Baboons (*Papio papio*), but not humans, break cognitive set in a visuomotor task.

Sarah M. Pope[1,2], Adrien Meguerditchian[2,3], William D. Hopkins[1,4], and Joël Fagot[2,3*]

[1] Neuroscience Institute and Language Research Center, Georgia State University, Atlanta, Georgia, USA
[2] Aix-Marseille University, Laboratory of Cognitive Psychology, UMR7290 CNRS and Federation 3C, , Marseille, France
[3] Brain and Language Research Institute, Marseille, France
[4] Division of Developmental and Cognitive Neuroscience, Yerkes National Primate Research Center, Atlanta, Georgia, USA

Dr. Joël Fagot,
Laboratory of Cognitive Psychology
CNRS and Aix Marseille University
3 Place Victor Hugo, Bât 9, Case C, 13331
Marseille cedex 3, France
Email: joel.fagot@univ-amu.f

Abstract

Cognitive set can be both helpful and harmful in problem solving. A large set of similar problems may be solved mechanically by applying a single solution method. However, efficiency might be sacrificed if a better solution exists and is overlooked. Despite half a century of research on cognitive set, there have been no attempts to investigate whether it occurs in non-human species. The current study utilized a non-verbal, computer task to compare cognitive set between 104 humans and 15 baboons (*Papio papio*). A substantial difference was found between humans' and baboons' abilities to break cognitive set. Consistent with previous studies, the majority of humans were highly impaired by set, yet baboons were almost completely unaffected. Analysis of the human data revealed that children (ages 7-10) were significantly better able to break set than adolescents (11-18) and adults (19-68). Both the evolutionary and developmental implications of these findings are discussed.

Introduction

As problems increase in complexity, so too do their solutions. To mediate the difficulties of solving these complex problems, rules may be established which give the correct answer yet bypass problem reanalysis. Rules allow many similar problems to be solved efficiently and are often used to teach problem-solving skills (Henderson and Pingry 1953; Rohrer and Taylor 2006). However, problem solving by rule-use is not solely beneficial. The difficulty arises when a learned strategy is not the only way to achieve the solution and may actually be less efficient than an alternative. Cognitive set, also termed 'einstellung' or 'mental set,' occurs when a subject successfully learns a rule to solve several problems but is unable to switch to a more efficient method when it becomes available (Luchins 1942; Ruscio and Amabile 1999). Thus, an inability to break cognitive set occurs when that learned solution occludes other problem-solving methods. In other words, once a rule is adopted, other options are not explored. An inability to break cognitive set leads to inefficient problem solving by preventing the use of alternative, sometimes better, problem-solving methods.

In 1942, Abraham Luchins showed that thousands of humans struggled to break cognitive set in order to use a more efficient solution. Luchins' (1942) task required participants to obtain a target quantity by adding and subtracting three given values. The first five of these problems could be solved by a single, somewhat complex, rule. However, these were followed by two hybrid problems, solvable both by the learned rule and a more efficient, direct method. Luchins (1942) found that 70-100% of subjects persisted in using the learned rule rather than switching to the direct method. However, despite its pervasiveness, the underlying causes of humans' susceptibility to cognitive set and potential methods to combat it are opaque.

Several factors influence, to varying degrees, subjects' abilities to break a mental set including instruction (Aftanas and Koppenaal 1962; Luchins and Luchins 1950), working memory availability (Beilock and De Caro 2007), speed requirements (Luchins 1942), amount of training (Luchins 1942; Crooks and McNeil 2009), and similarity between problems (Sweller et al. 1982). Further, although Luchins (1942) reported no substantial age effects on cognitive set, age has been shown to affect other similar types of problem solving. 'Functional fixedness' is described as occurring when an object's use as a tool is dramatically hindered by a subject's experience with it in another functional role (Duncker and Lees 1945). In a tool-use task, Defeyter and German (2003) reported that five-year-old children were unaffected by their previous experience with a tool, yet seven-year-olds and adults easily fell victim to functional fixedness. Despite these accounts, little is known of the differences in cognitive set between children, adolescents, and adults.

