (mean 2.3 years) committed more perseverative errors than adults (mean 10.3 years;[12]).

A slightly different adaptation of the WCST, referred to as the Conceptual Set-Shifting Task, (CSST) was also developed for rhesus macaques by Moore et al. [13,14]. In their study, the participants were presented with three visual stimuli selected from a set made from the combination of three possible colours (red, green, and blue) and three shapes (triangle, star and circle, nine stimuli altogether). Monkeys had to initially form a concept by establishing a pattern of responding to the red stimulus class, and then shift to a different stimulus class when the reward contingency changed. Three conceptual set shiftings were proposed in these studies, in which the new rule involved the selection of the triangle shape, the blue colour, and the star. The old macaques (24–30 years) were slower for learning the new rules after set shifting and revealed more perseverative errors than younger individuals (5–10 years), suggesting an age-related decline in the functioning of the prefrontal cortex [14]. The CSST of Moore et al. (2003) was further adapted in our laboratory to measure cognitive flexibility in a troop of 24 Guinea baboons (*Papio papio*, [15], Experiment 1). A first study [15] confirmed that cognitive flexibility diminishes with aging, as the rate of perseverative errors correlated positively with the age of the baboons. Interestingly, our baboons were younger on average than in Moore et al. (2003), with an age only varying from 2.2 to 14.6 years (baboons and macaques have similar life expectancies). Our study [15] therefore indicates a relatively early decline in cognitive flexibility that occurs in mid-adulthood, before animals reach older age.

In summary, the IE/ED and CSST adaptations of WCST in different species of monkeys all confirmed the existence of a cognitive flexibility decline with aging which might be related to reduced frontal cortex efficiency. However, inspection of this literature on aging suggests two remarks. First, the number of studies in that domain remains extremely limited, probably due to the necessity to have access to a large group of monkeys to assess age effects on cognitive flexibility. Second, most of the studies on nonhuman primates interested in age differences in conceptual shifting involved a very limited number of conceptual shifts, typically 3 or 4 shifts as in Moore et al. [13,14] or Bonté et al. [15], and 8 in the ED/ID task of Weed et al. [12] and Dias et al. [11]. Such a limited number of shifts per participant suggests that the measures of cognitive flexibility might not be optimal in these studies, because the animals were still in the process learning the task when tested .[2]

The current study alleviates these concerns. This study takes advantage of a unique feature of our laboratory in which a troop of Guinea baboons have had a free and unlimited access, since 2009, to a battery of operant conditioning test systems referred to as the Automated Learning Devices for Monkeys (ALDM, see [16]). In practice, the baboons from that group have been exposed to a large number of cognitive experiments addressing a variety of scientific questions (e.g., on the properties of their perceptual system [17], memory [18], reasoning [19], or social cognition [20]). As a baseline protocol in our laboratory, the same CSST task as in Bonté et al. [15] has been repeatedly used since 2012 as a filler task in between our different experimental programs. The vicissitude of our research (e.g., change in software and data loss) prevents statistical analyses of perseverative errors from the CSST data collected from 2012 to 2018, but one can

nevertheless infer from the data set that our baboons received a total number of 143 772 CSST trials on average (range 3469–327198, see supplemental Table 1) prior to the collection of the current data. This leaves the opportunity to study cognitive flexibility from the CSST data collected from 2018 to 2020, after years of exposure to this task. Although limited to two years of testing, our data set remains uniquely large in comparison to previously published studies in this domain, with exactly 1 664 118 CSST trials (an average of 69 338 trials per subject, SE = 11 248 trials).

Our goal in this context is threefold. First, we want to document an "expert" form of cognitive flexibility in non-human primates, which is measured behaviorally after the subjects have been exposed over 10 years to thousands of set shifts. Second, we want to document to what extent cognitive flexibility in our task depends on the age of the subjects. Our last goal is to compare our measures of cognitive flexibility in that situation with previously published behavioural data obtained with much less training by Bonté et al. [15], to document the reliability of previous conclusions.

2. Material and methods

2.1. Participants and living conditions

The participants are 24 Guinea baboons (*Papio papio*) from the CNRS primates Centre, Rousset-sur-Arc, France. This colony is divided in two different social groups of 6 (2 males and 4 females) and 18 baboons (6 males and 12 females). The small group is housed in 7.5 × 6.4 m outdoor enclosures connected to indoor housings. The biggest group is housed in a 25 × 30 m outdoor enclosure connected by tunnels to indoor housing used at night. The groups has ad libitum access to water, and feeding is provided daily at 4 pm. Table 1 provides information the sex and mean age of the participant during the study period. The females were neither pregnant nor lactating during the study period. Note that all age classes were represented in our baboons, ranging from 31 to 291 months (i.e., 2–24 years).

Table 1

Information on the age, sex, group of the participants, with the number of RS (rule sessions) exploitable after filtering data.

Name	Sex	MeanAge	Age Class	RS
LIPS	F	49	Young	987
LOME	M	54	Young	1324
MAKO	M	41	Young	–
MALI	F	48	Young	869
MUSE	F	42	Young	–
NEKKE	F	31	Young	85
EWINE	F	124	Adult	1576
FANA	F	117	Adult	1013
FELIPE	M	114	Adult	218
FEYA	F	111	Adult	1108
FLUTE	F	103	Adult	447
HARLEM	M	88	Adult	647
HERMINE	F	89	Adult	337
ANGELE	F	175	Middle-age	313
ARIELLE	F	170	Middle-age	1399
ARTICHO	M	168	Middle-age	253
BOBO	M	165	Middle-age	63
CAUET	M	145	Middle-age	875
DORA	F	133	Middle-age	464
DREAM	F	140	Middle-age	738
VIOLETTE	F	180	Middle-age	1672
ATMOSPHERE	F	263	Old	438
BRIGITTE	F	270	Old	39
KALI	F	292	Old	48
PETOULETTE	F	249	Old	37
PIPO	M	249	Old	55

---

[2] More extensive training was proposed in the context of electrophysiological studies in monkeys, but these studies remain limited for our purpose due to a too small number of subjects (N = 2; see [37–41]) or an absence of report of the age of the subjects preventing any comparison among age groups (e.g., [42, 43]).

2.2. Ethical statements

This research adhered to the applicable French and E.U rules for the ethical treatment of research animals. It received ethical approval from the national French ethics committee " Comité d'Ethique CE-14 " for experimental animal research, as well as the French Ministry of Education (approval APAFIS#2717–2015111708173794 10 v3).

