Simians in the Shape School: A comparative study of executive attention

Kristin French¹ · Michael J. Beran¹ · Kimberly Andrews Espy² · David A. Washburn¹

Published online: 8 January 2018
© Psychonomic Society, Inc. 2018

Abstract

Executive functions (EF) have been studied extensively in children and adults. However, EF tasks for young children can be difficult to administer and interpret. Espy (1997, *Developmental Neuropsychology*, 13, 495–499) designed the Shape School task to measure inhibition and switching in preschool-aged children. Shape School presents cartoon-like characters that children must flexibly name by their color, their shape, or both, depending on cues that indicate the appropriate rule. Shape School has been found to be age sensitive as well as predictive of performance on other EF tasks. We presented a computerized analogue of Shape School to seven rhesus macaques. Monkeys were trained to categorize characters by color or shape, or to inhibit this response, depending on whether the characters had eyes open, eyes closed, or wore hats. Monkeys performed above chance on the inhibition and switching components of the task. Long runs of a single classification rule and long runs of noninhibition trials had no significant impact on performance when the rule changed or inhibition was required. This nonverbal adaptation of Shape School can measure EF in nonhuman animals and could be used in conjunction with other EF tasks to provide a clearer picture of both human and nonhuman executive functions.

Keywords Executive attention · Shape school · Rhesus monkeys · Comparative cognition

Executive function (EF) is a higher order construct used to describe the processes by which attention, perception, memory, and other cognitive processes are controlled effortfully, intentionally, and typically consciously (e.g., Diamond, 2012, 2013; Logan, 1985). Linked closely to frontal-lobe functions and assessed by an array of tests (see reviews in Lyon & Krasnegor, 1996), there has been much debate in the literature as to what the multifactorial construct EF means and what the executive functions are. However, there is general agreement that three core executive functions are inhibition, flexibility, and updating; additional control processes such as monitoring and planning have also been proposed as executive functions that build on the core operations (e.g., Miyake & Friedman, 2012; Miyake et al., 2000). Executive

functions are the focus of an exponentially growing number of studies because they appear to be important in a wide range of everyday competencies, and failures of EF are implicated in a large number of behavioral and psychological disorders, such as attention-deficit/hyperactivity disorder (e.g., Barkley, 1997; Willcutt, Doyle, Nigg, Farone, & Pennington, 2005), autism spectrum disorders (e.g., Ozonoff, 1997; Ozonoff, Pennington, & Rogers, 1991), eating disorders (Fagundo et al., 2012), Tourette's syndrome, and obsessive-compulsive disorder (e.g., Watkins et al., 2005).

Testing the EF capabilities of children provides an understanding of how and when the various components might develop, and accurate EF assessment is critical in determining impairments and possible interventions. However, many tasks assess more than one aspect of EF simultaneously, and it can be difficult to distinguish performance on one component from another, to compare across measures, and to separate EF from lower level functions (Henry & Bettenay, 2010; Salimpoor & Desrocher, 2006). As a result, the connection between cognitive task performance and EF ability can be unclear. These challenges are further compounded when attempting to adapt EF tasks originally intended for adults or to create entirely new ones for children. This is because differences in verbal, motor, attention skills, and general knowledge between children and adults and between children at

✉ Kristin French
kfrench4@student.gsu.edu

David A. Washburn
dwashburn@gsu.edu

¹ Department of Psychology and Language Research Center, Georgia State University, P.O. Box 5010, Atlanta, GA 30302-5010, USA

² Department of Psychology, University of Arizona, Tucson, AZ, USA

different ages pose obstacles to age-appropriate and informative EF assessment in children (e.g., Isquith, Crawford, Espy, & Gioia, 2005). For example, EF regulates and modulates other cognitive processes and because these processes may develop at different points and rates throughout childhood, difficulties in assessment arise when development of each facet is not completed. Poor EF performance may reflect only one underdeveloped component rather than EF in its entirety, obscuring estimations of EF abilities for other aspects that are developed or developing (Pritchard & Woodward, 2011; Wiebe et al., 2011).

In an effort to improve EF tasks for use with children, Wiebe et al. (2011) stressed that tasks should have few extraneous and unrelated demands, simple rules and verbal requirements, and familiar concepts and knowledge universally understood and mastered by all children. Nonetheless, studies of EF often involve several different tasks of varying complexity and demand, which is beneficial for achieving multiple measures of EF components but can be time-consuming and tiring for children (e.g., Henry & Bettenay, 2010). Tasks can assess only one or two EF processes but may differ with respect to the aspects of EF studied, the measures taken to ensure age-appropriateness, and maturation sensitivity. To relieve these difficulties, most tasks have imposed restrictive age ranges for children and, as a result, are best suited primarily for older children, leaving few tasks available for preschool-aged children (Henry & Bettenay, 2010; Pritchard & Woodward, 2011).

Notwithstanding, studying EF in preschool-aged children is crucial to understanding the development of higher order cognitive processes, due to the rapid development of EF during the preschool years and its role in future educational, behavioral, and social success (Pritchard & Woodward, 2011). Along with the difficulties of EF assessment in children outlined above, preschoolers are a subgroup that requires still more consideration when studying EF. Some tasks that are appropriate for older children are too complex for preschoolers, involving multiple higher order cognitive processes to complete successfully (Espy, 1997; Henry & Bettenay, 2010). For example, many tasks require switching and inhibiting simultaneously, making it challenging to distinguish between age-sensitive components of EF (Espy, 1997), and furthering the importance of determining which aspects of EF develop when and how best to separate them when creating EF tasks for young children.

