



HAL
open science

Optional-switch cognitive flexibility in primates: Chimpanzees' (Pan troglodytes) intermediate susceptibility to cognitive set.

Sarah Pope, Joël Fagot, Adrien Meguerditchian, Julia Watzek, Sheina
Lew-Levy, Michelle Autrey, William D. Hopkins

► To cite this version:

Sarah Pope, Joël Fagot, Adrien Meguerditchian, Julia Watzek, Sheina Lew-Levy, et al.. Optional-switch cognitive flexibility in primates: Chimpanzees' (Pan troglodytes) intermediate susceptibility to cognitive set.. Journal of Comparative Psychology, 2020, 134, pp.98 - 109. 10.1037/com0000194 . hal-02412874

HAL Id: hal-02412874

<https://hal.science/hal-02412874v1>

Submitted on 22 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 MS Accepted version published in 2020 in *Journal of Comparative Psychology*, 134(1), 98-109.

2

3 **Optional-Switch Cognitive Flexibility in Primates: Chimpanzees' (*Pan troglodytes*)**

4 **Intermediate Susceptibility to Cognitive Set**

5

6 Sarah M. Pope^{1,2}, Joël Fagot², Adrien Meguerditchian², Julia Watzek¹, Sheina Lew-Levy⁴

7 Michelle M. Autrey³, William D. Hopkins^{1,3}

8

9 ¹ Neuroscience Institute and Language Research Center, Georgia State University,

10 Atlanta, Georgia, USA

11 ² Laboratory of Cognitive Psychology, CNRS, Aix-Marseille Univ, Marseille, France

12 ³ Department of Psychology, University of Cambridge, Cambridge, UK

13 ⁴ Division of Developmental and Cognitive Neuroscience, Yerkes National Primate

14 Research Center, Atlanta, GA, USA

15

16 *Current Affiliation for corresponding author:*

17 Sarah M. Pope, Ph.D.

18 Department of Comparative and Cultural Psychology

19 Max Planck Institute for Evolutionary Anthropology

20 Deutscher Platz 6

21 04103 Leipzig, Germany

22 E-mail: sarah_pope@eva.mpg.de

23

24

25

26

Abstract

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

Within human problem solving, the propensity to use a familiar approach, rather than switch to a more efficient alternative is pervasive. This susceptibility to ‘cognitive set’ prevents optimization by biasing response patterns toward known solutions. In a recent study, which utilized nonverbal touch screen task, baboons exhibited a striking ability to deviate from their learned strategy to utilize a more efficient shortcut. Humans, on the other hand, displayed the opposite response pattern and almost exclusively used a less efficient, but familiar, response. In the current study, we sought to further explore variation in susceptibility to cognitive set within the primate lineage by conducting the LS-DS task with ten chimpanzees (*Pan troglodytes*). Using multilevel multinomial modeling, we found that chimpanzees’ shortcut-use was intermediate to baboons’ and humans’. However, unlike either baboons or humans, there was pronounced inter- and intra-individual variability in chimpanzees’ shortcut-use. Additionally, a subset of chimpanzees employed a unique solution, wherein they switched strategies mid-trial. Further, we found that chimpanzees did not exhibit switch costs when switching between the learned strategy and the shortcut, but humans did. We propose that differences in abstract rule encoding may underlie differences in susceptibility to cognitive set on the LS-DS task within the primate lineage.

Keywords: Cognitive Flexibility, Evolution, Cognitive Set, Optional-Switch, Primates

48

Introduction

49

50

51

52

53

54

55

56

Adaptive behavior is predicated upon flexible strategy-use, yet such plasticity is complex. Flexible responses must integrate external environmental cues with internal inputs, such as past experience. Furthermore, when a past strategy is no longer the most appropriate, flexible behavior requires extinguishing that previous response and switching to a more efficient strategy. Here, we define cognitive flexibility as the ability to incorporate both known solutions and innovated or acquired novel solutions in a contextually appropriate manner (Buttelmann & Karbach, 2017; Lehner, Burkart, & Schaik, 2011).

57

58

59

60

61

62

63

64

65

66

67

Our current understanding of how cognitive flexibility may have evolved is limited. Cognitive flexibility is a multifaced construct, comprised of one's abilities to innovate, switch between, and adopt strategies; yet, it is often only studied in a single context. Typical forced-switch metrics, like reversal or card sorting tasks, *require* subjects to switch strategies. Subjects learn a solution that, at some later point, either a) stops working, or or b) they are instructed not to use anymore. Using forced-switch tasks, the role of cognitive flexibility in developmental and pathological variation in executive functioning has been studied extensively (Doebel & Zelazo, 2015; Luwel, Schillemans, Onghena, & Verschaffel, 2009; Rhodes, 2004; Sullivan et al., 1993; Zelazo et al., 2003). However, within the primate lineage, forced-switch measures yield somewhat conflicting results.

68

69

70

In reversal tasks, wherein a previously successful strategy stops working, humans (even 3-5 year old children) are less perseverative than apes, which in turn are less perseverative than monkeys (Harlow, 1949; Rumbaugh, Savage-Rumbaugh, &

71 Washburn, 1996). That said, with practice (~100 problems), even monkeys are able to
72 perform at or near ceiling, suggesting that their initial perseveration is due to slower
73 acquisition of the new rule rather than difficulty shifting between strategies (Harlow,
74 1949).

75 Card sorting tasks have been used extensively to measure switching ability across
76 primate (and other mammal) species (Brown & Tait, 2016). There are many versions of
77 this paradigm (e.g., Wisconsin Card Sorting Task, Dimensional Change Card Sort and
78 Conceptual Set Shifting Task); however, each requires subjects to sort a series of
79 multivalent cards, first according to one dimension (e.g., color) and then another (e.g.
80 shape; Berg, 1948; Zelazo et al., 2003). Adult humans master the first rule within a few
81 trials, while other non-human primate species take much longer on similar tasks
82 (baboons: $M = 178$; macaques: $M = 137$; chimpanzees: $M = \sim 300-600$ marmosets: $M \sim$
83 180); however, both human and non-human primates exhibit similar degrees of
84 perseveration following the first rule switch (Bonté, Flemming, & Fagot, 2011; Lacreuse,
85 Parr, Chennareddi, & Herndon, 2018; Moore, Killiany, Herndon, Rosene, & Moss, 2005;
86 Roberts, Robbins, & Everitt, 1988).