2.3. Apparatus

The enclosures available to each group are connected to experimental trailers providing free access to a total of ten (large group) and four (small group) automatized ALDM test systems. Each ALDM testing booth consists of a 70×70 cm automatic operant test chamber equipped with RFID tag reader. When a monkey enters an ALDM, it is identified by RFID microchips implanted in its forearms. This identification triggers the task which is displayed on a 19 in.' touchscreen. Correct responses are rewarded by a few grains of wheat delivered by a food dispenser, while incorrect responses results in a display of a green screen as time out for 3 s. At any time, monkeys can leave ALDM systems or decide to resume testing (see [21] and [16] for more details). When monkeys return voluntarily to the test system after a break, testing is resumed as the point it was left off.

2.4. CSST task

Our task was the same CSST task as in Bonté et al. [15]. It used a set of 9 stimuli resulting from the combination of three possible colours (green, pink and yellow) and three shapes ("Circle", "Triangle" and "Splash"). At the beginning of each session, the test program automatically selected either one stimulus shape (e.g., triangle) or one stimulus colour (e.g, green) that was used as target during the entire session. Each trial consisted in the display of 3 stimuli from the set randomly located in a matrix of 9 possible positions on the screen. The three stimuli were selected in such a way that the three shapes and three colours were all represented by only one stimulus. The task for the subject was to touch the stimulus containing the target. Each time the monkey reached the target, it was rewarded with a few wheat grains. If the subject touched one of the distractors, this behaviour triggered a 3 s timeout and the appearance of a green screen. All information concerning the subject (i. e., identity, age and sex), the task (i.e., type of trial, target) and response behaviour (score and response time) was automatically recorded at the end of each trial.

Testing was organized in sessions and blocks. Sessions corresponded to series of consecutive trials involving the same rule, for instance select the yellow colour or the triangle. For clarity, sessions will hereafter be referred to as "Rule sessions" (RS). RS were organized in blocks of 100 trials, at the end of each block the percentage of correct responses was automatically calculated. The criterion for rule shift was 80% correct response in the block of 100 trials. If that percentage was lower than 80%, the subject resumed testing with a new block of 100 trials conserving the same rule as previously (i.e. stayed in the same RS). Once the learning criterion has been reached in a given RS, the subject started a new RS with a new target which was selected at random from the set of five potential shape or colour targets different from the previous one. After a rule shift, the novel RS followed the same procedure as above, with the only difference that we avoided any intersection between the old and the new target during the first trial of the RS. Thus, if the previous target was "Yellow" for instance and the new one was "Triangle", the target could not be a yellow triangle during the first trial. Note that on this first trial the monkeys could not have known that the rule had changed. This constraint was removed from the second trials onward.

We acknowledge that our procedure for rule changes slightly differs from most studies in the field who used a sliding window to suddenly change the rule after the subjects had reached a criterion (typically after the subject has produced 10 consecutive correct responses [13]). Two reasons justify our choice to use trial blocks to compute the training criterion. First, a too rapid change in reinforcement contingencies after learning would hamper the free participation of the monkeys to the experiment, due to a too low reinforcement rate at the session level. Second, use of a trial block procedure allowed us to document the stabilisation of the rule, and not only focus on the initial learning process. Similar block procedures had already been used in Bonté et al. [15] and in Mansouri and Tanaka [22] who used a criteria of 90% correct responses in blocks of 40 trials, then 80% correct responses in blocks of 20 trials. To guaranty that our procedure allowed the same level of learning as with more traditional procedures using sliding windows, we computed the mean score obtained by the baboons during the last ten last of the block prior to the shift. At the group level, we found that the baboons performed 92.6% correct on average (SD = 0.05) during the last ten trials. This performance is therefore in the same range as in past studies of the field.

2.5. Data filtering

Data were collected on this CSST task from May 15th, 2018, to December 11th, 2020. During this period, the continuous use of CSST as a filler task allowed the recording of a uniquely large number of trials per animal, but also imposed (and allowed) stringent data filtering. First, the ALDM testing procedure implies that the monkeys can stop and resume testing. Long breaks can therefore sometime occur between two consecutive trials in a RS. Second, CCST testing was used as a filler task in our laboratory, and the priority given to the other tasks over our twelve years of testing created a situation in which many RS were interrupted by and then resumed after another experiment. During data filtering, we discarded all the RS which were interrupted by a break of three days or more, as well as all the incomplete RS interrupted by another experiment. Therefore, the RS retained for data analyses were all preceded by another complete RS during which the subjects had to select a different target. Secondly, to verify the coherence of our data, we reasoned that the chance of randomly finding the correct target during the first trial of a novel session (i.e., immediately after a switch) should be of the same order of magnitude as the probability of making an error during the last 10 trials of the preceding session (because the monkeys could not know that the rule had just changed). We computed the average score of the first trial of the RS retained for data analysis and compared that score to the percentage of incorrect responses obtained at the end of the preceding session. From this analysis, we found that all baboons behaved as expected during the first trial, except for two monkeys (Mako and Muse) that produced for reasons that we could not explain more correct responses on the first trial than expected by chance. We therefore cautiously removed these two subjects from analysis, considering that a sufficiently large number of subjects (N = 22) remained available for meaningful conclusions. Finally, we removed from analysis all the RS that started with a first successful trial because (1) these correct responses might have been produced by chance and (2) the rewarding of the first response after the shift might affect learning curve in the subsequent trials of the RS [23]. The data analysed in this study have an average of 577 RS per participant (see details in Supplemental Table 2), and 14997 set shiftings at the group level, which is well beyond the order of magnitude of past behavioural studies.

2.6. Data analyses

Our aim was to explore the hypothesis that cognitive flexibility as inferred by our CSST task is age dependent in baboons. To do so, we calculated the average age of the subjects during the study period and used individual mean ages to assign each subject to an age class. Age classes were defined as follows. "Young": up to 60 months (5 years old max); "Adult": from 61 to 130 months (5–10.8 years old); "Middle-age": from 131 to 200 months (10.8–16.7 years old), and "Old": from 201 to 300 months (16.7 years old and more). The "Young" class corresponds to

Table 2

Results of Perseveration. Bayesian model considering the Age class, Sex, the Current dimension and Number of trials of the session as factors. SD: standard Deviation, SE: standard error, Conf.low: Lower confidence interval, Conf.high: Higher confidence interval, confidence intervals that do not overlap zero are in bold.