Of the tasks that do exist for young children, many measure a variety of aspects of EF, including inhibition, switching, working memory, and planning, or any combination thereof. Inhibition is the suppression of goal-irrelevant information and prepotent responses, and it is an aspect of EF that allows for adaptive responding in the face of response competition where an individual is faced with conflicting information about what they should do (Diamond, 2013; Friedman &

Miyake, 2004; Roberts & Pennington, 1996). Some tasks have used inhibition of natural responses, such as the day–night Stroop task in which children must respond with “night” when shown a white card with a sun and with “day” when shown a black card with moon and stars (Gerstadt, Hong, & Diamond, 1994). However, this task has been found to be difficult for children under the age of 5 years due to the associations of pictures and words not being automatic. The go/no-go task can include both arbitrary and natural response suppression, and its stimuli and presentation have taken many forms (Wiebe et al., 2011). The dimensional change card sorting (DCCS) task imposes an arbitrary prepotent response, requiring children to sort cards by color, shape, or number, depending on a predetermined rule (Isquith et al., 2005). The DCCS also requires switching from one rule to another. Switching is indicative of cognitive flexibility and involves switching or shifting from one previously engaged response set to another. Children must simultaneously inhibit the previous sorting rule used in DCCS while also shifting to the new sorting rule. These tests are appropriate measures of executive function but have proven difficult for younger children, and many present an indistinguishable overlap of switching, inhibition, attention, verbal ability, and working memory, creating confusion over a pure measurement of executive functions in young children.

Espy (1997) sought to overcome these issues through development of the Shape School task. Shape School was created specifically for preschool-aged children to verbally measure inhibition and switching both independently and simultaneously. Many of the challenges detailed above are addressed in the Shape School task, as it measures several components of EF but does so in a cohesive and simple way, so participants feel as if they are completing one task told as a story rather than multiple distinct tasks. Additionally, the separation of baseline, inhibition, and switching measures allows for independent assessment of these EF processes.

In the Shape School task, inhibition and switching are introduced in steps. The task is presented in a storybook fashion either manually or on a computer. The stimuli initially are characters with arms, legs, and faces that differ on shape (circle or square) and color (blue or red). In the first condition, children are instructed to name the characters by their color, and the resulting data act as a baseline for color naming speed and ability. The second condition introduces characters with either happy or sad faces, and children are told to only name the happy-faced characters by their color while ignoring the sad-faced characters. This allows for a measure of inhibition. Switching is measured in the third condition in which some characters have hats and some do not. Children are instructed to name the hatted characters by their shape and to name the unhatted characters by their color. In the final condition, inhibition and switching are measured simultaneously. Happy-faced characters are named by either their color or their

shape, depending on whether they are hatted or unhatted, while sad-faced characters are ignored. The storybook method and staged rule introduction is more age appropriate for preschoolers than other executive function tasks, and the separation of inhibition and switching by condition allows for an assessment of each measure independently in order to determine developmental differences and age effects.

Espy (1997) found that Shape School performance is age sensitive, with children under the age of 5 years struggling with the switching and combined conditions but showing little trouble with the inhibition condition. This suggests that inhibition may develop earlier than switching and that the development of inhibition may be a necessary foundation before switching can occur successfully (also see Pritchard & Woodward, 2011). James, Choi, Wiebe, and Espy (2016) reported that Shape School performance also was demonstrated to be significantly correlated with other executive-function tasks, including tests of updating (e.g., Nine Boxes test, Delayed Alternation test), inhibition (e.g., Stroop, go/no-go), and set-switching (trail-making), for children as young as 3 years. Shape School is also related to later math and reading abilities (Clark et al., 2013; Espy, Bull, Martin, & Stroup, 2006).

Our research focus was on the utility of the Shape School task to assess executive function across species, toward the goal of better understanding the phylogenetic emergence of EF as well as its developmental emergence. To understand EF as a regulatory system requires understanding how it may have emerged through evolutionary selection, and a comparative approach to studying EF and other aspects of so-called higher order cognition is the obvious way to do this (e.g., Dias, Robbins, & Roberts, 1996; Smith, Redford, Beran, & Washburn, 2010; Washburn, 2016). Taking this perspective requires that paradigms that have been demonstrated to be diagnostic of EF in human adults and children must also be administered to nonhuman primates, in tests that approximate as much as possible the validated procedures. This was the goal of the present project.

Nonhuman primate cognition shares many similarities with human cognitive processes (Maestriper, 2003; Roitblat, Bever, & Terrace, 1984; Shettleworth, 2009; Tomasello & Call, 1997; Zentall & Wasserman, 2012), and executive functioning likely is no exception. Indeed, some of the earliest and most influential studies of the neural bases of EF reflect biobehavioral studies with nonhuman primates (e.g., Goldman-Rakic, 1987, 1995). A variety of tasks, such as the Wisconsin Card Sort Task (WCST) and the Stroop task (e.g., Beran, Washburn, & Rumbaugh, 2007; Michelet et al., 2015; Washburn, 1994, 2016) have indeed made their way from human testing to monkey and ape testing, revealing much about the commonalities in EF between human and nonhuman primates. For example, Nakahara, Hayashi, Konishi, and Miyashita (2002) compared fMRIs of monkeys and humans and found that similar brain areas

involved in set shifting in both species were activated during completion of a WCST. Moore, Schettler, Killiany, Rosene, and Moss (2009) described the effects of damage to the macaque prefrontal cortex on WCST performance, and reviewed the results of ablation studies on memory updating and other aspects of EF. Michelet et al. (2015) described activity in the macaque dorsal anterior cingulate cortex corresponding to executive control in performance of a Stroop-task analog. Additionally, monkeys and humans exhibited similar responses to stop signals during a motor inhibition task and both species changed their behavior depending on the probabilities of those signals, suggesting that human and nonhuman primates share some similarities when inhibiting motor movement (Wattiez, Poitou, Rivaud-Péchoux, & Pouget, 2016).