Understanding the evolutionary origins of cognitive set may aid in understanding it's pervasiveness in human problem solving. However, previous research on how the mechanization of set may have evolved is nonexistent. This is likely due to the impossibility of a comparative analysis using Luchins' task, which uses arithmetic problems. Studies comparing adults and nonhuman primates using computer paradigms have noted differences in problem-solving performance that may be relevant to set-breaking behavior. First, differences in sequence perception have been reported between non-human primates and adults. Ohshiba (1997) noted that macaques' response times to a simultaneous chaining task increase as they progress through the sequence, indicating that they are using a 'serial search strategy.' Conversely, human adults' responded slowly to the first item in the sequence but quickly to the rest of the items, indicating that they were using a 'collective search strategy' and were mentally identifying the entire sequence before reproducing it (Conway and Christiansen 2001). These results may be applicable to problem-solving in general, with macaques operating in a more local manner (each step is independent) and humans in a more global one (each step is part of the sequence). This is in line with findings comparing perceptual biases between humans and baboons, another old world monkey species. Baboons were found to respond more quickly to local stimuli while human adults responded more quickly to global stimuli (Deruelle and Fagot 1998). If we consider that the key to avoiding cognitive set is likely rooted in an ability to see and utilize the individual steps within a rule, humans' holistic approach may be what is driving their inability to break set. Further, Stoet and Snyder found that macaques' problem solving was more affected by distractions than human adults' (2003), suggesting that they may be less focused on the problem-solving rule and more attentive to individual variation between problems. Conceptually, this may provide old world monkeys with an increased awareness of the alternative method in a cognitive set task.

The current research has two main goals. First, it investigated the evolutionary origins of cognitive set by comparing humans to baboons *(Papio papio*) in a computerized, nonmathematical cognitive set task. We hypothesized that the ability to break set would be different between the two species due to the differences in perceptual and sequential processing between old world monkeys and humans.

Indeed, extreme differences were found between the two species. Baboons were almost entirely immune to the effects of set, while the majority of humans did not break away from the learned rule. Second, the developmental trajectory of cognitive set in humans was analyzed by comparing children, adolescents, and adults. Children were significantly more likely to break cognitive set than either adolescents or adults. These findings are discussed from both evolutionary and developmental perspectives.

Methods:

Subjects and Materials

Baboon data were collected from 15 subjects (ages 1.8-9.3 years), including six males (mean age = 5.3, SD = 2.68) and nine females (mean age = 5.1, SD = 2.36), living in a larger social group of 24 individuals located at the CNRS "Station de Primatologie", Rousset-sur-Arc, France. Baboons were tested via 10 automated learning devices for monkeys (ALDMs; Fagot and Paleressompoulle 2009; Fagot and Bonté 2010), which were directly attached to an outside 700 m^2 enclosure. Subjects had unrestricted access to the ALDMs which consisted of a 70 cm × 70 cm × 80 cm testing chamber with a view port and two hand ports. The view port allowed subjects to see the 19-inch LCD touchscreen monitor (1939L Open-Frame Touchmonitor, Elo Touch Solutions). As subjects reached through the hand ports, a microchip was read for subject identification, which prompted the program to resume the trial list at the appropriate place for that subject. For correct responses, the ALDMs automatically dispensed several grains of dry wheat. The experiment was programmed using EPrime (Version 2.0, Psychology Software Tools, Pittsburgh). The local "Provence Alpes Côte d'Azur" ethic committee for experimental animal research approved the use of the ALDM procedure.

Human data were collected from 104 subjects (ages 7-68), including 40 males (mean age = 26.85, SD = 17.7) and 64 females (mean age = 25, SD = 17.7). Subjects were recruited via a sign which read "Would you like to be a part of scientific study?" and tested at Zoo Atlanta, in Georgia, USA. Humans were tested behind a curtain in a 'booth' along a main path at Zoo Atlanta on a 19-inch LCD touchscreen monitor (1915L Desktop Touchmonitor, Elo Touch Solutions). The experimenter was nearby, but separated from the subject by a curtain and remained inattentive. Additionally, family members often remained in the general vicinity but were asked to remain inattentive and out of sight of the participant.

Participants were given headphones (Koss On-Ear KPH Headphones, KPH7W) to hear sounds elicited by incorrect or correct responses. Correct responses were followed by a cartoon of a present, which increased in size with each correct response. After the fourth correct response, subjects were allowed to choose a sticker and the present size was reset. The same instructions were given to both children and adults explaining the correct/incorrect response screens and stated that they "would need to touch the shapes to figure out the right answer." Subjects could choose to stop testing at any point and 27 (6 adults, 6 adolescents, and 15 children) either stopped of their own volition or were casually stopped (by displaying the end screen) if they had not passed the training within 15 minutes. All human methods were approved by the Zoo Atlanta Research Committee and the Georgia State University Institutional Review Board prior to testing.