87 Moreover, in some ways monkeys have been found to be more efficient than
88 humans at switching between known strategies. Switch costs, or deficits in response time
89 or accuracy, occur when switching from one familiar strategy to another (Brass, Derrfuss,
90 & von Cramon, 2007). They are thought to arise from the shift in neural activity
91 associated with how each strategy guides behavior (Lemaire, Luwel, & Brun, 2017;
92 Luwel et al., 2009; Meiran, 1996). Both human children and adults exhibit switch costs
93 on trials which require them to switch from one familiar strategy to another (Arrington &

94 Logan, 2004; Ionescu, 2012; Rogers & Monsell, 1995; Zelazo, 2008), even after
95 extensive practice (Stoet & Snyder, 2007). Yet, rhesus macaques (*Macaca mulatta*)
96 seemingly do not (Stoet & Snyder, 2003; however see: Caselli & Chelazzi, 2011; Huguet,
97 Barbet, Belletier, Monteil, & Fagot, 2014). However, how switch costs affect other non-
98 human primate species remains unclear.

99 The current understanding of strategy-switching is dominated by forced-switch
100 tasks. However, the ability of forced-switch metrics to capture the breadth of cognitive
101 flexibility is up for debate (Arrington & Logan, 2004; Deak & Wiseheart, 2015; Ionescu,
102 2012, 2017). Forced strategy switches are certainly not the only contexts in which an
103 animal might require cognitive flexibility. Often, multiple solutions are present within a
104 problem space and must be selected between for optimal behavior (Brosnan & Hopper,
105 2014). *Optional-switch* paradigms account for some of this complexity by allowing for
106 multiple ‘correct’ solutions that differ in overall efficiency (e.g., time spent, reward
107 value, reduced risk of predation, etc.).

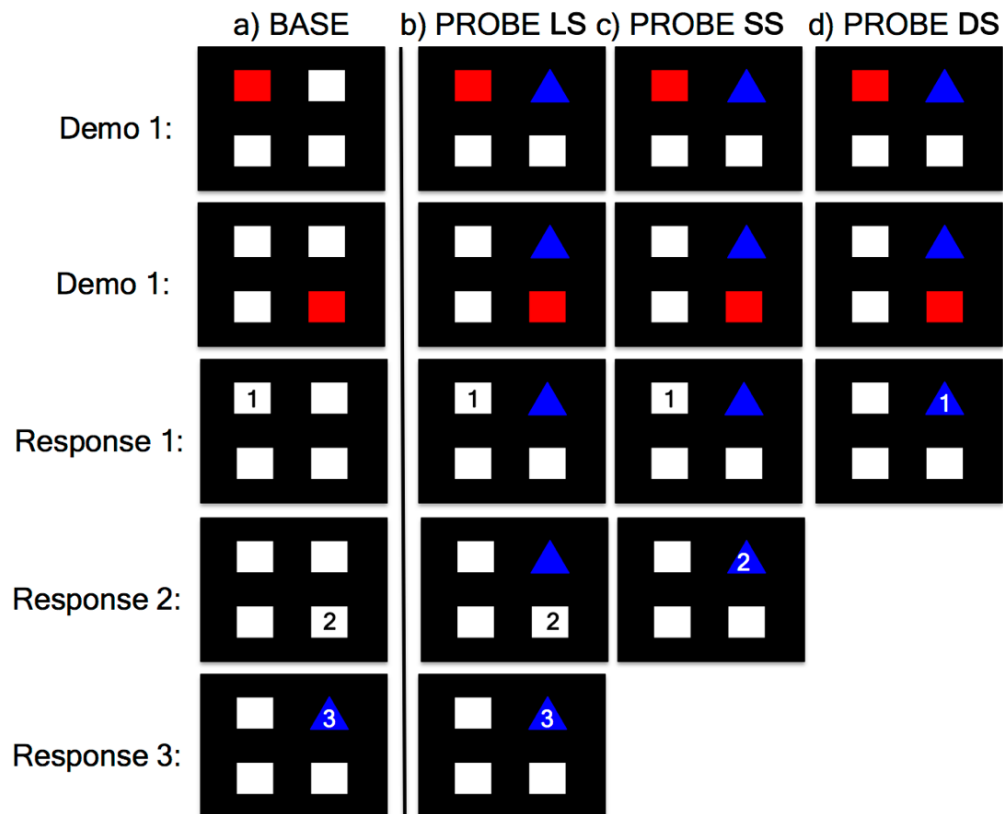
108 Interestingly, humans often fail to select the most appropriate response on
109 optional-switch tasks, primarily when adopting a better alternative would require them to
110 deviate from a learned or familiar method (Adamson, 1952; Aftanas & Koppenaal, 1962;
111 Bilalić, Mcleod, & Gobet, 2008; Chrysikou & Weisberg, 2005; Crooks & McNeil, 2009;
112 Duncker & Lees, 1945; Luchins, 1942; Luchins & Luchins, 1950; Ruscio & Amabile,
113 1999; Sweller, Mawer, & Howe, 1982). We will refer to this ability of learned solutions
114 to block better alternatives, as cognitive set. In a previous study, Pope, Meguerditchian,
115 Hopkins, & Fagot (2015) compared baboons’ and humans’ susceptibility to cognitive set
116 using a nonverbal ‘Learned Strategy-Direct Strategy’ (LS-DS) touch screen task. In this

117 task, subjects become familiar with a three-step (Square1 → Square2 → Triangle) learned
118 strategy (LS; see Figure 1a and Video S1) and are then presented with experimental trials
119 in which they can either use the LS (See Figure 1b) *or* a more direct strategy (DS or the
120 shortcut) by skipping the Square1 → Square2 sequence and going straight for the Triangle
121 (See Figure 1d and Video S2). The study found that, remarkably, all 15 baboon subjects
122 immediately switched to the DS when it became available, using it in 99% of trials. By
123 comparison, only 6.7% of humans used the DS in more than 50% of trials, demonstrating
124 that humans but not baboons were highly susceptible to cognitive set on the LS-DS task.