Nonhuman primate research thus offers the opportunity to study EF from not only a developmental perspective but also from an evolutionary one. Differences in human and nonhuman primate EF can be just as informative as similarities, as evidenced by Avdagic, Jensen, Altschul, and Terrace's (2014) finding that rhesus macaques differ from humans on set switching in terms of interference from irrelevant stimuli and delays when switching from one rule to the next, or Washburn's (2016) demonstration that incentive differentially improves humans', compared to macaques', cognitive control. Awareness of where commonalities in EF across species begin and where they end serves to not only inform on EF as a whole but also on measures of the components of EF. And comparing EF performance across species and across developmental stages can provide insight into whether certain parts of EF are foundational and required for the emergence of others, and whether improvement in one aspect can be trained or learned. Thus, we presented rhesus monkeys (*Macaca mulatta*) with their own version of Shape School.

Given that monkeys had never been trained on this task before, we designed a series of parameters to make the task progressively harder, to determine how far monkeys could advance. We first trained monkeys to classify stimuli according to either the color or their shape. Stimuli to be classified by color were those with closed eyes. Stimuli to be classified by shape were those with open eyes. Monkeys learned to make both classifications, and then we introduced inhibitory stimuli. Because we expected it to be difficult to get monkeys to refrain from responding to stimuli, we initially began with 50% of stimuli being inhibitory to help train that new response class. These stimuli looked just like the shape-designated and color-designated stimuli, except that those characters had hats. Thus, the presence of a hat was the cue to make no response to that stimulus. When we realized that monkeys could learn to do this, we reduced the relative frequency of those stimuli to approximately 20%, and then reduced it again to approximately 5% of trials. At that point, monkeys were making many color or shape classifications in a row between

inhibitory stimuli, and so this condition was the strongest test of the monkeys' ability to inhibit responses to specific stimuli.

Method

Subjects

We tested seven adult male rhesus monkeys between the ages of 16 and 26 years. All monkeys previously had been trained to use the computerized apparatus (Rumbaugh, Richardson, Washburn, Hopkins, & Savage-Rumbaugh, 1989). Monkeys had continuous access to the task in test sessions that ranged from 4 to 8 hours. They could engage the computerized apparatus as they chose. Monkeys always had access to water and were given a daily meal each afternoon independent of whether and how much they worked on the computer task. Monkeys also spent time each week with outdoor access, during which time they did not work on computerized tasks.

Apparatus

The monkeys were tested using the Language Research Center's Computerized Test System. This system is composed of a personal computer, digital joystick, color monitor, and pellet dispenser (Richardson et al., 1990). Monkeys manipulated the joystick with their hands so as to move a cursor onscreen during tasks. Monkeys earned 94-mg banana-flavored chow pellets as food rewards (Bio-Serv, Frenchtown, NJ), and those pellets were delivered via a pellet dispenser that was connected to the computer. The software for the tasks was written in Visual Basic 6.0.

Design and procedure

On each trial, a monkey first moved the cursor upward onscreen into contact with a rectangle stimulus in the center of the screen. This was the orienting and initiation response. That stimulus disappeared when contacted, and then the monkey was presented with a visual stimulus in the center of the computer screen (see Fig. 1). This stimulus was an animated figure that was colored blue or red, had the body shape of a circle or a square, and had open eyes or closed eyes. In some cases, the stimulus figure might also be wearing a hat, which required an inhibition response (see Fig. 2 for all stimuli that were to be classified, located at the top, and some example inhibitory stimuli, located on the bottom). In addition, four other response loci stimuli were presented (see Fig. 1). The top stimulus was the "red" response, the bottom stimulus the "blue" response, the left stimulus the "square," response and the right stimulus the "circle" response. All four trial types were randomized for each trial. The correct choice in each of these cases was dictated by the central character. If that character had open eyes, the goal was to classify it as being a circle

or a square. If the eyes were closed, it was to be classified as red or blue. Monkeys made a response by moving that figure directionally to one of the four choice icons. A correct response led to a food pellet reward, and then the trial initiation stimulus appeared to begin the next trial. An incorrect response led to a 20-s time-out during which the screen was blank, and then the trial initiation stimulus appeared. There were four phases to the experiment. Inhibition phases required monkeys to refrain from classification for 5 s. More details are provided below about those trials.

In all phases, monkeys worked as many or as few trials as they chose to complete. We did not have a minimum number required for completion in a session given that we knew they would complete many hundreds (or thousands of trials) per session, which were more than enough to evaluate performance.

Training phase The training phase involved the four trial types as outlined above (see Table 1). Monkeys performed test sessions in this phase until they completed a session at 80% or greater performance for each of the character classifications (red, blue, circle, square).