Testing

The testing phase consisted of two conditions, baseline (BASE) and probe (PROBE). Trials began after the subject touched a fixation cross. First, two demonstration slides (150 ms each) displayed a 300 x 300 pixels red square in one location and then again in another location (out of four possible locations) on a light blue background (Fig 1). Subjects were then given a response screen, consisting of four white squares in the four available locations on a dark blue background. For correct responses, subjects were required to reproduce the demonstration by touching the two white squares located where the red squares had been in the demonstration, in the correct order. If subjects touched an incorrect square they were shown a 3 second 'time-out' screen before the next trial was cued. In the BASE condition, if subjects accurately reproduced the demonstration, they were presented with a blue triangle in place of one of the two remaining white squares (Fig 1a). To answer correctly, subjects needed to touch the blue triangle. This sequence (touch square1, touch square2, touch triangle) constituted the learned strategy (LS). However, in the PROBE condition, the blue triangle was shown throughout the two demonstration squares and remained visible in the same location on the response screen (Fig 1b). Therefore, subjects could continue to use the LS, but were also able to touch the triangle directly and receive a more immediate reward. This more efficient response (touch triangle) constituted the direct strategy (DS). Importantly, in BASE trials, the triangle was revealed after the first two correct responses. However, if subjects happened to touch where the triangle was located during any of the response screens, they were rewarded. This enabled us to capture the baseline number of time subjects might accidentally touch the triangle's location regardless of it being visible.

Baboons were given 720 testing trials (576 BASE and 144 PROBE), humans 11 years and older were given 96 testing trials (48 BASE and 48 PROBE) and humans under 11 were given 48 testing trials (24 BASE and 24 PROBE). The number of trials presented to humans was reduced to minimize fatigue (as they had to complete the experiment in one sitting). It should be recognized that for baboons, the ratio of BASE to PROBE was 4:1, while for humans it was 1:1. The 4:1 ratio promotes the use of the LS in baboons. This is due to BASE trials not being readily solvable with the DS method, as the triangle is hidden. Therefore, more BASE trials means more LS use. After collecting pilot data indicating that humans preferred the LS even with the 1:1 ratio, we opted to collect an even number of PROBE and BASE trials for humans to minimize the duration of the test sessions, while keeping the number of test trials large enough for statistical analyses and cross species comparison. See Supplementary Figure 1 for Pilot data.

Training

The three training procedures were carried out over the course of 14 days for baboons and immediately preceding the testing phase for humans. Baboons completed an average of 12,945 training trials (SD = 4346), while humans completed an average of 35.2 trials (SD = 18). One of the baboons did not pass training level 2 but successfully passed the more difficult training level 3. Thus, his data were included.

For baboons, training 1 consisted of 96-trial blocks. Trials were randomly selected from 24 possible square/triangle configurations. Each trial began with a fixation cross, followed by two demonstration slides (each 150 ms) showing a red square move from one location to another (out of four possible locations). Next, a response screen was presented, consisting of two white squares in the same locations as the demonstrated squares. The correct response was to touch the squares in the demonstrated order. Baboons completed an average of 5545 (SD = 1947) training 1 trials. To pass training 1, subjects needed to achieve 80% accuracy within a training block, two times (non-consecutively).

For humans, training 1 consisted of 8-trial blocks and demonstration slides were 350 ms each. This is the same training procedure that was used for baboons and the 8 trials were randomly selected from the same 24 possible square/triangle configurations. After each block, the subject's accuracy was assessed. If below 80%, the subject repeated the training level. Accuracy criteria were the same for all training levels. Humans required an average of 13.06 (SD = 7.68) trials to pass training 1.

Training 2 was conducted immediately after Training 1. Demonstration slides' display times decreased to 250 ms for humans. Trials and block composition were identical to Training 1 except that four white squares were given as options during the response phase instead of two (See Fig 1 for example of four square setup). Baboons and humans completed an average of 6095 (SD = 2141) and 10.42 (SD = 8.35) training 2 trials, respectively.

Training 3 is identical to the BASE condition in the testing phase. Demonstration slides' display times were equivalent for baboons and humans at 150 ms (Fig 1a). The experimental phase began immediately after subjects passed Training 3. Baboons and humans completed an average of 1574 (SD = 1000) and 11.70 (SD = 9.18) training 3 trials, respectively.

Data Analysis:

Trials were analyzed to determine whether the LS or the DS was used. Trials in which the subject sequentially touched all three response stimuli (Square1 + Square2 + Triangle) to achieve the reward were classified as having been solved by the LS. Trials in which the subject touched only two (Square1 + Triangle) or one (Triangle) response stimuli to achieve the reward were classified as having been solved by the DS. The Square 1 + Triangle response was included as a DS in an attempt to conservatively maintain the dichotomous LS vs DS paradigm. For each subject, the number of trials in which the DS were used was divided by the number of correct trials completed. This yielded a DS-use ratio for both BASE and PROBE trials. Next, for each trial-type (BASE and PROBE) subjects were classified as preferring the DS (DSer) or the LS (LSer) based on this DS-use ratio. For percent DS use, the median was 0% for both BASE and PROBE conditions and the mean was 1.01% (stdev = 2.16) for BASE and 8.01% (stdev = 20.44) for PROBE trials. Thus, subjects who used the DS in more than 5% of BASE trials were classified as DSers in the BASE condition. Alternatively, subjects who used the DS in fewer than 5% of BASE trials were classified as LSers in the BASE condition. The same highly conservative criteria were used to classify subjects as DSers and LSers in the PROBE condition.