Perseveration Bayesian Model				
Term	Group	Estimate		
sd_(Intercept). Name	Name	0.412		
sd_Ntrial_Session.Name	Name	0.0158		
cor_(Intercept). Ntrial_Session. Name	Name	0.185		
Term	Estimate	SE	Conf.low	Conf. high
(Intercept)	-0.651	0.214	-1.09	-0.220
Age Class: Adult	-0.567	0.262	-1.11	-0.0321
Age Class: Middle-age	-0.239	0.258	-0.772	0.296
Age Class: Old	0.612	0.279	0.0483	1.18
Ntrial_Session	-0.0599	0.00794	-0.0764	-0.0436
Current_Dimension: Shape	0.110	0.0753	-0.0350	0.255
Sex: Male	0.284	0.190	-0.115	0.670
Age Class: Adult by Ntrial_Session	-0.0139	0.0103	-0.0347	0.00694
Age Class: Middle-age by Ntrial_Session	0.00267	0.00974	-0.0172	0.0227
Age Class: Old by Ntrial_Session	0.0166	0.0104	-0.00575	0.0388
Ntrial_Session by Current_Dimension: Shape	-0.00847	0.00207	-0.0126	-0.00436
Age Class: Adult by Current_Dimension: Shape	-0.00711	0.0886	-0.177	0.165
Age Class: Middle-age by Current_Dimension: Shape	-0.163	0.0800	-0.319	-0.00844
Age ClassOld by Current_Dimension: Shape	-0.166	0.0816	-0.321	-0.00931

Table 3

Results of Learning latency. Bayesian model considering the Age class, Sex and the Current dimension of the session as factors. SD: standard Deviation, SE: standard error, Conf.low: Lower confidence interval, Conf.high: Higher confidence interval, confidence intervals that do not overlap zero are in bold.

Learning latency bayesian model				
Term	Group	Estimate		
sd_(Intercept). Name	Name	13.0		
sd_Observation.Residual	Residual	14.6		
Term	Estimate	SE	Conf. low	Conf. high
(Intercept)	**36.5**	**6.49**	**23.3**	**49.6**
Age Class: Adult	-11.9	7.81	-28.6	3.81
Age Class: Middle-age	-4.81	7.80	-20.6	10.6
Age Class: Old	**25.1**	**8.39**	**7.53**	**42.4**
Current_Dimension: Shape	**-1.44**	**0.519**	**-2.45**	**-0.427**
Sex: Male	-3.09	5.78	-14.7	8.74
Age Class: Adult by Current_Dimension: Shape	-0.230	0.647	-1.47	1.09
Age Class: Middle-age by Current_Dimension: Shape	**-1.84**	**0.656**	**-3.09**	**-0.572**
Age Class: Old by Current_Dimension: Shape	**-3.29**	**1.30**	**-5.80**	**-0.693**

Table 4

Results of the 5 first Response times. Bayesian model considering the Age class, Sex, the Current dimension and Number of trials of the session as factors. SD: standard Deviation, SE: standard error, Conf.low: Lower confidence interval, Conf.high: Higher confidence interval, confidence intervals that do not overlap zero are in bold.

Response times of 5 first trials Bayesian Model				
Term	Group	Estimate		
sd_(Intercept). Name	Name	346		
sd_Ntrial_Session.Name	Name	65.4		
cor_(Intercept). Ntrial_Session. Name	Name	-0.789		
sd_Observation.Residual	Residual	735		
Term	Estimate	SE	Conf. low	Conf. high
(Intercept)	**1331**	**195**	**944**	**1724**
Age Class:Adult	68.5	238	-407	537
Age Class: Middle-age	138	228	-347	592
Age Class: Old	-311	248	-811	193
Ntrial_Session	64.1	39.8	-17.3	145
Current_Dimension: Shape	30.8	85.2	-137	199
Sex: Male	30.7	102	-181	236
Age Class: Adult by Ntrial_Session	-38.8	47.7	-136	57.6
Age Class: Middle-age by: Ntrial_Session	-18.0	46.2	-114	79.8
Age Class: Old by Ntrial_Session	**118**	**51.6**	**12.4**	**220**
Ntrial_Session by Current_Dimension: Shape	-2.96	19.1	-40.8	34.5
Age Class: Adult by Current_Dimension: Shape	-72.4	66.5	-202	57.9
Age Class: Middle-age by Current_Dimension: Shape	-29.8	66.4	-157	96.1
Age Class: Old by Current_Dimension: Shape	36.3	75.2	-106	179

the juvenile period until puberty. The "Adult" class includes sexually mature individuals with the development of secondary sexual characteristics, ready for reproduction, which corresponds to young adults. The "Middle-age" class corresponds to older adults, and the "Old-class" corresponds to the period of life approaching and exceeding the life expectancy in the natural environment.

We explored three dependent variables that are related to task performance to assess age effects in cognitive flexibility. The first one was perseveration represented by the number of times the target from the previous session had been selected after the shift. The response at the very first trial of the session was not included in the calculation of perseverative error, because for the first trial the subject had not yet received feedback indicating that the rule had changed. The second dependent variable was learning latency described the time it took the individual to learn the new rule (we used the number of trials before ten successful trials were performed in a row). The last one was Response time (RT), which was defined as the time (in ms) it took the participant to select a response.

The independent variables were Age class (described previously, see Table 1), Sex (male or female), and Current dimension of the target (either shape or colour). Age was included because our main hypothesis was that cognitive flexibility is age dependent in baboons. The sex was included because previous studies found subtle differences between sex in reversal learning [24]. The current dimension was also included, because previous studies have shown that the baboons performed better with the shape than with the colour dimensions[15]. Note that the comparison between intra- (e.g., from one colour to another one) and inter-dimensional shifts (e.g., from colour to shape) was made impossible because of uneven baseline probabilities in these two conditions. This is because there were three possible targets (e.g., all three colours) in the case of an interdimensional shift, whereas there were only two possible targets in case of intra-dimensional (e.g., green or pink, if the previous target was yellow). To account for repeated measures, we used a random intercept per individual and a random slope where appropriate (see details of the models in Tables 2–5).

We used the Rstan package [25] of R to study mixed models of the variables of interest using a Bayesian approach [26]. For perseveration, we used a binomial GLMM with logit link function in which each trial was coded as one, if it was a perseveration, or zero otherwise. For learning latency and reaction times we used LMM models. For perseverative errors and reaction times, the computation time using the entire dataset was excessive, we therefore decided to randomly select a maximum of 50 RS per individual and verified that successive runs of the model with different random sets gave similar results. For learning latency, we had only one value per session (indicating the number of trials until ten successive trials were correct) and therefore included the full

Table 5

Results of the last 45 Response times. Bayesian model considering the Age class, Sex, the Current dimension and Number of trials of the session as factors. SD: standard Deviation, SE: standard error, Conf.low: Lower confidence interval, Conf.high: Higher confidence interval, confidence intervals that do not overlap zero are in bold.