50% inhibition trials When this phase began, a new class of characters could appear (see Fig. 2). Those characters were the same as the training stimuli, except that they had on hats. This was the indication that the monkey was to inhibit making any of the four classifications for 5 s. Doing so led to a pellet reward. Classifying "hatted" characters as red, blue, circle, or square led to a time-out, as outlined above. Failing to classify "unhatted" characters within 5 s also was scored as an error. Monkeys performed test sessions in this phase until they completed a session at 80% or greater performance for each of the character classifications (red, blue, circle, square) and for successfully inhibiting responses on 80% or more of the trials with the hatted characters. Or, if a monkey failed to reach criterion after eight sessions, that monkey still moved to the next phase, but we considered this performance to be indicative of failure at this level. Each trial had a probability of .50 of being a hatted character rather than one of the characters to be classified.

20% inhibition Trials All details of this phase were identical to the previous phase, except that each trial had a probability of .20 of being a hatted character rather than one of the characters to be classified. If a monkey failed to reach criterion after five sessions in this phase, that monkey still moved to the next phase, but we considered this performance to be indicative of failure at this level.

5% inhibition trials All details of this phase were identical to the previous phases, except that each trial had a probability of .05 of being a hatted character rather than one of the characters to be classified. Each monkey completed only one test session in this phase.

Fig. 1 An example trial. In this example, the stimulus at center was colored blue and had closed eyes, and so the classification rule was for color. The correct response was to move that character to the bottom edge

of the screen where a blue response icon was located. There was a red response icon at the top of the screen. (Color figure online)

Data analysis

Unless otherwise noted, statistical significance was designated as $p < .05$. In each phase, we reported the percentage of trials correctly completed by each monkey, comparing that performance to chance levels using binomial tests. We also analyzed the mean response times of monkeys for each trial type in each phase, using within-subjects ANOVA, and we followed this with paired t tests to assess differences in response time for specific response classifications. In those cases, we used the Bonferroni correction to account for multiple comparisons, and thus a significant effect was set at $p < .008$. To assess the effects of how many noninhibition trials intervened before inhibition stimuli appeared in the 5% inhibition phase, we reported a Pearson correlation coefficient of this number of trials and percentage correct in inhibiting a classification when inhibition stimuli appeared. We also reported the Pearson correlation coefficients for runs of consecutive trials of the same classification rule and percentage of correct classifications on those trials within the runs.

Results

Training phase Because monkeys worked at their own pace, completing as many or as few trials per session as they chose to complete, we collected different numbers of trials from each monkey. During the training phase, the criterion to move to the first test phase was completion of a session with performance at 80% or better in classifying each of the four different stimulus types that could be presented (red stimuli, blue stimuli, circle stimuli, and square stimuli). Table 1 shows the

number of sessions and total number of trials each monkey completed in this phase and all subsequent phases. For one monkey, Willie, we had to institute a correction procedure halfway through this phase because his performance had plateaued at a subcritical level. This correction procedure consisted of re-presenting incorrectly completed trials immediately following the incorrect trial until the correct response was given, so that he could learn each of the classification rules. No other monkey required this, and after this phase with Willie, this procedural modification was discontinued. Figure 3 shows the acquisition learning curves for each monkey during the training phase. Figure 4a shows the performance of each monkey in their final session in this phase of the experiment. Given that on each trial, chance performance was 25% (four response options were available), all of these performance levels were significantly better than chance, as assessed with a binomial test, all $ps < .001$. All monkeys thus progressed to the next phase. Note that in one case, for monkey Han, experimenter error led to progressing him to the next phase despite being only 73% correct with square stimuli. This was still significantly better than chance ($p < .05$, binomial test).

We also examined response times in correctly classifying the stimuli into the two color and two shape categories. The mean response times were 1.07 s ($SD = 0.13$) for square responses, 1.00 s ($SD = 0.26$) for blue responses, 0.94 s ($SD = 0.19$) for circle responses, and 0.72 s ($SD = 0.13$) for red responses. A within-subjects ANOVA indicated that there was an effect of trial type on response time, $F(3, 18) = 6.07$, $p = .005$, partial $\eta^2 = .50$. Red responses were significantly faster than square responses, paired $t(6) = 6.69$, $p = .001$, whereas none of the other comparisons showed significant

Fig. 2 Stimuli used in this experiment. The training stimuli all were to be classified as red or blue in color, or as circle or square shapes. At the bottom of the figure are some examples of inhibition stimuli, with their

hats. Those stimuli required monkeys to not classify them for 5 s for food reward to be delivered. (Color figure online)

differences between response types, all $t(6) < 2.87$, $p > .02$. The most likely reason for this difference was the orienting response to begin a trial was an upward movement, as was the red response, and so it was likely that this was the more

prepotent direction of cursor movement for the monkeys. In any case, these data highlight that by the end of the training phase, monkeys were responding very quickly to all four stimulus types.