To look at the effects of age, humans were classified into three age groups: Young Children: ages 7-10 (n=27, mean=8.44, stdev=1.15), Adolescents: ages 11-18 (n=25, mean=13.52, SD=2.22), and Adults: ages 19-68 (n =52, mean=40.48, SD=13.18). Age effects were not investigated in baboons, as there was extremely little response variation between subjects.

Results:

Baboons: Mean percentages correct for BASE and PROBE were 80.7% (SD = 4.5) and 82.9% (SD = 11.0), respectively. Combined, subjects used the LS in only 3 PROBE trials (.02 %) of the 1790 PROBE trials compared to 6898 uses (1.02 %) in BASE trials out of the 6969 BASE trials. Additionally, in 20 PROBE trials (.11 %), the baboons touched the correct first red square but then skipped the second and proceeded to touch the blue triangle. Further, all baboons immediately switched to the DS on the first PROBE trial; the three times subjects failed to use the DS were trial numbers 22, 37, and 49. All the tested baboons therefore showed a pronounced and immediate preference for the more efficient, DS method in the PROBE condition and were classified as DSers (see Fig 2).

Humans: Mean percentages correct for BASE and PROBE were 91.2% (SD = 10.1) and 89.5% (SD = 11.1) respectively. Among the104 subjects, only 21 (20.2%) used the DS in greater than 5% of PROBE trials. Of these, only 7 (6.7%) used the DS in over 50% of trials, indicating that they were able to overcome cognitive set and use the more efficient alternative method consistently. Interestingly, 50 humans (48%) used the DS at least once. Thus, even after discovering the more efficient alternative, their set was unbroken.

A Yates' continuity corrected chi-square (used due to an expected value smaller than 5) compared the frequencies of LSers vs DSers in the two species. A significant chi-square (χ^2 (1)= 35.88, p = .000) confirmed that the number of DSers was greater in baboons than in humans (see Fig 2). Additionally, another Yates' continuity corrected chi-square, indicated that there was no significant association between BASE solution strategy classification and species χ^2 (1)= .105, p = .746 (Fig 2).

To investigate the impact of age on DS-use, human subjects were reclassified as DSers or LSers based on their first 48 trials (24 BASE, 24 PROBE). This was done to eliminate the difference in trial number between children (who received 48 trials) and adolescents and adults (who received 96 trials). A Pearson's chi-square revealed that there was a significant association between age group and PROBE solution strategy classification in humans χ^2 (2)= 13.32, p = .001 (Fig 3). Further, the only category in which the standardized residual was significant (2.8) was Children DSers, indicating that they were driving the effect. The association between BASE trials and age group was not significant χ^2 (2)= 1.60, p = .923 (Fig 3); however, the three expected values associated with DS-use for each age group were under 5 (1.6, 1.8, 3.6) indicating that the BASE age results should be cautiously interpreted.

Discussion

In this study, the first main finding was that baboons and humans responded differently on a cognitive set task. Baboons immediately broke set and adopted the more efficient DS when it became available, while the majority of humans failed to deviate from the LS. Our second finding was that humans' ability to break cognitive set is associated with their age. Children were 3 times more likely to be classified as DSers in the PROBE condition than adolescents and 2.4 times more likely than adults. As far as we know, this is the first study to investigate cognitive set in a non-human species and it is one of very few to look at developmental differences in susceptibility to cognitive set in humans (Luchins 1942; Cunningham 1965; Janzen et al. 1976).

Previous findings regarding age effects and cognitive set are inconsistent. Luchins (1942) found a trend (of unreported statistical significance) indicating that public school children (ages 9-14) were less able to recover from cognitive set compared to adults (ages 16-52). Cunningham (1965) tested children ages 7-12 on modified cognitive set tasks and found (minimally reported) trends indicating that older subjects were better able to overcome set. However, Janzen, Maguire, and Boersma (1976) tested children (ages 5-12) on visual set tasks and found no significant age effects. We propose that the LS-DS task is better able to compare cognitive set across ages (and species) than previous methods. While previous set tasks have involved arithmetic (Luchins 1942; Cunningham 1965) and alphabetic rules (Cunningham 1965), the LS-DS task required the use of a spatiotemporal rule: Identify the two demonstrated squares in their demonstrated order. This paradigm allowed us to a) compare set between baboons and humans and b) compare across age groups where all subjects were naïve to the task and it's rules prior to testing. Thus, the LS-DS task may be a better test of cognitive set across age groups because it does not rely on math or language skills, which are very different between children, adolescents, and adults.