Response times of 45 last trials Bayesian Model				
Term	**Group**	**Estimate**		
sd_(Intercept). Name	Name	217		
sd_Ntrial_Session.Name	Name	4.55		
cor_(Intercept). Ntrial_Session. Name	Name	-0.159		
sd_Observation.Residual	Residual	664		
Term	**Estimate**	**SE**	**Conf. low**	**Conf. high**
(Intercept)	**1530**	**111.**	**1306**	**1757**
Age Class: Adult	**-285**	**132**	**-557**	**-12.4**
Age Class: Middle-age	-82.6	132	-342	187
Age Class: Old	**488**	**141**	**200**	**781**
Ntrial_Session	**-10.1**	**2.23**	**-14.6**	**-5.58**
Current_Dimension: Shape	**-60.1**	**19.0**	**-97.1**	**-23.2**
Sex: Male	103	94.7	-83.6	296
Age Class: Adult by Ntrial_Session	-2.48	2.75	-8.15	3.44
Age Class: Middle-age by Ntrial_Session	-2.60	2.72	-8.22	2.91
Age Class: Old by Ntrial_Session	3.87	3.04	-2.27	9.94
Ntrial_Session by Current_Dimension: Shape	**1.18**	**0.454**	**0.310**	**2.09**
Age Class: Adult by Current_Dimension: Shape	11.8	17.9	-24.2	46.6
Age Class: Middle-age by Current_Dimension: Shape	9.09	17.5	-25.6	43.7
Age Class: Old by Current_Dimension: Shape	**-46.8**	**19.7**	**-86.1**	**-8.36**

dataset. Note that the goal of the Bayesian approach is not significance testing but to find confidence intervals of model's parameters given the data, therefore giving an estimate of the importance and precision of the parameters of interest (for an introduction to Bayesian statistics, see Kruschke, [27]).

2.7. Data availability

The data and code to reproduce the figure and results is available at: DOI 10.17605/OSF.IO/E267K.

3. Results

3.1. General trends

Results at the group level are illustrated in Fig. 1. Fig. 1A shows that the proportion of perseveration is high following the change of rule and decreases extremely rapidly during the first five trials, mean = 0.48 (SE = 0.03) on the 2nd trial, mean = 0.31(SE = 0.03) on the 5th. This proportion then decreases more progressively, mean = 0.21 (SE = 0.2) on the 10th to reach a mean = 0.05 (SE = 0.01) on the 50th trial.

Fig. 1B shows that the success rate of the group is very low at the onset of the session and increases drastically during the first 5 trials after the rule change, mean = 0.31 (SE = 0.01) at the second trial, mean = 0.46 (SE = 0.02) at the 5th. In a complementary way to the perseverative errors, that increase slows down afterwards to reach the average score of 0.91 (SE = 0.014) at the 50th trial.

Finally, Fig. 1C indicates a pronounced slowing down of RT during the first 5 trials after the rule change, with mean = 1421 ms (SE = 40 ms) on the second trial until mean = 1644 ms (SE = 49 ms) on the 5th. RT then accelerates during the learning of the new target (mean = 1475 ms, SE = 65 ms on the 10th trials) before stabilizing when learning is achieved (mean = 1095 ms, SE = 85 ms) on the 50th trial.

3.2. Age effect

Results on age effects are reported in Fig. 2. They show the same trends for perseveration, learning latency, and RTs. We found that the Adult group has the best performance with the smallest mean perseveration on the 50 first trials (mean = 3.65, SE = 0.38), the fastest learning latency (mean = 22.87 trials, SE = 1.16) and the shortest RTs (mean = 961.25 ms, SE = 34.42). Young group is actually worse, with higher perseveration (mean = 6.85, SE = 1.67), longer learning latency (mean = 35.23 trials SE = 7.96) and longer RTs (mean = 1277.78 ms SE = 88.26). Worse than adults and better than juveniles, Middle-aged baboons show an average number of perseverations of 5.34 (SE = 0.64), a learning latency of 28.86 trials (SE = 1.79) and RT of 1163.48 ms (SE = 69.10). Finally, older baboons show the worst results. Old baboons have the highest mean number of perseverations of 12.5 (SE = 2.08), the longest learning latency of 58.86 trials, and the longest RT of 1790.14 ms (SE = 157.58). Said differently, inspection of the above three dependent variables suggest that baboons experience strong age effects in cognitive flexibility, with the following ordering of cognitive flexibility capacities: Old < Young < Middle age < Adult. Fig. 3.

The results of the Bayesian models are provided in Tables 2 to 5. They support inspection of the results, revealing similar age effects for perseveration, learning latency, and response times. They indicate that the adults perform best with a significant difference from the younger and Middle-age groups, who perform worse. In turn, older baboons had significantly higher perseveration, learning latency and response times than the other three groups (note that due to the dynamic change in RT (slowing down then acceleration), we studied independently the first five trials and the remaining 45 ones).

There is finally a significant interaction between trial position and age class in later RTs. From the 1st to the 5th trial, the oldest individuals experience the greatest slowing down after the rule change, with a steeper slope than the other 3 age classes (see Tables 4 and 5).

3.3. Sex effect

None of the models revealed an effect of sex on performance (see Tables 2 to 5). Males and females showed similar results for perseveration (mean Females = 6.86, SE = 1.13; mean Males = 5.94, SE = 1.03), for learning latency (mean Females = 36.21 trials, SE = 4.96; mean Males = 30.11 trials, SE = 2.41) and for response times (mean Females = 1250.28 ms, SE = 98.35; mean Males = 1263.37 ms, SE = 95.14).

3.4. Current dimension effect

There is a small but consistent interaction between the dimension of the object and age classes. Overall, the results suggest that for older individuals the performance is slightly better when the discrimination of the object is done on the shape rather than the colour, whereas there is no such effect for younger classes (see Fig. 4 and Tables 2–5).