125 In the current study, we explored how non-human apes fit into this dialogue.
126 Indeed, many social learning studies have investigated non-human apes' abilities to
127 retrieve food from an apparatus by switching away from a familiar strategy, after
128 watching a demonstration of a more efficient alternative. A large number of these studies
129 find that non-human apes struggle to switch to the demonstrated strategy (Davis, 2017;
130 Davis, Vale, Schapiro, Lambeth, & Whiten, 2016; Gruber, Muller, Reynolds, Wrangham,
131 & Zuberbuhler, 2011; Hrubesch, Preuschoft, & van Schaik, 2009; Manrique & Call,
132 2015; Marshall-Pescini & Whiten, 2008; Price, Lambeth, Schapiro, & Whiten, 2009).
133 However, within the context of social learning, it is difficult to parse out perseverative
134 behaviors that arise from an inability to learn from a demonstration, from those that are
135 products of cognitive inflexibility. In fact, other studies report compelling evidence for
136 flexible strategy-use in apes, especially when the alternative solution requires relatively
137 simple modifications or reaps a better reward (Davis, Schapiro, Lambeth, Wood, &
138 Whiten, 2018; Harrison & Whiten, 2018; Jacobson & Hopper, 2019; Lehner et al., 2011;
139 Manrique, Völter, & Call, 2013; Van Leeuwen, Cronin, Schutte, Call, & Haun, 2013;

140 Yamamoto, Humle, & Tanaka, 2013). Thus, the extent to which non-human apes are
 141 capable of flexible strategy-use within an optional-switch context is also unclear.

142 The current study had two aims. First, we wanted to explore chimpanzees'
 143 susceptibility to cognitive set on the LS-DS task. Importantly, in the LS-DS task, the
 144 shortcut is a familiar component of the learned sequence; therefore, using it does not
 145 require learning new affordances or watching a demonstration. Second, we sought to
 146 contextualize these findings by comparing chimpanzees' response patterns on the LS-DS
 147 task to those of baboons and humans, previously reported in (Pope et al., 2015).



148

149 **Figure 1.** The LS-DS Task. a) BASE and b-d) PROBE trials depicting the Top left →
 150 Bottom right → Top right configuration. On PROBE trials, subjects may utilize any of
 151 three response strategies to be correct: b) the learned strategy (LS), c) the switch strategy

152 (SS), or d) the direct strategy (DS). Numbers indicate the correct Response 1, Response
153 2, and Response 3 for this configuration. Also see Video S1.

154

155

Methods

156 Chimpanzee Methods

157 **Subjects and General Procedure.** We tested 10 captive-born chimpanzee
158 subjects, five females ($M \pm SD = 30.9 \pm 8.0$ years, range: 21-40 years) and five males
159 ($M \pm SD = 25.4 \pm 2.7$ years, range: 23-30 years) located at the Yerkes National Primate
160 Research Center in Atlanta, Georgia, USA. The LS-DS was programmed using EPrime
161 (version 2.0, Psychology Software Tools, Pittsburgh, PA, USA) and administered using a
162 19-inch touch monitor (1939L Open-Frame Touchmonitor, Elo Touch Solutions,
163 Milpitas, CA, USA) affixed to a metal housing, which was temporarily attached to
164 subjects' home enclosure during testing. Each testing session lasted roughly 20 minutes.
165 Subjects' correct responses were rewarded with juice or small pieces of fruit or
166 vegetables, depending on their preferences or dietary restrictions. All testing was
167 approved by the Emory University IACUC.

168 **LS-DS Training Phase.** Chimpanzees completed four LS-DS training levels
169 (Table 1 & Video S2). During LS-DS training, subjects are presented with a
170 demonstration phase wherein two squares (Square1 and Square2) flash red in sequence,
171 and a response phase wherein they select the locations of Square1 and Square2 in the
172 demonstrated order. For each trial, the locations of Square1 and Square2 are randomly
173 assigned to a screen quadrant (e.g. top left, top right, bottom left, bottom right). In the
174 final training level, subjects first select Square1 and Square 2, then a triangle which

175 appears in one of the remaining quadrants. Incorrect responses resulted in a 3 second
 176 green delay screen. Accuracy was assessed after each block of 24 trials. To progress to
 177 the next training level, subjects were required to achieve >80% accuracy, twice. As levels
 178 progress, the time each square is demonstrated decreases and the number of response
 179 options increases (up to four). Note that Training 1.5 was a deviation from previous LS-
 180 DS training methods and consisted of trials identical to Training 1 (Training1.5a)
 181 randomly interspersed with trials in which four squares were shown but only one flashed
 182 red (Training1.5b). By the end of training, subjects were proficient in using the three step
 183 Square1→Square2→Triangle sequence (i.e. the learned strategy or **LS**).
 184

Table 1. *Learned Strategy–Direct Strategy Training Levels*

Level	Demonstration	Response	Required Responses
	Speed	Options	
Training 1	250ms/square	2 Squares	Square 1 + Square 2
Training 1.5a	250ms/square	2 Squares	Square 1 + Square 2
Training 1.5b	250ms/square	4 Squares	Square 1
Training 2	250ms/square	4 Squares	Square 1 + Square 2
Training 3	150ms/square	4 Squares	Square 1 + Square 2 + Triangle

185

186 We encountered several difficulties during training that required procedural
 187 adjustments. First, at the beginning of Training 3, five of the ten chimpanzee subjects
 188 would not select the triangle. In order to avoid their giving up on the task altogether, after
 189 several unsuccessful sessions we opted to briefly modify the program (see Video S2).

190 Thus, for five of the ten subjects, after they had correctly selected Square1 → Square2,
191 the triangle appeared in one of the two remaining places (like normal) but the other
192 response options disappeared, such that all that was left on the screen was the triangle.
193 Once subjects consistently touched the triangle (# of modified trials per subject = 36, 18,
194 41, 153, 21), they were switched back to the regular version of Training3 for all
195 remaining trials. Second, training duration varied from several months to more than a
196 year, depending on the subject, and at some points motivation appeared to dwindle
197 substantially. If this occurred, we enhanced their food reward or instated a 5 second (as
198 opposed to the normal 3 second) delay following incorrect responses. Third, for Training
199 3, even if subjects were excluding their previous selections, the likelihood of being
200 correct in any given trial simply by chance is 4.17% (Response 1 = 1/4, Response 2 =
201 1/3, and Response 3 = 1/2). Therefore, evidenced by considerably higher than chance
202 accuracy scores (in one case a subject had achieved above 75% accuracy nine times)
203 many subjects grasped the LS, yet failed to reach the >80% criterion twice. After several
204 months, we opted to adjust the Training 3 accuracy requirements such that a subject
205 needed to achieve either greater than 80% twice, greater than 75% three times, or greater
206 than 70% five times in order to progress to the experimental trials.