Although the LS-DS task was extremely similar between humans and baboons, it was not identical. However, differences in methodology between species should have promoted the opposite of our observed effects and thus, strengthen our findings. Humans received longer display times during training, which could have conferred increased salience to the LS. Yet, overall LS salience was heavily weighted towards baboons as they received an average of 12,915 more training trials than humans. Research suggests that increased training with a rule decreases the likelihood of participants' breaking set (Crooks and McNeil 2009). Thus, baboons should have been less able to break set than humans based on LS experience, which was not the case.

The differential abilities of baboons and humans to break cognitive set are extreme and yet, an underlying cause is not immediately apparent. Why did the baboons immediately consider the DS, whereas humans ignored it? One hypothesis is that differences in visual and sequential processing may have conferred increased perceptual awareness of the DS to baboons. Indeed, the baboons used the DS the very first time it was available and then continued to use it in nearly every subsequent PROBE trial. To do this, they must have (a) been aware of the triangle's premature presence in PROBE trials and (b) associated it, not the sequence as whole, with the reward. In line with Conway and Christiansen's (2001) findings illustrating the serial and collective search strategies of old world monkeys and humans respectively, baboons may have perceived the task's solution as a series of individual stimuli [(Square1) + (Square2) + (Triangle) = Reward] and humans may have perceived it as a collective rule [i.e. (Square1 + Square2 + Triangle) = Reward]. Thus, if baboons solved the LS-DS task with a serial search strategy, it might have allowed the DS [(triangle) = reward] to be visually disentangled from within the LS and thus, used more effectively. Humans, on the other hand, may have used a collective search strategy and been less attentive to the triangle's premature presence in PROBE trials. Further, Bilalić, McLeod, and Gobet (2008) found that previous experience with a solution strategy biased visual attention towards that strategy in expert chess players. However, the possible differences in visual and sequential processing of the LS-DS task between baboons and humans does not explain why only 14 % of humans who used the DS at least once were able to break set. Even if it was accidental, what prevented the majority of humans from switching to the DS after discovering it?

Another explanation for humans' inability to break set is that they simply did not understand that they were allowed to. Humans' notions of how they should respond might block the use of alternative solutions. Since the classic Milgram shock experiments (1974), obedience to authority has been known to affect human behavior and this has been extended to experimenter presence and the experimental environment in general (see Rosenthal and Rosnow 1969 for discussion). For the current study, humans' responses may have been affected by the presence of the experimenter and/or the knowledge that the task was a scientific study. It is possible that they saw the LS as the way they should solve the task based on their experience with the training and the experimental environment. Baboons, on the other hand, had free access to the testing apparatuses, without the presence of an experimenter and are likely unaffected by the experimental environment. This species difference in 'obedient' responses is supported by findings showing that following a live demonstration of how to access food from a box, humans but not chimpanzees imitated superfluous actions (Horner and Whiten 2005). While the current study did not measure humans' conceptual understanding of the task directly, pilot participants were asked if they had thought about touching the triangle directly after task completion. Responses varied from "I didn't see a triangle" to "I thought it was a trap" to "Yes, and I tried it once." However, even the pilot-subject who tried the DS continued to use the LS afterwards, which is consistent with the 43% of non-pilot participants who "discovered" the DS yet continued to use the LS. The question now becomes: If a subject is able to see the early onset of the triangle in PROBE trials and is willing to try touching it directly, what prevents the majority of them from adopting it as a consistent strategy?

We propose that working memory availability plays an important role in humans' persistent use of the LS. In 2007, Beilock and DeCaro found that, when under stress, humans with lower working memory availability used the direct response in Luchins' (1942) task more than humans with higher working memory. They posited that those with higher working memory were better able to remember and enact the learned rule, while those with lower working memory favored the less memory-intensive, direct response. Although the current task did not appear to induce stress in subjects, if we consider that the same working memory constraints might also have driven the increased DS preference in children, who show lower working memory skills than adults (Miles et al. 1996; Thomason et al. 2009), then our age effect becomes more coherent. The LS requires subjects to remember the locations of Square1 and Square2, while the DS only requires the subject to touch the visible triangle. Simply stated, the LS requires working memory and the DS does not. Thus when we consider their lower working memory availability, it seems logical that more children favored the DS than adults who are presumably better equipped to handle the working memory load necessitated by the LS. This is corroborated by the comment of a 7-year-old pilot subject after he discovered the DS, "I like it when the triangle is already there because I don't have to remember the squares!" Further, baboons show overall lower working memory skills than humans (Fagot and De Lillo 2011) suggesting that, while a serial search strategy may allow them to see the DS more readily than humans, their limited working memory could provide increased incentive to use the DS. Adults' and adolescents' persistent use of the LS may simply be a combination of inherent cognitive set and a lack of working-memory based incentive to deviate from what they've learned. This hypothesis should be explored in future studies.