4. Discussion

Not surprisingly, our study confirms Bonté et al. [15]'s conclusions that the baboons are endowed with efficient cognitive flexibility. They understood very quickly (at the group level within two or three trials after their first error) that the rule had changed, and that they had to adapt their response to the new rule. This was evidenced by their score in the task, as well as the transient slowing down of the response times which was only observed during the 5 trials that occurred immediately after the rule change. It remains unclear at this stage if this slowing down in response time reflects the inhibition of the previously learned rule, the process of mentally searching for the novel correct response, or both.
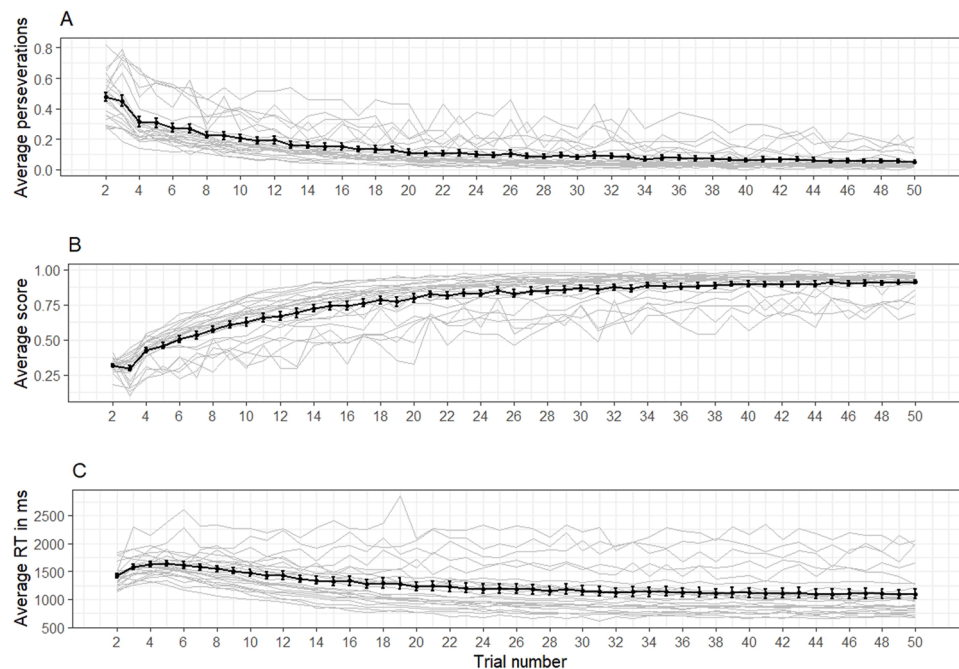
Fig. 1. General performance at the group level during the first 50 trials after the change of rule. A: Number of perseverative errors, B: Score, C: Response time in ms. In light grey, average for each individual, in black group mean (error bars represent standard errors).
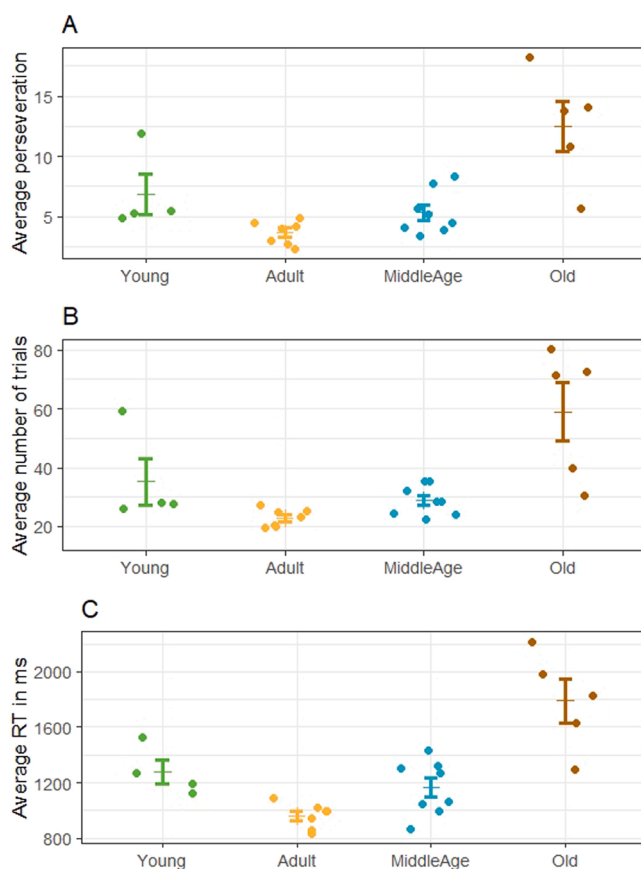

Fig. 2. Results by age class on the first 50 trials, A: average number of errors due to perseveration, B: mean learning latency is the number of trials necessary to cumulate 10 consecutive successful trials, C: Average response time in ms. Dots represent individual means, crosses with error bars are group means (+/- SE).

4.1. Effects of expertise

Our study is unique regarding the length of exposure of the subjects to the CSST task, and size of the dataset. Never had a group of non-human primates been tested for so long on a cognitive flexibility task. The undeniable advantage of this was that the measures of cognitive flexibility were obtained well after learning had occurred, providing an accurate measure of an "expert" form of cognitive flexibility independent of learning processes. Because our CSST task served as a filler task in our laboratory, strict data filtration was essential. Objective criteria independent of the expected results were applied for data filtering, eliminating around half of the RS collected, but guaranteeing reliable and accurate results.

What was the added values of years of training in our task? To address that question, we compared our perseveration results with those of Bonté et al. [15] which used the same task and species as in the current study, moreover on some of the same individuals. Fig. 5 reports individual data on perseverative error for the subgroup of 13 baboons who had served as subjects in both Bonté et al. [15] and in the current study. Because Bonté et al. [15] computed perseverative errors over the first 25 trials after a shift, Fig. 5 only considers these first 25 post-shift trials in both studies. Remember that our subjects are now 10 years older than in Bonté et al., and that they have thus either moved from young-hood or adult-hood to middle-age, or from middle age to old-hood (see the x-axis of Fig. 5).