207 **LS-DS Experimental Phase.** Immediately after training, subjects completed 96
208 experimental trials, comprised of 48 PROBE and 48 BASE condition trials. In the
209 PROBE condition, the Triangle appears alongside the Square1 → Square2 demonstration
210 and remains visible on the response screen (see Fig. 1b-d). Thus, in PROBE trials,
211 subjects can then either continue to use their learned strategy,
212 Square1 → Square2 → Triangle (the LS) or they can simply ignore the demonstration and

213 select the Triangle (i.e. the DS or shortcut). Additionally, the task permits a third strategy
214 wherein subjects seemingly initiate the LS by selecting the first square but then switch to
215 the DS by skipping the second square and instead selecting the triangle
216 (Square1 → Triangle; Figure 1c). This ‘switch strategy’ (SS) was not reported in previous
217 studies, as it occurred so rarely (Pope et al., 2015). In the BASE (baseline) condition,
218 trials appear identical to Training 3 trials. However, if the subjects select the triangle’s
219 hidden location, they are scored and rewarded as if they had used the DS or the SS. This
220 was done to ensure that trends in shortcut use could not be attributed to accident.

221 In summary, subjects were extensively trained to reproduce the two-square
222 demonstration and then touch the triangle but when the triangle was already present
223 (PROBE trials) they could forego this learned strategy and directly select the triangle for
224 a more immediate reward.

225

226 **Comparative Methods**

227 A primary aim of this study was to compare the chimpanzees’ responses on the
228 LS-DS task to those of baboons and humans, which were previously collected and
229 reported in Pope et al. (2015).

230 **Subjects and General Procedure.** Baboon data were collected from 15 socially
231 housed baboons, nine females ($M \pm SD = 5.1 \pm 2.3$ years, range: 1.8-9.1 years) and six
232 males ($M \pm SD = 5.3 \pm 2.5$ years, range: 2.3-9.3 years) located at the CNRS Station de
233 Primatologie in Rousset-sur-Arc, France. The task was administered via ten automated
234 learning devices for monkeys (ALDMs; Fagot & Paleressompouille, 2009) and was
235 approved by the local Provence Alpes Cote d’Azur ethics committee for experimental

236 animal research. Human data were collected from 104 humans; 63 females ($M \pm SD =$
237 27.9 ± 17.6 years, range: 7-68 years), 40 males ($M \pm SD = 29.2 \pm 17.2$, range: 7-68 years)
238 and 1 participant who preferred to identify as neither male nor female (17 years) at a
239 temporary testing booth set up at Zoo Atlanta in Atlanta, Georgia, USA. Human testing
240 was approved by the Zoo Atlanta Research Committee and the Georgia State University
241 Institutional Review Board.

242 **LS-DS Training and Experimental Phases.** As reported in Pope et al. (2015),
243 baboons and humans completed Training levels 1, 2, and 3 (Table 1). The experimental
244 phase consisted of 720 trials for baboons (576 BASE and 144 PROBE, this is a 4:1 ratio),
245 96 trials for humans 11 years and older (48 BASE and 48 PROBE) and 48 trials for
246 humans under 11 (24 BASE and 24 PROBE).¹ As in the original analysis, we included
247 only the first 48 BASE and 48 PROBE trials for the baboon data in order to compare
248 initial strategy-use across species; however baboon strategy-use remained consistent
249 throughout the entire experimental phase.

250

251 **Analyses**

252 **General**

253 To measure differences between continuous variables such as response time or
254 number of trials, we used linear mixed-effects models. To measure differences in
255 accuracy error patterns, we used logistic mixed-effects models with a binomial error
256 structure. Subject ID was included as a random effect in all models. We used the *lme4*

¹ As before, we argue that the 4:1 ratio of BASE to PROBE trials that was received by baboons would only serve to enhance LS-use.

257 package (Bates, Mächler, Bolker, & Walker, 2015) in R 3.5.2 (R Core Team, 2019) to fit
258 the models, likelihood ratio tests using single-term deletions to assess the test predictors'
259 importance, and the *emmeans* package (Lenth, 2018) to compute pairwise contrasts, with
260 the Tukey multiple comparison correction.

261

262 **Training Phase**

263 **Training and Species.** To analyze differences in the number of training trials
264 between species, we calculated the number of trials each subject completed. Species and
265 training level (1, 2, or 3) were included as fixed effects.²

266 **Training Errors.** To analyze differences in accuracy between each step (Step 1:
267 Square 1, Step 2: Square 2, and Step 3: Triangle) of the response in Training 3, we
268 calculated conditional accuracies for each step. For example, if the participant correctly
269 selected Square 1, we calculated the proportion of trials in which they then correctly
270 selected Square 2. Species and step (Square 1, Square 2, or Triangle) were included as
271 fixed effects.

272 To analyze baboons' and chimpanzees' Training 2 errors, we isolated incorrect
273 Response 1 trials that were preceded by a correct trial. For each incorrect Response 1, we
274 determined whether the erroneous selection was a repetition of either Response 1 or
275 Response 2 of the preceding trial. For example, if the correct response to the preceding
276 trial was Bottom Left → Top Left and the erroneous selection was Bottom Left, this was
277 classified as a previous Response 1 error (Previous1); similarly, if the erroneous selection

² Chimpanzees' Training 1.5 trials were grouped with their Training 1 trials; however, the results were unchanged when they were combined with Training 2 or excluded.

278 was Top Left, this was classified as a previous Response 2 error (Previous2). We also
279 identified trials in which subjects selected the Square2 location first. For example, if the
280 demonstration depicted Bottom Left → Bottom Right and the erroneous Response 1 was
281 Bottom Right this would be classified as a reversal error (Reversal). If driven by chance,
282 we would expect errors to fall within these error types 33.3% of the time (one out of the
283 three erroneous response options). We compared the rates of Previous1, Previous2, and
284 Reversal error types for each species to chance ($\mu = .33$) using one-way t-tests.

285

286 **Experimental Phase**

287 **Strategy-Use Models.** For each trial, subjects' responses could fall into one of
288 four categories: incorrect, LS, SS, or DS. Thus, for each subject, the proportion of trials
289 in which each strategy was used is constrained by the proportion of trials in which each
290 other strategy was used. To best understand how strategy-use differed between subjects
291 and across species, we decided to use Bayesian multilevel multinomial modeling. This
292 allowed us to assess the probability of observing LS, SS, DS responses (excluding
293 incorrect trials) within the same model.³ We fit three models to these data. Each model
294 was fit using Hamiltonian Monte Carlo estimations, using the *Rstan* and *brms* packages
295 (Bürkner, 2017; Stan Development Team, 2018). For all models, we ran 3 chains of 4000
296 iterations to optimize model convergence, judged by the R-hat Gelman and Rubin
297 convergence diagnostic (McElreath, 2015).