In summary, the current study presents findings suggesting that baboons are less susceptible to the negative effects of cognitive set than humans. This is, as far as we know, the first comparative cognitive set study. It should be noted that in Luchins' original cognitive set task, an 'extinction problem,' where the only possible solution was the direct one, was sometimes used to enhance subjects' ability to break set. The current study did not incorporate an extinction problem but this might have an interesting effect on the observed differences. While future studies are required to more fully understand these species and age differences in ability to break cognitive set, the current study proposes that:

(1) Baboons' immediate use of the DS is facilitated by an increased ability to see the difference between the PROBE and BASE trials, which is a result of independently processing the individual components of the task sequence. Further, continued use of the DS is promoted by its minimal working memory requirements. Free from experimenter effects, baboon responses were unaffected by the experimental environment and their training with the LS.

(2) After extracting the collective LS from the training, humans' persistent use of it may have been governed by a combination of a) difficulty visually differentiating between the PROBE and BASE trials, b)

consideration of how they should respond as dictated by the experimental environment and their training, and c) differences in working memory availability, with lower working memory availability promoting DS-use and higher working memory enabling LS-use.

Cognitive set facilitates complex problem solving. While non-human primates may encounter complex ecological, physical, or social problems, they are likely variable and not easily solvable by a single governing rule. Humans, on the other hand, are regularly faced with complex similar problems, which readily lend themselves to rule-based solutions. The adaptive benefits (or detriments) of cognitive set are not fully understood but it seems logical that set facilitates humans' ritualized problem solving. It would be interesting to address the presence of cognitive set in non-traditionally educated human populations and/or other non-human primate species.

Conflict of interest statement

The authors certify that this research was conducted with no financial, commercial, or other pursuits, which could be construed as potential conflicts of interest.

Acknowledgements

We are grateful to the staff of the CNRS Station de Primatologie (Rousset, France), specially Romain Lacoste and Jean-Christophe Marin, for technical assistance. Marianne Jover and Gérard Meguerditchian are acknowledge for their help in the pilot study. We thank the Zoo Atlanta staff, especially our staff liason Joseph Mendelson, for help and hospitality. S Pope is funded by the Chateaubriand Fellowship from the Ministry of Foreign Affairs and International Development as well as Georgia State University's 2nd Century Initiative Primate Social Cognition, Evolution, and Behavior fellowship. A Meguerditchian is funded by the French National Ambassy Agency (ANR "LangPrimate") grant reference ANR-12-PDOC-0014_01. J. Fagot is funded by the Premilang2" ANR grant ANR-13-BSH2-0002-01

References

Aftanas MS, Koppenaal RJ (1962) Effects of instructional problems and jar position variation on the water-jar Einstellung test. Psychol Rep,10, 359-362.

Beilock SL, DeCaro MS (2007) From poor performance to success under stress: working memory, strategy selection, and mathematical problem solving under pressure. J Exp Psychol Learn, 33, 983.

Bilalić M, McLeod P, Gobet, F (2008) Why good thoughts block better ones: The mechanism of the pernicious Einstellung effect. Cognition, 108, 652-661.

Conway CM, Christiansen MH (2001) Sequential learning in non-human primates. Trends Cogn Sci, 5, 539-546.

Crooks NM, McNeil NM (2009) Increased practice with 'set' problems hinders performance on the water jar task. Proceedings of the 31st Annual Conference of the Cognitive Science Society. Amsterdam, the Netherland, July 29-1. http://csjarchive.cogsci.rpi.edu/proceedings/2009/papers/114/Index.html

Deruelle C, Fagot J (1998) Visual search for global/local stimulus features in humans and baboons. Psychon Bull Rev, 5, 476-481.

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

Fagot J, Bonté E (2010) 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, 507–516.

Fagot J, De Lillo C (2011) A comparative study of working memory: Immediate serial spatial recall in baboons (*Papio papio*) and humans. Neuropsychologia, 49, 3870–3880.