Fig. 5 shows a decrease in the number of perseverative errors after 10 years of exposure to the task. This finding confirms that the small number of shifts in Bonté et al. [15] (N = 3) were insufficient for revealing optimal performance. We believe that the same limitation may have occurred in most previous studies inferring cognitive flexibility from a limited number of set shifting (e.g., Lacreuse et al. [26] and Moore et al. [13,25]). In our study, it is likely that the extremely small numbers of perseverative errors probably correspond to the best possible performance of our baboons, thanks to the use of a self-service cognitive testing device, the ALDM, allowing long-term testing. Fig. 5 also indicates that this decrease in the number of perseverative errors was observed in all age groups, and that it was even found in the oldest baboons such as Brigitte or Kali who made less perseveration errors in
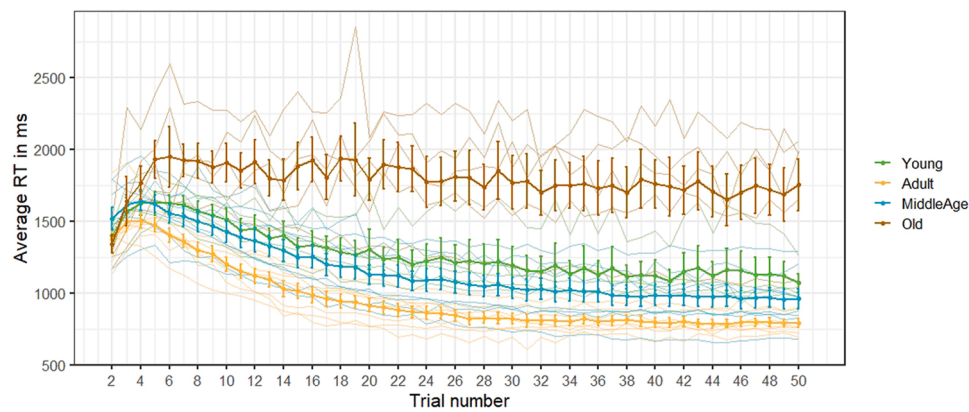
Fig. 3. Average response times in ms by age class on the 50 first trials. In light colour, individual means, in darker colour group mean (error bars represent standard errors).
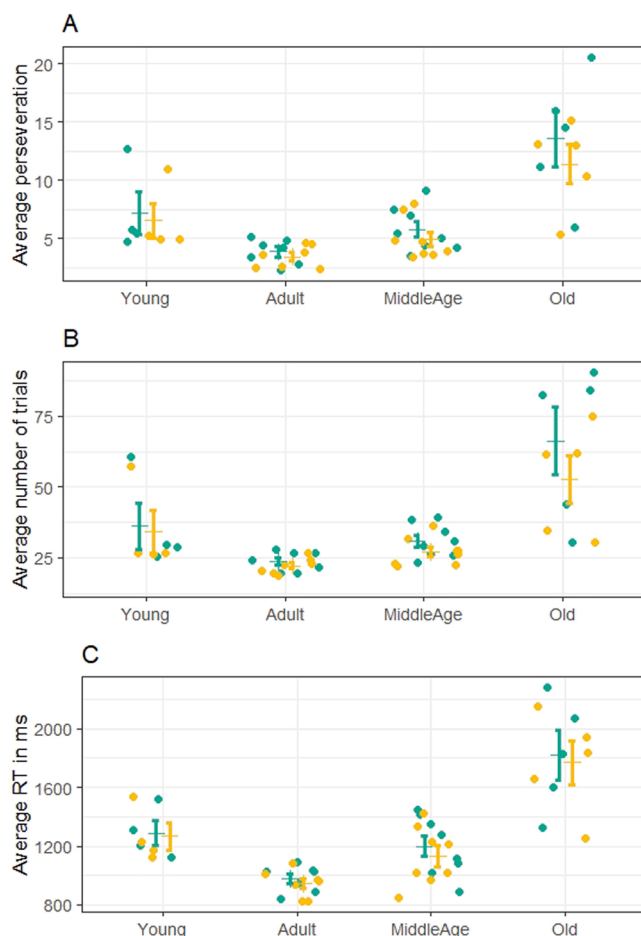

Fig. 4. Age Class and Dimension Interaction effect on our measures of cognitive flexibility, A: Perseveration, B: Learning latency, C: RT in ms, Colour dimension results are green points, Shape dimension results are yellow. Dots represent individual means, crosses with error bars are group means (+/- SE).

the current study in spite of their age than in Bonté et al..[3] This important finding suggests that the exposure to the task for 10 years improved cognitive flexibility in our monkeys, and that training had

beneficial effects in all age groups.

4.2. Age effects

Our study also indicates important effects of age on cognitive flexibility. Overall, we found that the Adult group performed better than the other three age groups, as inferred by their lower rate of perseveration errors after a rule change, faster learning latencies, and shorter response times. When compared to the Adult group, we found a significant decrease in all of these measures of performance in the Old group. Similar results were already reported in macaques of comparable age [13]. In addition, these effects of ageing were already visible in the middle-aged baboons, whose perseveration rates, learning latencies and response time were intermediate between Adults and Older groups. These results are consistent with those of Moore et al. [28] in rhesus macaques. They also converge with the findings from a female chimpanzee (*Pan troglodytes*, [29]), although a strict comparison between apes and monkeys is difficult due to different lifetimes. Bonté et al. [15] suggested that the decline in executive control occurred earlier in baboons (approximatively 6–8 years), but their analyses were different from ours and their conclusions were drawn from a linear regression between age and the number of perseverative errors. We believe that the current analyses are more sensitive for revealing age difference on cognitive flexibility, and this assumption is confirmed by Fig. 5 showing identical age effects in both studies, with the lowest mean of perseverative errors in Adults, and its decline in the Middle-aged class. In a review of the literature on the ageing of the prefrontal areas, Luebke et al. [30] (for a complementary review, see Peters and Kemper [31]) reported that ageing induces a structural change of the myelin sheaths in macaques leading to a reduced "conduction velocity and timing in neuronal circuits" (page 212). Luebke et al. [30,31] further described a regression of dendritic trees in the upper layers of the prefrontal cortex of aged macaques, as well as a loss of dendritic spines and synapses, and an alteration of transmitters and receptors leading to a reduction of inputs to cortical neurons. Our behavioural findings suggest that similar effects might have occurred in our baboons.

Our analyses further reveal poorer performance in the Young group, in comparison to Adults. Admittedly, this could be explained by a shorter exposure to the task and a shorter training period, compared to other age classes who worked on this task for longer period of time. However, given the number of RS completed by these individuals during the two years of testing (range 85–1324, see Table 1), we think that the amount of exposure does not fully explain this difference. These poorer performance for young baboons could rather be explained by their still developing cognitive flexibility and the underlying brain structure at that age (for a study on macaques, see [32]). This conclusion is reminiscent of Weed et al. [12], who showed similar differences between
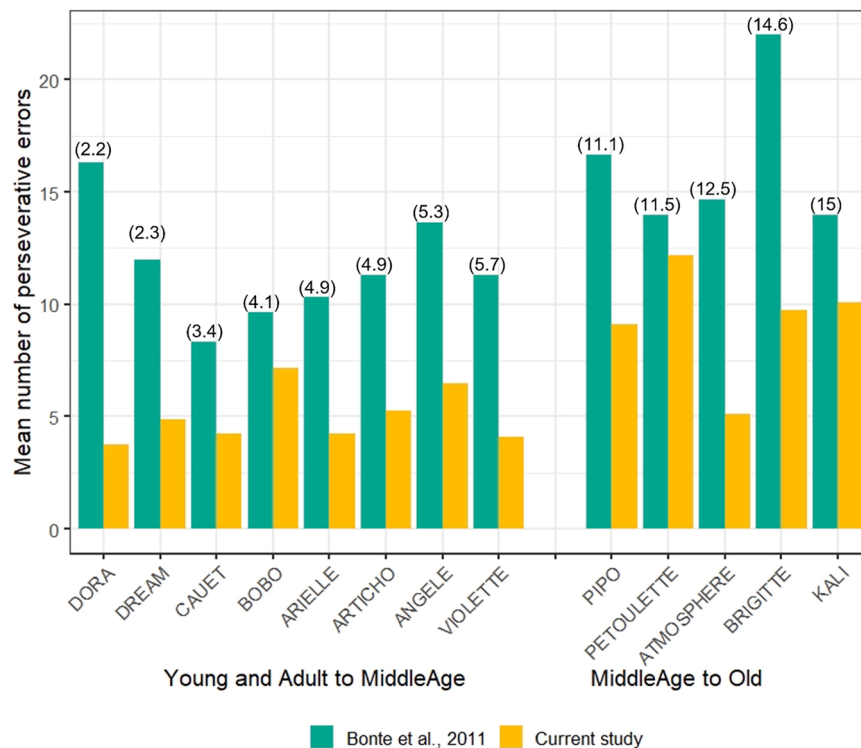
---