³ We report frequentist statistics similar to our previous analyses (Pope, Fagot, Meguerditchian, Washburn, & Hopkins, 2018; Pope et al., 2015) alongside Figure S3.

298 Model 0 included only the random effect of subject ID. In addition to the random
299 effect of subject ID, Model 1 included the main effects of species (Baboon, Chimpanzee,
300 or Human), condition (BASE or PROBE), sex (Female⁴ or Male), age (continuous), and
301 training experience (continuous). To account for the substantial species difference in ages
302 sampled, age values were centered around the age of sexual maturity averaged across
303 males and females, for each species: baboons = 4.5, chimpanzees = 8.6, humans = 13.5
304 years (Figure S1; De Magalhaes & Costa, 2009). To account for the large species
305 differences in total number of training trials, this variable was z-transformed. Finally,
306 Model 2 included the random and fixed effects from Model 1, and the interaction of
307 species*condition. Models were compared using the Widely Applicable Information
308 Criteria (WAIC; McElreath, 2015).

309 **Strategy and Accuracy.** To assess the impact of strategy on performance, we
310 investigated average accuracies in BASE and PROBE trials. Recall that LS, SS, and DS
311 responses are each considered correct, only varying in their relative efficiencies. Species,
312 condition and the interaction of species*condition were included as fixed effects.

313 **Strategy and Trial Time.** To assess the impact of strategy on efficiency, we
314 investigated average total trial times for PROBE trials in which subjects used the DS, SS,
315 or LS. Due to an error in recording baboon response times, this analysis was only
316 possible for human and chimpanzee data. For each subject, response time outliers
317 (outside of 1.5 x the inter-quartile range of the first and third quartile) were excluded.
318 Species, condition and the interaction of species*condition were included as fixed effects.

⁴ The human participant that preferred to identify as neither male nor female was randomly included as female. However, excluding this participant did not change any results.

319 **Switch Costs.** We also analyzed costs associated with switching between the LS
320 and the DS. Baboon data were excluded from all switch cost analyses because they
321 received a different ratio of PROBE:BASE trials (4:1 instead of 1:1). Only subjects who
322 used the DS in greater than 50% of PROBE trials were included (humans: $N = 6$;
323 chimpanzees: $N = 7$). Trials that were preceded by an incorrect response were excluded
324 because it is not possible to assess which strategy was used in incorrect trials.

325 **Accuracy.** First, we compared subjects' Response 1 accuracies between trials in
326 which the condition repeated or was different. For example, a PROBE trial preceded by
327 another PROBE trial would be classified as condition = PROBE and trial type = *repeat*,
328 and a PROBE trial preceded by a BASE trial would be classified as condition = PROBE
329 and trial type = *different*. Species, condition, trial type and their interactions were
330 included as fixed effects.

331 **Response Time.** Second, we determined the first response time (RT1) for each
332 trial, by isolating the time between the fixation response and the first response. All trials
333 in which the first response was incorrect, or did not correspond to either LS or DS
334 strategies were excluded. Additionally, for each subject, response time outliers (outside
335 of 1.5 x the inter-quartile range of the first and third quartile) were excluded. For each
336 trial, subjects' strategy choice in relation to their previous strategy choice was
337 determined. For example, if a PROBE trial wherein the subject used the DS was preceded
338 by another PROBE trial wherein the subject had used the DS, this would be classified as
339 condition = PROBE and shift type = *stay*, and if a PROBE trial wherein the subject used
340 the DS was preceded by a BASE trial wherein the subject had used the LS, this would be
341 classified as condition = PROBE and shift type = *switch*.

342

343

Results

344 Training Phase

345 **Training and Species.** Chimpanzees' training required an of 295.8 days (range:
346 27-465 days). Across all training levels, chimpanzees required significantly more training
347 trials than baboons, which required significantly more training trials than humans [$\chi^2(2) =$
348 8.61, $p = 0.014$; Table 2].

349 **Training Errors.** To better understand the immense number of training trials that
350 chimpanzees and baboons needed to learn the LS, we analyzed for error patterns during
351 Training 2. We chose to analyze Training 2 errors because, unlike Training 1, subjects
352 selected between all four response options and even by Training 3, chimpanzees' and
353 baboons' (but not humans') Response 1 (chimpanzees: $M = 69.3\%$, baboons: 74.1%) and
354 Response 2 (chimpanzees: $M = 65.7\%$, baboons: 85.5%) accuracies were significantly
355 lower than Response 3 [(chimpanzees: $M = 93.5\%$, baboons: 88.7%); $\chi^2(4) = 2626.4$, $p <$
356 0.001; Figure S2]. Thus, chimpanzee and baboon subjects' difficulty learning the LS
357 involved their representation of the Square1 \rightarrow Square2 portion of the solution, not the use
358 of the Triangle.

359

360

361

362

363

364

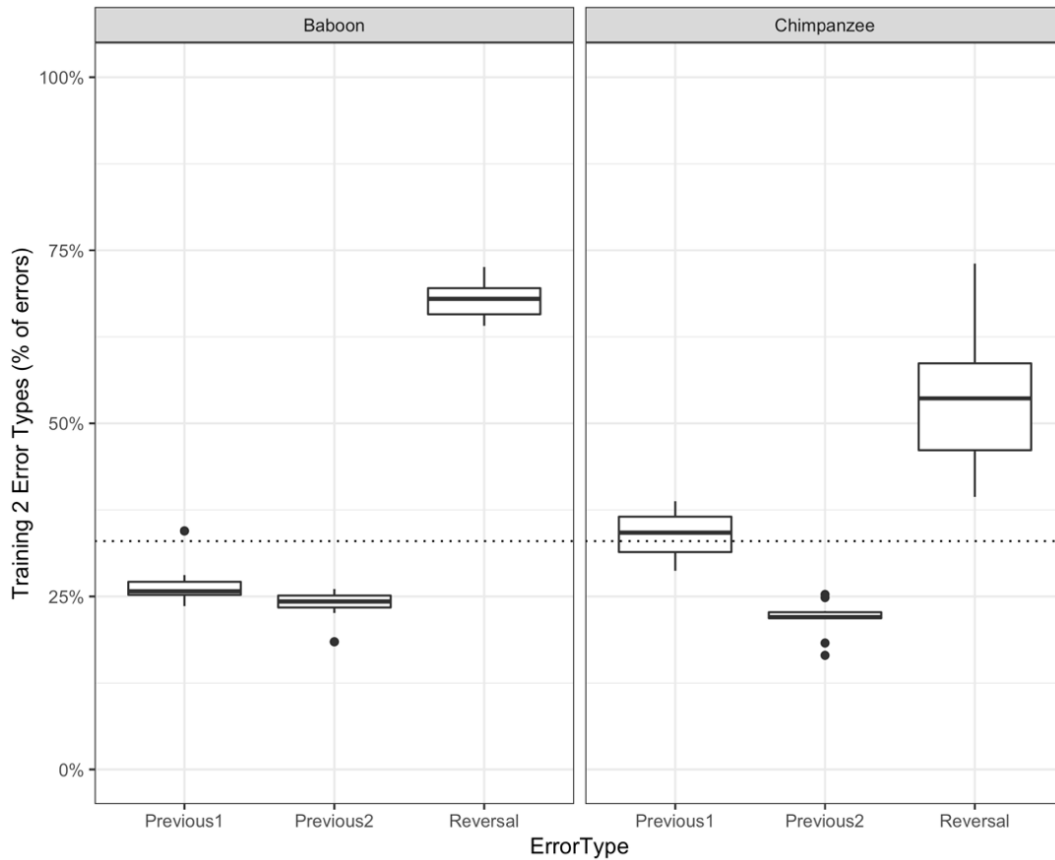
Table 2. Median, Interquartile Range (IQR) for the Number of Training Trials Required for All Species