Janzen HL, Maguire TO, Boersma FJ (1976) A Developmental Analysis of Set Patterns in Children: A Normative Study. Proceedings of the American Educational Research Association. San Francisco, California, April 19-23. http://eric.ed.gov/?id=ED124861

Henderson KB, Pingry RE (1953) Problem-Solving in Mathematics. In: Fehr HF (ed) The Learning of Mathematics Its Theory and Practice. Reston, Virginia: National Council of Teachers of

Mathematics, pp 228-270.

Horner V, Whiten A (2005) Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). Anim Cogn 8, 164–181.

Luchins AS (1942) Mechanization in problem solving: The effect of Einstellung. Psychol Monogr, 54, 1-95.

Luchins AS, Luchins EH (1950) New experimental attempts at preventing mechanization in problem solving. J Gen Psychol, 42, 279-297.

Miles C, Morgan MJ, Milne AB, Morris EDM (1996) Developmental and individual differences in visual memory span. Curr Psychol, 15, 53–67.

Milgram S (1974) Obedience to authority. New York: Harper and Row.

Ohshiba N (1997) Memorization of serial items by Japanese monkeys, a chimpanzee, and humans. Jpn Psychol Res, 39, 236-252.

Rohrer D, Taylor K (2006) The effects of overlearning and distributed practise on the retention of mathematics knowledge. Appl Cognitive Psych, 20, 1209-1224.

Rosenthal R, Rosnow RL (1969) Artifact in behavioral research, Vol 121, New York: Academic Press.

Ruscio AM, Amabile TM (1999) Effects of instructional style on problem-solving creativity. Creativity Res J, 12, 251-266.

Stoet G, Snyder LH (2003) Executive control and task-switching in monkeys. Neuropsychologia, 41, 1357-1364.

Sweller J, Gee W (1978) Einstellung, the sequence effect, and hypothesis theory. J Exp Psychol Learn, 4, 513-526.

Thomason ME, Race E, Burrows B, Whitfield-Gabrieli S, Glover GH, Gabrieli JD (2009) Development of spatial and verbal working memory capacity in the human brain. J Cogn Neurosci*,* 21, 316-332.

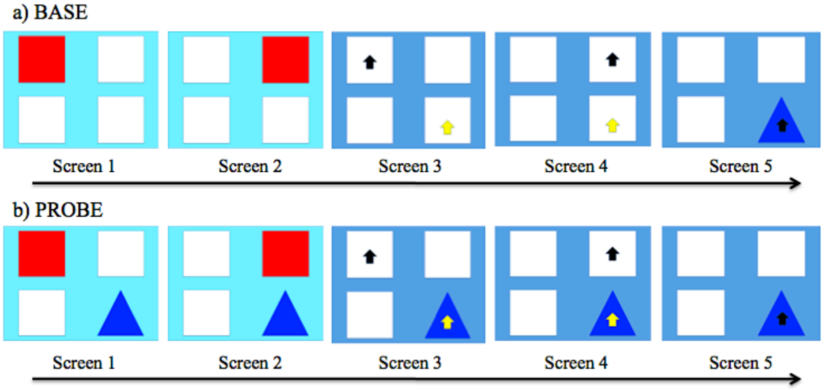


Fig 1. The demonstration (light blue) and response (dark blue) screens for a) BASE and b) PROBE conditions. Black arrows indicate the LS. Yellow arrows indicate where the subject could touch to use the DS. Arrows were not visible during testing. Upon touching the blue triangle, the subject is rewarded.

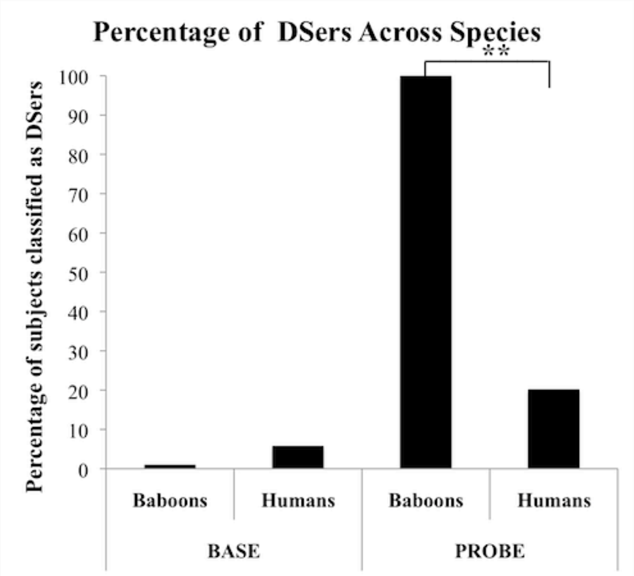


Fig 2. The percentage of subjects classified as DSers in the BASE and PROBE conditions across baboons and humans. ** p ≤.001.