Table 1 Number of sessions and number of trials completed by each monkey in each phase

Phase	Training		50% inhibition		20% inhibition		5% inhibition	
	Sessions	Trials	Sessions	Trials	Sessions	Trials	Sessions	Trials
Chewie	2	3,592	1	2,000	1	2,000	1	573
Han	3	2,729	1	1,158	1	1,562	1	701
Lou	3	2,190	5	3,690	1	1,841	1	1,083
Luke	2	2,296	8	5,471	1	862	1	2,000
Murph	4	3,132	5	2,360	1	1,044	1	1,480
Obi	5	2,297	4	2,332	1	672	1	1,796
Willie	6	5,061	1	2,000	5	4,533	1	1,216

Fig. 3 Learning curves for each monkey during the training phase

50% inhibition trials Monkeys again completed variable numbers of trials in sessions given the self-paced aspect of testing. The criterion for moving to the next phase was 80% correct during the final session in all trial types, including the

inhibition trials in which none of the four primary responses (red, blue, circle, square) was to be made. As in the training phase, the use of this criterion ensured that performance exceeded chance levels and was statistically significant as

Fig. 4 **a** Performance of each monkey with all four classification trial types. **b** Performance of each monkey when 50% of the trials were hatted characters that required inhibition of a classifying response. **c** Performance of each monkey when 20% of the trials were hatted

characters that required inhibition of a classifying response. **d** Performance of each monkey when 5% of the trials were hatted characters that required inhibition of a classifying response

assessed with a binomial test, all $ps < .001$. All monkeys except for Luke met that criterion by the completion of eight sessions. Figure 4b shows the performance of each monkey for their final session in this phase of the experiment. As shown in the figure, Luke did not meet 80% correct with red characters but was moved to the next phase despite this, given his otherwise overall good performance.

The mean response times in the last test session for correct responses were 1.07 s ($SD = 0.21$) for square responses, 0.91 s ($SD = 0.31$) for blue responses, 1.04 s ($SD = 0.34$) for circle responses, and 0.73 s ($SD = 0.16$) for red responses. A within-subjects ANOVA indicated that there was an effect of trial type on response time, $F(3, 18) = 4.71$, $p = .013$, partial $\eta^2 = .44$. Given the multiple comparisons and the correction to alpha ($p < .008$), however, no pair of primary response types differed significantly from each other.

20% inhibition trials Monkeys again completed variable numbers of trials in sessions. The criterion for moving to the next phase was the same as in the previous phases. Again, the use of this criterion ensured that performance exceeded chance levels and was statistically significant as assessed with a binomial test, all $ps < .001$. Failure to meet criterion after five sessions still led to a monkey being moved to the next phase, and we again considered this performance to be indicative of failure at this level. Only monkey Willie failed to meet criterion even in his last session, and as shown in Fig. 4c, he never classified circle stimuli correctly despite performing well with all other stimuli.

The mean response times in the last test session for correct responses were 0.95 s ($SD = 0.17$) for square responses, 0.81 s ($SD = 0.22$) for blue responses, 0.97 s ($SD = 0.27$) for circle responses, and 0.61 s ($SD = 0.10$) for red responses. A within-subjects ANOVA indicated that there was an effect of trial type on response time, $F(3, 18) = 11.1$, $p < .001$, partial $\eta^2 = .69$. Red responses were significantly faster than square responses ($p < .001$), whereas none of the other comparisons showed significant differences.

5% inhibition trials Monkeys each completed one test session in this phase, as it was the final phase of the experiment. Figure 4d shows that five of the seven monkeys again were over 80% correct with all trial types in this test session. Luke showed some difficulty with red stimuli and with the inhibitory stimuli. Willie again was excellent with all trial types, except for circle stimuli.

The mean response times in the last test session for correct responses were 0.96 s ($SD = 0.08$) for square responses, 0.76 s ($SD = 0.16$) for blue responses, 0.93 s ($SD = 0.13$) for circle responses, and 0.61 s ($SD = 0.15$) for red responses. A within-subjects ANOVA indicated that there was an effect of trial type on response time, $F(3, 18) = 14.9$, $p < .001$, partial $\eta^2 = .75$. Red responses were significantly faster than circle and

square responses, and blue responses were significantly faster than circle responses, all paired t tests, $p < .008$. By this phase, monkeys were responding very quickly, and with high accuracy in nearly all trial types, including the inhibition trials with the hatted characters.

Transitions between phases It was possible that some of the effects of shifting the proportion of trials that involved the inhibitory nonresponse (i.e., the hatted characters) may have been masked by the large trial counts in each phase of the experiment. For this reason, we focused on the first and last 100-trial block in each phase. Specifically, we assessed performance in the last 100 trials of the training phase, the first and last 100 trials of the 50% inhibition phase, the first and last 100 trials of the 20% inhibition phase, and the first 100 trials of the 5% inhibition phase. In this way, we could determine if any of these shifts from the very end of one phase to the very beginning of the next phase led to significant shifts in performance. These data are presented in Fig. 5. With only one exception, performance tended to remain at the same levels as the monkeys transitioned from the last 100 trials in one phase to the first 100 trials in the next phase. The exception occurred for the first 100 trials in the 50% inhibition phase, which was expected, given that this was the first time the monkeys had to learn to not respond to the hatted stimuli. But, otherwise, the decreasing probability of inhibition trials appearing did not disrupt the monkeys' performances.