	Baboons	Chimpanzees	Humans
Training 1	6084 (4794-6650)	817 (310-1362)	8 (8-16)
Training 1.5	-----	2343 (980-3099)	-----
Training 2	6108 (4167-7456)	8771 (8283-10276)	8 (8-8)
Training 3	2133 (2016-2294)	4852 (3260-11531)	8 (8-16)
Total	14235 (11767-16544)	17960 (14776-25284)	32 (24-48)

365

366 For chimpanzees, Previous1 errors did not significantly differ from chance ($M =$
367 33.9%; $t(9) = 0.80, p = .442, d = .25$), Previous2 errors occurred significantly less often
368 than expected by chance ($M = 21.8%; t(9) = -13.36, p < .001, d = 4.22$), and Reversal
369 errors occurred significantly more often than chance ($M = 53.6%; t(9) = 6.64, p < .001, d$
370 $= 2.10$; Figure 2b). For baboons, Previous1 errors occurred significantly less often than
371 expected by chance ($M = 26.5%; t(14) = -10.06, p < .001, d = 2.60$), Previous2 errors
372 occurred significantly less often than expected by chance ($M = 24.0%; t(14) = -18.63, p <$
373 $.001, d = 4.81$), and Reversal errors occurred significantly more often than chance ($M =$
374 $68.0%; t(14) = 52.40, p < .001, d = 13.53$; Figure 2). Only a small subset of humans ($N =$
375 38) committed errors of this type, and those comprised very few trials ($M \pm SD:$
376 Previous1 = 2.58 ± 4.67 trials, Previous2 = 2.14 ± 4.50 trials, Reversal = 4.51 ± 6.35
377 trials). Thus, human Training 2 errors were not eligible for further analysis.

378



379

380 **Figure 2.** The median number of trials for each error type in Training 2 (solid horizontal
 381 line), inter-quartile range (IQR; box), values within 1.5 x IQR (whiskers), and outliers
 382 (solid circles).

383

384 **Experimental Phase**

385 Table 3 reports the percentage of trials in which the Learned, Direct, and Switch
 386 strategies were used during experimental trials. Strikingly, 15/15 baboons, 6/10
 387 chimpanzees, and 5/104 humans used the DS the very first time it was available. Indeed,
 388 baboons used the DS for the first time on *Mdn* trial = 1 (*IQR*: 1-1, *range*: 1-1),
 389 chimpanzees on *Mdn* trial = 1 (*IQR*: 1-2, *range*: 1-5), and humans on *Mdn* trial = 7 (*IQR*:
 390 2-17.5, *range*: 1-44). In fact, 8/10 baboon and 2/10 chimpanzee subjects used the DS in

391 every single PROBE trial. In contrast, baboons used the SS for the first time on *Mdn* trial
 392 = 18 (*IQR*: 10.8-38%, *range*: 3-47%), chimpanzees on *Mdn* trial = 1.5 (*IQR*: 1-5.75%,
 393 *range*: 1-10%), and humans on *Mdn* trial = 14 (*IQR*: 8-22.2%, *range*: 3-46%).
 394

Table 3. Median, Interquartile Range (*IQR*), and Range Values for the Percentage of Trials in Which Learned, Direct, and Switch Strategies Were Used in BASE and PROBE Trials for All Species.

Note. LS = learned strategy; DS = direct strategy; SS = switch strategy. There was no difference in shortcut use between chimpanzees who received the brief Training 3 modification (N = 5; Mdn = 54.2%) and those who did not (N = 5; Mdn = 60.4%).

		Baboons		Chimpanzees		Humans	
BASE	LS	89.60%		66.70%		91.70%	
		<i>IQR</i> : 80.2-90.6	<i>range</i> : 68.8-97.9	<i>IQR</i> : 65.1-70.3	<i>range</i> : 56.2-77.1	<i>IQR</i> : 85.4-97.9	<i>range</i> : 52.1-100
	DS	0%		2.10%		0%	
		<i>IQR</i> : 0-0	<i>range</i> : 0-2.1	<i>IQR</i> : 0-2.1	<i>range</i> : 0-6.3	<i>IQR</i> : 0-0	<i>range</i> : 0-8.3
PROBE	SS	0%		7.30%		0%	
		<i>IQR</i> : 0-1.0	<i>range</i> : 0-4.2	<i>IQR</i> : 4.7-9.9	<i>range</i> : 2.1-12.5	<i>IQR</i> : 0-0	<i>range</i> : 0-2.1
	LS	0%		10.40%		89.60%	
		<i>IQR</i> : 0-0	<i>range</i> : 0-2.1	<i>IQR</i> : 0.5-16.1	<i>range</i> : 0-27	<i>IQR</i> : 79.2-95.8	<i>range</i> : 4.8-100
PROBE	DS	97.90%		57.30%		0%	
		<i>IQR</i> : 94.8-100	<i>range</i> : 89.6-100	<i>IQR</i> : 41.7-80.2	<i>range</i> : 22.9-100	<i>IQR</i> : 0-0	<i>range</i> : 0-85.4
	SS	0%		17.70%		0%	
	<i>IQR</i> : 0-2.1	<i>range</i> : 0-8.3	<i>IQR</i> : 7.8-28.6	<i>range</i> : 0-50	<i>IQR</i> : 0-2.1	<i>range</i> : 0-16.7	

395

396

397

398

399

400

401

402

403 **Strategy-Use Models.** Table 4 reports the parameter estimates for all models.
 404 Model 2 had the lowest WAIC, and a weight of 1, suggesting the best out-of-sample fit;
 405 we interpret only Model 2 in what follows.
 406

Table 4. $M \pm SE$ parameter estimates for Model 0, Model 1, and Model 2.