[3] The correlation between the age of the subjects and the difference between the number of the perseveration errors in the two studies was not significant (Spearman correlation, rs = 0.04).

Fig. 5. Average number of perseverative errors for the 13 baboons who participated to both Bonté et al. [15] and the current study. Data corresponds to perseverative errors obtained during the first 25 post-shift trials. Subjects are ordered by Age class, with the baboons considered in Middle-age and Old-age class in the current study being represented on the right-hand side of the graph. Numbers in bracket indicates the age of the subjects (in years) when they were tested by Bonté et al.

juveniles and adults in the rhesus macaques in the ID/ED task. The class of 24 juvenile macaques made more perseverative errors and more errors in the set formation and set-shifting components of the ID/ED task than the 16 adult macaques, and this difference between age groups emerged despite an identical amount of exposure to the ID/ED task. Fig. 5 further reports individual data on perseverative error for the subgroup of 13 baboons who had served as subjects in both Bonté et al. [15] and in the current study. All subjects showed the same pattern of results, corresponding to a systematic decline in the number of perseverative errors after 10 years of exposure to the task. This finding on baboons suggests that exposure to tasks or situations requiring cognitive flexibility has the potential of reducing cognitive impairment that normally occurs with aging. Additional studies will be required to determine if that effect of exposure remains limited to cognitive flexibility mechanisms, or if it also pervades other cognitive domains.

In a different perspective, our results are reminiscent of the findings previously reported on humans. Thus, 3-year-olds children have the ability to sort cards on one dimension but fail when the relevant dimension changes [6,7]. However, at 4 years of age, their WCST performance begins to improve until late adolescence [8,33], which is consistent with the performance of our youngest baboons. Performance is then optimal between the ages of 20 and 40 for all executive functions, but it then declines with ageing [34–36], with deficits already appearing in middle-aged between 40 and 50 years of age [9], again in accordance with the current study on baboons. These declines can also be explained by structural changes in the prefrontal cortex that become more pronounced after age 65 [10]. Taken together, results obtained on humans [9], chimpanzees [29], macaques [13,28], and baboons (current study) all indicate parallel developmental and aging trends regarding cognitive flexibility. This conclusion implies that (1) cognitive flexibility and underlying brain mechanisms were already present in our common ancestor in a similar form as in modern humans, 30–40 million years ago, and that (2) these non-human primate species can serve as

reasonable animal models for both behavioural and brain studies on that function.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data Availability

The data and code to reproduce the figure and results is available at: DOI 10.17605/OSF.IO/E267K.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.bbr.2022.114043.

References

[1] A. Miyake, N.P. Friedman, M.J. Emerson, A.H. Witzki, A. Howerter, T.D. Wager, The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis, Cogn. Psychol. 41 (2000) 49–100.
[2] F. Borjon, Flexibilité cognitive et résolution de problèmes, PhD Thesis, Université Lumière Lyon 2, 2016.
[3] E.A. Berg, A simple objective technique for measuring flexibility in thinking, J. Gen. Psychol. 39 (1948) 15–22.

[4] B. Milner, Effects of different brain lesions on card sorting: the role of the frontal lobes, Arch. Neurol. 9 (1963) 90–100.

[5] A.R. Luria, Higher cortical functions in man, Basic Books, Oxford, England, 1966.

[6] P.D. Zelazo, U. Müller, D. Frye, S. Marcovitch, G. Argitis, J. Boseovski, J.K. Chiang, D. Hongwanishkul, B.V. Schuster, A. Sutherland, , The development of executive function in early childhood, Monogr. Soc. Res. Child Dev. (2003) i–151.

[7] N. Chevalier, A. Blaye, Le développement de la flexibilité cognitive chez l'enfant préscolaire: enjeux théoriques, L'année Psychol. 106 (2006) 569–608.

[8] N.J. Cepeda, A.F. Kramer, J. Gonzalez de Sather, Changes in executive control across the life span: examination of task-switching performance, Dev. Psychol. 37 (2001) 715.

[9] S. Daigneault, C.M. Braun, H.A. Whitaker, Early effects of normal aging on perseverative and non-perseverative prefrontal measures, Dev. Neuropsychol. 8 (1992) 99–114.

[10] H. Haug, R. Eggers, Morphometry of the human cortex cerebri and corpus striatum during aging, Neurobiol. Aging 12 (1991) 336–338.

[11] R. Dias, T.W. Robbins, A.C. Roberts, Dissociation in prefrontal cortex of affective and attentional shifts, Nature 380 (1996) 69–72, https://doi.org/10.1038/380069a0.

[12] M.R. Weed, R. Bryant, S. Perry, Cognitive development in macaques: attentional set-shifting in juvenile and adult Rhesus monkey, Neuroscience 157 (2008) 22–28.

[13] T.L. Moore, R.J. Killiany, J.G. Herndon, D.L. Rosene, M.B. Moss, Impairment in abstraction and set shifting in aged Rhesus monkeys, Neurobiol. Aging 24 (2003) 125–134, https://doi.org/10.1016/S0197-4580(02)00054-4.

[14] T.L. Moore, R.J. Killiany, J.G. Herndon, D.L. Rosene, M.B. Moss, A non-human primate test of abstraction and set shifting: an automated adaptation of the Wisconsin Card Sorting Test, J. Neurosci. Methods 146 (2005) 165–173, https://doi.org/10.1016/j.jneumeth.2005.02.005.