Performance as a function of immediately preceding trials

Given the success that the monkeys showed all of the way through the phases of this experiment, we next assessed performance more closely in the 5% inhibition phase. This was in an effort to better understand the effects of having to switch classification rules and to inhibit responding. We chose this phase because it was most similar to human studies of inhibition, in which inhibitory stimuli are presented infrequently. First, we examined performance on inhibition trials as a function of how many noninhibition trials had occurred prior to an inhibition stimulus appearing. The idea here was that monkeys perhaps would have made more errors on these trials after longer runs in which they had not had to inhibit their responses. As noted above, by this point in the experiment, the monkeys were responding to color and shape in under 1 s, on average, and so they were very fast at making those judgments. Figure 6 presents performance on the inhibition trials as a function of how many trials had occurred since the last inhibition trial. Performance remained high no matter the number of trials that occurred between inhibition trials, including in cases where more than 50 trials intervened between those trials with hatted characters. There was no significant relation between the trial runs represented on the x -axis in Fig. 6 (recoded as one to 12 for each of the 12 bars) and the percentage correct, $r(10) = -.41$, $p = .18$.

Fig 5 Performance in selected early or late parts of each phase. The figure shows mean performance on 100-trial blocks that ended or began each of the phases in the experiment. Error bars reflect 95% confidence intervals

With regard to classification rule switching, we examined performance on trial types where stimuli had to be classified according to color or shape as a function of how many trials in a row previous to that trial also required the same classification rule. The idea here is that if monkeys struggled with the switching component of the task, this would be most evident when the previous trial required a different classification than the current trial, whereas monkeys should have performed best (and fastest) for longer runs of trials with the same classification rule in place. With regard to performance, Fig. 7 highlights that overall the monkeys performed as well in each trial type across all run values of previous trials

having the same classification rule. For runs of one to seven consecutive trials with the same classification rule (runs of eight or more trials were rare within the data set), we found no correlation for any stimulus type between percentage of trials correct and length of run of the same classification rule: blue $r(5) = .55$, $p = .20$; red, $r(5) = .49$, $p = .27$; circle $r(5) = .07$, $p = .88$; square $r(5) = -.33$, $p = .47$. Likewise, we found no correlation for any stimulus type between mean response times for trials correctly completed and length of run of the same classification rule: blue $r(5) = .06$, $p = .89$; red, $r(5) = -.44$, $p = .31$; circle $r(5) = .06$, $p = .90$; square $r(5) = .005$, $p = .99$ (see Fig. 8).

Fig. 6 Performance on trials that required making no classification response from the final phase of the experiment. Each bar represents the total performance of all monkeys in the 5% Inhibition phase, and the bars show each possible number of trials that occurred prior to the target

inhibition trial. Thus, the “0” bar reflects trials in which an inhibition trial came immediately after another inhibition trial, whereas a “5” would reflect that the most recent five trials had not required the inhibition of classification

Fig. 7 Performance combined for all trials across monkeys with all four classification trial types in the 5% inhibition phase. Each bar represents performance for a given stimulus type as a function of how many trials in a row had required that same classification discrimination. For example, when a stimulus was to be classified as being blue, the bars show whether this was the first trial in this run to require the monkey to make that classification (red or blue) or this trial was in a run of two to more than

seven such trials in which color also had been the discrimination that was required. Collapsing across trial type, the following numbers of trials occurred for each of the numbers of trials in a row: one in a row: 4,440 trials; two in a row: 2,099; three in a row: 995; four in a row: 441; five in a row: 210 trials; six in a row: 115 trials; seven in a row: 54 trials; more than seven in a row: 44 trials

Discussion

We tested seven adult male rhesus macaques on a computerized analogue of the Shape School task. We were interested in whether the monkeys could learn to perform this cognitive neuropsychological assessment that was originally designed for preschool children and, if so, how their performance would compare to response patterns reported in the developmental literature. The monkeys did learn to classify stimuli according to color or shape, to switch across trials between these two response rules, and to inhibit responses to hatted figures. The task probably should not be considered “simple” for the monkeys, however, as reflected in the performance variability

sometimes seen in certain individuals. For example, monkey Willie varied across phases in terms of which response classes gave him the most trouble. He seemed to simply discard one specific response from various phases, never using that response, and that pattern also changed across the experiment. Perhaps this reflects a working memory deficit, for each of the classification rules, or some other constraint on how he viewed the task and the range of productive responses that could be made.

Whereas children under the age of 5 years typically do well at the inhibition portion of the task, they have displayed difficulty with switching flexibly between response rules—a pattern not seen in the monkeys. Of course, the monkeys required

Fig. 8 Mean response times on correctly completed trials with all four classification trial types in the 5% inhibition phase. Data are presented as in Fig. 6 with regard to how long the run of making the same classification (color or shape) was for each trial type. Error bars reflect 95% confidence

intervals. Collapsing across trial type, the following numbers of trials occurred for each of the numbers of trials in a row: one in a row: 4,047 trials; two in a row: 1,915; three in a row: 905; four in a row: 407; five in a row: 192 trials; six in a row: 103 trials; seven in a row: 50 trials

thousands of training trials to master what children are able to learn from verbal instructions and the Shape School story. Children in other attention shifting and inhibition tasks such as the Wisconsin Card Sort Task also experience a situation in which there is the formation of an attentional set before switching occurs (i.e., they learn to classify according to one rule only before learning the conditional discrimination). Our monkeys learned the conditional discrimination from the outset, and then learned when to inhibit. Thus, it is difficult to compare directly between children and the rhesus macaques due to the vast differences in the number of trials presented, differences in the early training phases, and the benefits of language that are afforded to children, but we do conclude that rhesus macaques can become skilled in both the set switching and the inhibition executive functions that are assessed in this task. It is also noteworthy that the first and last 100 trials of each phase showed little to no decline in performance.