Note. LS = learned strategy; SS = switch strategy; DS = direct strategy; WAIC = widely applicable information criteria. Reference categories are LS (strategy), Baboon (species), BASE (condition), and Male (sex). Estimates whose 95% credible intervals did not include zero are bolded.

Effects	LS vs. SS	LS vs. DS
<i>Model 0; WAIC = 5523.74 ± 123.01</i>		
Intercept	-5.31 ± .29	-4.93 ± .43
<i>Model 1; WAIC = 3008.76 ± 147.50</i>		
Intercept	-4.77 ± .42	-3.37 ± .74
Species – Baboon vs. Chimpanzee	3.14 ± .53	-0.87 ± 1.10
Species – Baboon vs. Human	-2.33 ± .42	-8.11 ± .80
Condition – BASE vs. PROBE	3.05 ± .23	6.37 ± .23
Age	-0.04 ± .01	-0.01 ± .02
Total Training	0.09 ± .13	0.37 ± .26
Sex – Male vs. Female	-0.05 ± .27	0.37 ± .56
<i>Model 2; WAIC = 2795.38 ± 126.32</i>		
Intercept	-5.18 ± .55	-9.72 ± 1.90
Species – Baboon vs. Chimpanzee	3.59 ± .65	5.31 ± 2.11
Species – Baboon vs. Human	-2.58 ± .96	1.37 ± 1.85
Condition – BASE vs. PROBE	7.86 ± 1.45	18.04 ± 2.37
Age	-0.04 ± .01	-0.01 ± .02
Total Training	0.09 ± .13	0.37 ± .28
Sex – Male vs. Female	0.05 ± .27	0.53 ± .60
Species*Condition		
Baboon vs. Chimpanzee, BASE vs. PROBE	-4.99 ± 1.48	-11.70 ± 2.42
Baboon vs. Human, BASE vs. PROBE	-4.17 ± 1.67	-15.10 ± 2.37

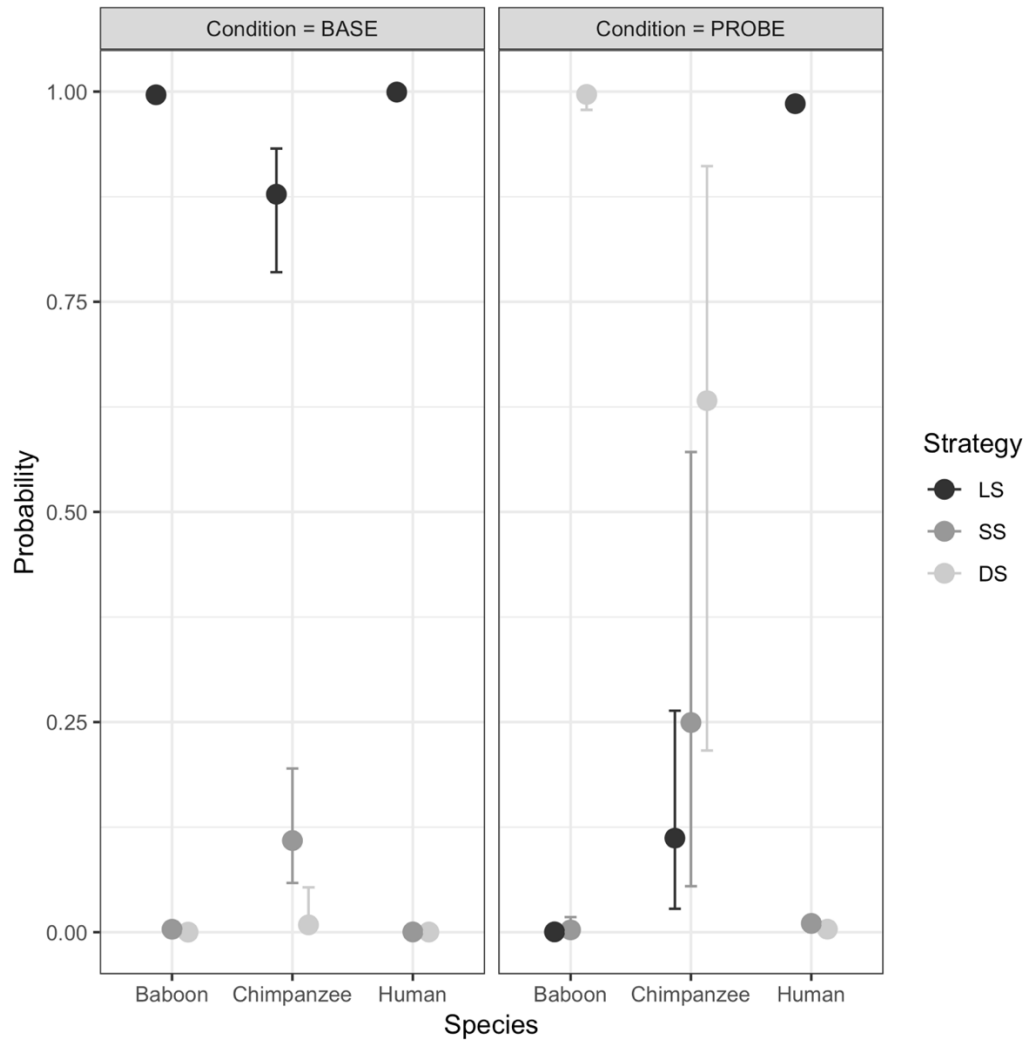
407

408 Number of training trials and sex were not strong predictors for using either the
 409 SS or DS. Age was a strong predictor of SS-use, but the effect was small; with every one-
 410 year increased in age, subjects were 1.04 (risk ratio = $1/e^{-.04}$) times less likely to select SS
 411 compared to the LS. Age was not a strong predictor for DS-use. We did not test for a
 412 three-way interaction between species, condition and age because of the large differences

413 in ages-sampled across species; however, our previous report noted an age-effect on
414 shortcut-use in humans wherein children under age 11 were more likely to be classified
415 as shortcut-users than adolescents (12-17) or adults (18+; Pope et al., 2015).