[15] E. Bonté, T. Flemming, J. Fagot, Executive control of perceptual features and abstract relations by baboons (Papio papio, Behav. Brain Res. 222 (2011) 176–182, https://doi.org/10.1016/j.bbr.2011.03.034.

[16] J. Fagot, E. Bonté, Automated testing of cognitive performance in monkeys: use of a battery of computerized test systems by a troop of semi-free-ranging baboons (Papio papio, Behav. Res. Methods 42 (2010) 507–516, https://doi.org/10.3758/BRM.42.2.507.

[17] I. Barbet, J. Fagot, Processing of contour closure by baboons (Papio papio), J. Exp. Psychol.: Anim. Behav. Process. 37 (2011) 407–419, https://doi.org/10.1037/a0025365.

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

[19] J. Fagot, R. Thompson, Generalized relational matching by Guinea Baboons (Papio papio) in two-by-two-item analogy problems, Psychol. Sci. 22 (2011) 1304–1309, https://doi.org/10.1177/0956797611422916.

[20] A. Formaux, D. Paleressompoulle, J. Fagot, N. Claidière, The experimental emergence of convention in a non-human primate, Philos. Trans. R. Soc. B: Biol. Sci. 377 (2022) 20200310, https://doi.org/10.1098/rstb.2020.0310.

[21] J. Fagot, D. Paleressompoulle, Automatic testing of cognitive performance in baboons maintained in social groups, Behav. Res. Methods 41 (2009) 396–404.

[22] F. Mansouri, K. Tanaka, Wisconsin card sorting test with macaque monkeys, Int. Congr. Ser. 1250 (2003) 105–118, https://doi.org/10.1016/S0531-5131(03)00975-0.

[23] R. Cook, J. Fagot, First trial rewards promote 1-trial learning and prolonged memory in pigeon and baboon, Proceedings of the National Academy of Sciences. 106 (2009) 9530–9533. https://doi.org/10.1073/pnas.0903378106.

[24] M. LaClair, M. Febo, B. Nephew, N.J. Gervais, G. Poirier, K. Workman, S. Chumachenko, L. Payne, M.C. Moore, J.A. King, A. Lacreuse, Sex differences in cognitive flexibility and resting brain networks in middle-aged marmosets,

ENEURO.0154-19.2019, ENeuro 6 (2019), https://doi.org/10.1523/ENEURO.0154-19.2019.

[25] Stan Development Team, R Interface to Stan, (2020). https://mc-stan.org/rstan/ (Accessed 21 December 2021).

[26] C. Muth, Z. Oravecz, J. Gabry, User-friendly Bayesian regression modeling: a tutorial with rstanarm and shinystan, Quant. Methods Psychol. 14 (2018) 99–119, https://doi.org/10.20982/tqmp.14.2.p099.

[27] J. Kruschke, Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, (2014).

[28] T.L. Moore, R.J. Killiany, J.G. Herndon, D.L. Rosene, M.B. Moss, Executive system dysfunction occurs as early as middle-age in the rhesus monkey, Neurobiol. Aging 27 (2006) 1484–1493, https://doi.org/10.1016/j.neurobiolaging.2005.08.004.

[29] A. Lacreuse, L. Parr, L. Chennareddi, J.G. Herndon, Age-related decline in cognitive flexibility in female chimpanzees, Neurobiol. Aging 72 (2018) 83–88, https://doi.org/10.1016/j.neurobiolaging.2018.08.018.

[30] J. Luebke, H. Barbas, A. Peters, Effects of normal aging on prefrontal area 46 in the Rhesus monkey, Brain Res Rev. 62 (2010) 212, https://doi.org/10.1016/j.brainresrev.2009.12.002.

[31] A. Peters, T. Kemper, A review of the structural alterations in the cerebral hemispheres of the aging rhesus monkey, Neurobiol. Aging 33 (2012) 2357–2372, https://doi.org/10.1016/j.neurobiolaging.2011.11.015.

[32] L. Malkova, E. Heuer, R.C. Saunders, Longitudinal magnetic resonance imaging study of rhesus monkey brain development, Eur. J. Neurosci. 24 (2006) 3204–3212, https://doi.org/10.1111/j.1460-9568.2006.05175.x.

[33] M.C. Davidson, D. Amso, L.C. Anderson, A. Diamond, Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching, Neuropsychologia 44 (2006) 2037–2078.

[34] K.R. Ridderinkhof, M.M. Span, M.W. van der Molen, Perseverative behavior and adaptive control in older adults: performance monitoring, rule induction, and set shifting, Brain Cogn. 49 (2002) 382–401, https://doi.org/10.1006/brcg.2001.1506.

[35] M.M. Span, K.R. Ridderinkhof, M.W. van der Molen, Age-related changes in the efficiency of cognitive processing across the life span, Acta Psychol. 117 (2004) 155–183, https://doi.org/10.1016/j.actpsy.2004.05.005.

[36] N.S. Wecker, J.H. Kramer, B.J. Hallam, D.C. Delis, Mental flexibility: age effects on switching, Neuropsychology 19 (2005) 345–352, https://doi.org/10.1037/0894-4105.19.3.345.

[37] G. Stoet, L.H. Snyder, Single neurons in posterior parietal cortex of monkeys encode cognitive set, Neuron 42 (2004) 1003–1012.

[38] F.A. Mansouri, K. Tanaka, Behavioral evidence for working memory of sensory dimension in macaque monkeys, Behav. Brain Res. 136 (2002) 415–426, https://doi.org/10.1016/S0166-4328(02)00182-1.

[39] F.A. Mansouri, K. Matsumoto, K. Tanaka, Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog, J. Neurosci. 26 (2006) 2745–2756.

[40] T. Kamigaki, T. Fukushima, Y. Miyashita, Cognitive set reconfiguration signaled by macaque posterior parietal neurons, Neuron 61 (2009) 941–951, https://doi.org/10.1016/j.neuron.2009.01.028.

[41] T. Kamigaki, T. Fukushima, Y. Miyashita, Neuronal signal dynamics during preparation and execution for behavioral shifting in macaque posterior parietal cortex, J. Cogn. Neurosci. 23 (2011) 2503–2520.

[42] M.J. Buckley, F.A. Mansouri, H. Hoda, M. Mahboubi, P.G.F. Browning, S.C. Kwok, A. Phillips, K. Tanaka, Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions, Science 325 (2009) 52–58, https://doi.org/10.1126/science.1172377.

[43] F.A. Mansouri, M.J. Buckley, D.J. Fehring, K. Tanaka, The role of primate prefrontal cortex in bias and shift between visual dimensions, Cereb. Cortex 30 (2020) 85–99.