Further, long runs of trials with the same classification rule and long runs with no inhibition trials did not result in the performance costs typically observed in humans. Switch costs are common in set-switching tasks performed by humans. Both children and adults show performance and/or response time costs when switching between rule sets (e.g., Davidson, Amso, Anderson, & Diamond, 2006). Adult humans typically show longer responses when switching between tasks or rules, whereas children show switch costs in accuracy. In the current study, the monkeys were not susceptible to these same performance costs, moving seamlessly from phase to phase, and from response rule to response rule across trials. This replicates the finding by Stoet and Snyder (2006), who also showed no switch costs for monkeys, suggesting that the monkeys may not form response sets in the way that humans do, but rather reset response strategy on trial-by-trial or stimulus-by-stimulus basis. Conversely, Caselli and Chelazzi (2011) did report significant switch costs for macaques and humans, and offered no clear explanation for the difference between their findings and the earlier report. Further complicating the picture, Buschman, Denovellis, Diogo, Bullock, and Miller (2012) found significant switch costs when monkeys changed response sets from orientation to color but not in the opposite direction. It seems clear that the effects of set switching on the accuracy and speed of responses by rhesus monkeys requires further investigation, particularly under conditions that discourage stimulus–response associative responding rather than rule-like relational responding, and thus, potentially rule switching. And, as noted, under conditions in which the training phases are designed to most closely match those used in traditional tests given to human children and adults.

Children show pronounced individual differences in performance on both Shape School and on other EF tasks (Espy et al., 2006). These individual differences in EF

persist into adulthood, and genetics has been shown to play a large part (Friedman et al., 2008). Some individual differences were observed between the monkeys in the present study, although the sample size is too small to explore the reliability or implications of these differences. In the current study, each monkey exhibited differences in the amount of training required to switch and inhibit successfully. However, all were efficient at the task when training was complete. Despite the differences between the monkeys in the number of trials required to reach criterion, most of the animals reached 80% accuracy in each phase quickly, and 5 of 7 monkeys maintained 80% performance with 5% inhibition trials in only one session. It will be important for future studies to determine whether the individual differences in performance on Shape School relate to differences in set-switch and inhibitory skills measured in other tests of executive functioning as well as to examine the relationship of these things with measures of working memory capacity in monkeys. Such studies with non-human primates will allow researchers to disentangle the biological and experiential underpinnings of EF and the interaction of both with development.

For example, because inhibition and switching are carefully measured separately and then together, results from nonhuman primates on Shape School can be compared with their performance on other inhibition and set-switching tasks, such as the Stroop task or card sorting. The Stroop effect is a robust measure of inhibition in humans (MacLeod, 1991, 1992) and has been found to exist also in monkeys and chimpanzees (Beran et al., 2007; Washburn, 1994, 2016). Stroop testing of human adults and children, and of monkeys and apes, has been used to elucidate the cognitive and neural processes by which organisms come to control attention, as well as the performance deficits that occur when the capacity for executive control of attention is lacking. Comparisons of measures like the Stroop task with Shape School can provide new information about EF along with validating measures of inhibition or set switching across species. Switching is a main feature of the WCST, and many studies of EF in nonhuman primates have used analogues to this task, especially for research into the brain areas responsible. For example, Mansouri, Matsumoto, and Tanaka (2006) used a WCST analogue to study the relationship between prefrontal cell activity and performance. Using Shape School on a variety of species and in parallel with other EF measures can provide a greater understanding of EF and a reasonable avenue of cross-species comparisons.

Author note This research was supported by grants from the National Institute of Child Health and Human Development (HD-060563 and HD-38051). Additional support was provided by Georgia State University and the Second Century Initiative. The authors thank Ted Evans for his assistance in collecting these data, as well as the Language Research Center animal care staff.