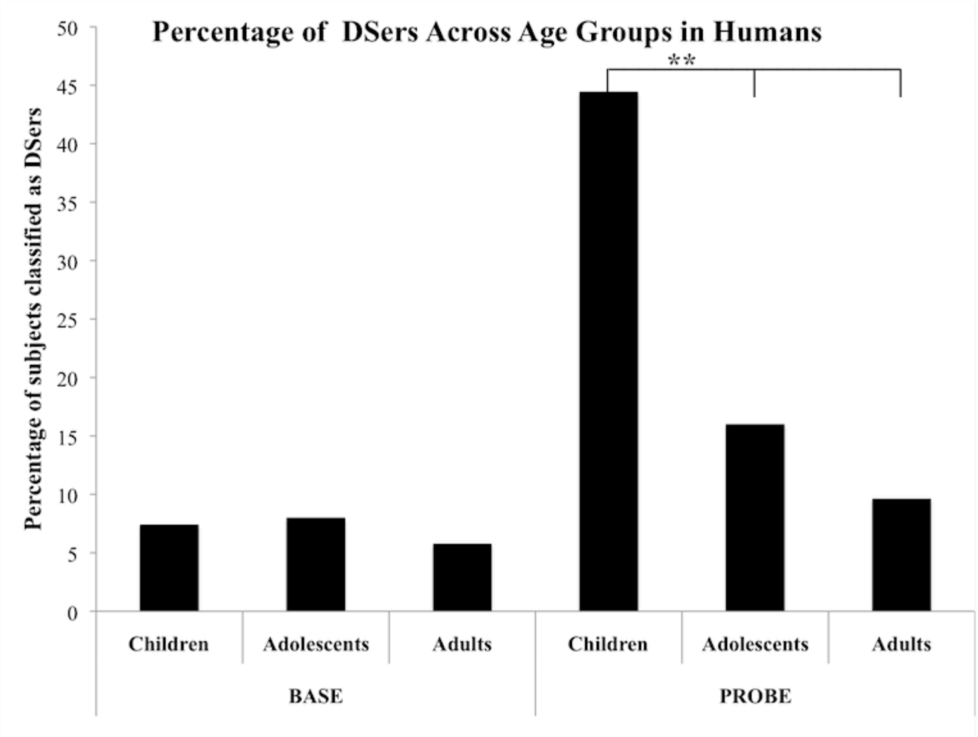
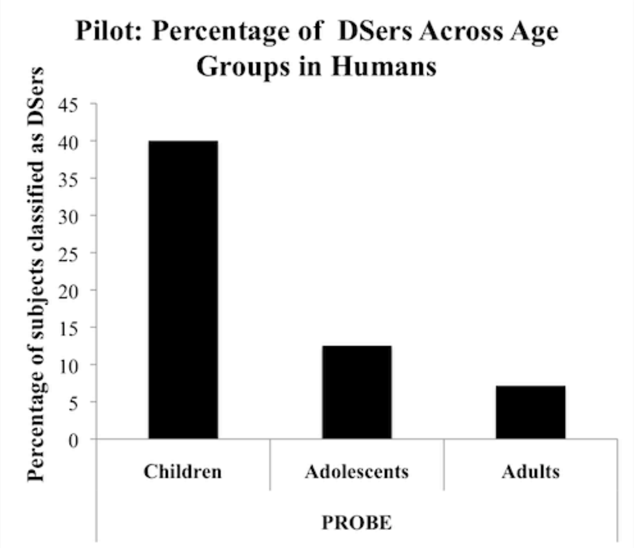


Fig 3. The percentage of subjects classified as DSers in the BASE and PROBE conditions across human children, adolescents, and adults. ** p ≤.001



Supplemental Fig 1. Pilot data was collected on 32 humans (ages 6-51), including 5 children (mean age = 6.4, SD=.55), 8 adolescents (mean age = 14.13, SD=.35), and 14 adults (mean age 36.36, SD= 10.02). Methods were highly similar to those previously described; however, children were given 500ms demonstration slides during testing. Additionally, the first 10 adults were only given 48 testing trials. After a subject noted that she “figured it out at the very end,” the trial numbers were doubled. Once participants had completed all trials, they were asked if they had thought about touching the triangle directly and their responses were recorded. Our results showed that 1 out of 14 (7.14%) adults, 1 out of 8 (12.5%) adolescents and 2 out of 5 (40%) children would be classified as DSers. This is consistent with our later findings.