References

- Avdagic, E., Jensen, G., Altschul, D., & Terrace, H. S. (2014). Rapid cognitive flexibility of rhesus macaques performing psychophysical task-switching. *Animal Cognition*, 17, 619–631.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, 121, 65–94.
- Beran, M. J., Washburn, D. A., & Rumbaugh, D. M. (2007). A Stroop-like effect in color-naming of color-word lexigrams by a chimpanzee (*Pan troglodyte*). *Journal of General Psychology*, 134, 217–228.
- Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D., & Miller, E. K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron*, 76, 838–846.
- Caselli, L., & Chelazzi, L. (2011). Does the macaque monkey provide a good model for studying human executive control? A comparative behavioral study of task switching. *PLOS ONE*, 6, e21489.
- Clark, C. A., Sheffield, T. D., Chevalier, N., Nelson, J. M., Wiebe, S. A., & Espy, K. A. (2013). Charting early trajectories of executive control with the shape school. *Developmental Psychology*, 49, 1481–1493.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44, 2037–2078.
- Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science*, 21, 335–341. doi:<https://doi.org/10.1177/0963721412453722>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.
- Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Primate analogue of the Wisconsin Card Sorting Test: Effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behavioral Neuroscience*, 110, 872–876.
- Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology*, 13, 495–499.
- Espy, K. A., Bull, R., Martin, J., & Stroup, W. (2006). Measuring the development of executive control with the shape school. *Psychological Assessment*, 18, 373–381.
- Fagundo, A. B., De la Torre, R., Jiménez-Murcia, S., Agüera, Z., Granero, R., Tárrega, S., ... & Forcano, L. (2012). Executive functions profile in extreme eating/weight conditions: From anorexia nervosa to obesity. *PLoS ONE*, 7, e43382.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133, 101–135.
- Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, 137, 201–225.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2–7 years old on a Stroop-like day-night test. *Cognition*, 53, 129–153.
- Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. *Child Development*, 52(3), 601–622.
- Goldman-Rakic, P. S. (1995). Architecture of the prefrontal cortex and the central executive. *Annals of the New York Academy of Sciences*, 769(1), 71–84.
- Henry, L. A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and Adolescent Mental Health*, 15, 110–119.
- Isquith, P. K., Crawford, J. S., Espy, K. A., & Gioia, G. A. (2005). Assessment of executive function in preschool-aged children. *Mental Retardation and Developmental Disabilities Research Reviews*, 11, 209–215.
- James, T. D., Choi, H. J., Wiebe, S. A., & Espy, K. A. (2016). II. The preschool problem solving study: Sample, data, and statistical methods. *Monographs of the Society for Research in Child Development*, 81(4), 30–46.
- Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica*, 60, 193–210.
- Lyon, G. R., & Krasnegor, N. A. (1996). *Attention, memory, and executive function*. Baltimore: Brookes.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203.
- MacLeod, C. M. (1992). The Stroop task: The “gold standard” of attentional measures. *Journal of Experimental Psychology: General*, 121, 12–14.
- Maestripietri, D. (Ed.). (2003). *Primate psychology*. Cambridge: Harvard University Press.
- Mansouri, F. A., Matsumoto, K., & Tanaka, K. (2006). Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *The Journal of Neuroscience*, 26, 2745–2756.
- Michelet, T., Bioulac, B., Langbour, N., Goillandeau, M., Guehl, D., & Burbaud, P. (2015). Electrophysiological correlates of a versatile executive control system in the monkey anterior cingulate cortex. *Cerebral Cortex*, 26(4), 1684–1697.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21, 8–14.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Moore, T. L., Schettler, S. P., Killiany, R. J., Rosene, D. L., & Moss, M. B. (2009). Effects on executive function following damage to the prefrontal cortex in the rhesus monkey (*Macaca mulatta*). *Behavioral Neuroscience*, 123(2), 231.
- Nakahara, K., Hayashi, T., Konishi, S., & Miyashita, Y. (2002). Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science*, 295, 1532–1536.
- Ozonoff, S. (1997). Components of executive function in autism and other disorders. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 179–211). New York: Oxford University Press.
- Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, 32, 1081–1105.
- Pritchard, V. E., & Woodward, L. J. (2011). Preschool executive control on the Shape School task: Measurement considerations and utility. *Psychological Assessment*, 23, 31–43.
- Richardson, W. K., Washburn, D. A., Hopkins, W. D., Savage-Rumbaugh, E. S., & Rumbaugh, D. M. (1990). The NASA/LRC computerized test system. *Behavior Research Methods*, 22, 127–131.
- Roberts, R. J., Jr., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology*, 12, 105–126.
- Roitblat, H. L., Bever, T. G., & Terrace, H. S. (1984). *Animal cognition*. Hillsdale: Erlbaum.
- Rumbaugh, D., Richardson, W., Washburn, D., Hopkins, W., & Savage-Rumbaugh, E. S. (1989). Rhesus monkeys (*Macaca mulatta*), video tasks, and implications for stimulus-response spatial contiguity. *Journal of Comparative Psychology*, 103, 32–38.
- Salimpoor, V. N., & Desrocher, M. (2006). Increasing the utility of EF assessment of executive function in children. *Developmental Disabilities Bulletin*, 34, 15–42.
- Shettleworth, S. J. (2009). *Cognition, evolution, and behavior*. New York: Oxford University Press.

- Smith, J. D., Redford, J. S., Beran, M. J., & Washburn, D. A. (2010). Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. *Animal Cognition*, 13, 93–101.
- Stoet, G., & Snyder, L. H. (2006). Effects of the NMDA antagonist ketamine on task-switching performance: Evidence for specific impairments of executive control. *Neuropsychopharmacology*, 31, 1675–1681.
- Tomasello, M., & Call, J. (1997). *Primate cognition*. New York: Oxford University Press.
- Washburn, D. A. (1994). Stroop-like effects for monkeys and humans: Processing speed or strength of association? *Psychological Science*, 5, 375–379.
- Washburn, D. A. (2016). The Stroop effect at 80: The competition between stimulus control and cognitive control. *Journal of the Experimental Analysis of Behavior*, 105, 3–13.
- Watkins, L. H., Sahakian, B. J., Robertson, M. M., Veale, D. M., Rogers, R. D., Pickard, K. M., ... & Robbins, T. W. (2005). Executive function in Tourette's syndrome and obsessive-compulsive disorder. *Psychological Medicine*, 35, 571–582.
- Wattiez, N., Poitou, T., Rivaud-Péchoux, S., & Pouget, P. (2016). Evidence for spatial tuning of movement inhibition. *Experimental Brain Research*, 234, 1957–1966.
- Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108, 436–452.
- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, 57, 1336–1346.
- Zentall, T. R., Wasserman, E. (Eds.). (2012). *Oxford handbook of comparative cognition*. Oxford: Oxford University Press.