416 There was a strong effect of the interaction between species and condition on SS-
417 and DS-use. Figure 3 shows this interaction; in BASE trials, the probabilities of using the
418 SS and the DS were lower than LS-use for all species; however, chimpanzees exhibited
419 more SS-use compared with humans and baboons. In PROBE trials, baboons exhibited an
420 enhanced likelihood of using the DS, compared to both other species. Humans exhibited
421 an enhanced likelihood of using the LS, compared to both other species. However,
422 chimpanzees exhibited an inconsistent probability of using all three strategies. Indeed,
423 when considering *Mdn* strategy-use for each subject's PROBE trials, most baboons and
424 humans consistently selected either LS or DS strategies, whereas chimpanzees showed a
425 much more varied response pattern (Figure S3).

426



427

428 **Figure 3.** The relative probabilities of selecting the LS, SS, and DS in BASE and PROBE

429 conditions, across species. Error bars indicate 95% credible intervals.

430

431 **Strategy and Accuracy.** We found a significant interaction between species and432 condition on accuracy ($\chi^2(2) = 117.4, p < .001$). For baboons, PROBE ($M \pm SD = 98.2 \pm$ 433 2.1%) accuracy was significantly higher than BASE ($M \pm SD = 86.2 \pm 8.8%$) accuracy434 ($Z = -4.18, p < .001$). Similarly, for chimpanzees, PROBE ($M \pm SD = 90.4 \pm 10.2%$)435 accuracy was significantly higher than BASE ($M \pm SD = 77.1 \pm 5.0%$) accuracy ($Z = -$ 436 3.76, $p < .001$). But for humans, BASE ($M \pm SD = 90.6 \pm 9.8%$) accuracy was

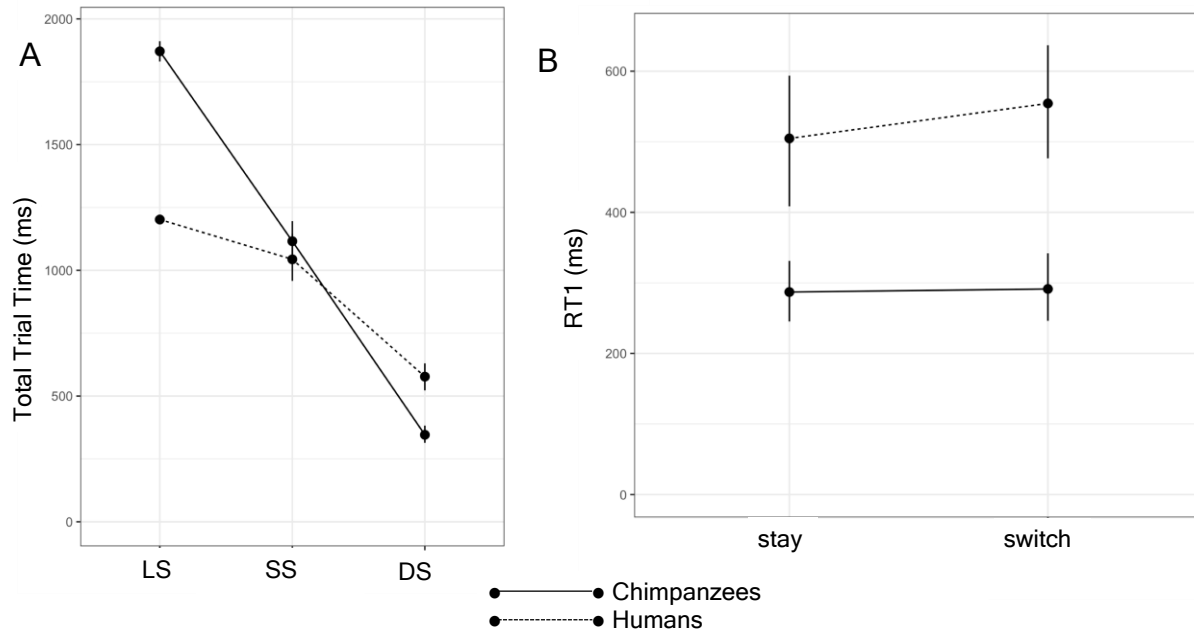
437 significantly higher than PROBE ($M \pm SD = 88.8 \pm 11.1\%$) accuracy ($Z = 2.78, p =$
438 $.005$), indicating that although they typically did not use the DS, the presence of the
439 triangle in PROBE trials may have been distracting for humans.

440 **Strategy and Total Trial Time.** We found a significant interaction effect of
441 species*strategy ($\chi^2(2) = 276.7, p < .001$; Figure 4A). For chimpanzees, DS ($M \pm SD =$
442 $346 \pm 304\text{ms}$) total trial times were significantly faster than SS ($M \pm SD = 1116 \pm 468\text{ms};$
443 $t(120.1) = 12.71, p < .001$) total trial times, which were significantly faster than LS ($M \pm$
444 $SD = 1871 \pm 414\text{ms}; t(129.0) = 12.75, p < .001$) total trial times. For humans, DS ($M \pm$
445 $SD = 577 \pm 420\text{ms}; t(112.3) = 18.06, p < .001$) and SS ($M \pm SD = 1044 \pm 355\text{ms};$
446 $t(871.6) = 2.38, p = .046$) total trial times were significantly faster than LS ($M \pm SD =$
447 $1202 \pm 398\text{ms}$) total trial times. Finally, humans' DS total trial times were significantly
448 slower than chimpanzees' ($t(112.3) = 6.89, p < .001$) and humans' LS total trial times
449 were significantly faster than chimpanzees' ($t(129.0) = -2.33, p = .021$).

450 **Switch Costs.**

451 **Accuracy.** We found no significant main or interaction effects of species ($\chi^2(1) =$
452 $2.68, p = 0.102$) or trial type (repeat and different; $\chi^2(1) = .215, p = 0.643$) on Response 1
453 accuracies.

454 **Response Time.** We found a significant interaction between species and shift type
455 (stay and switch; $\chi^2(4) = 3.85, p = 0.050$), wherein humans but not chimpanzees took
456 significantly longer to start responding when switching between LS and DS strategies ($t =$
457 $2.81, p = .012$; Figure 4B).



459 **Figure 4.** A) Total trial times by LS, SS, and DS. B) RT1(ms) by shift type.

460

461

461 Discussion

462 The current study found that chimpanzees exhibited an intermediate susceptibility
 463 to cognitive set on the LS-DS task. Their shortcut-use broadly resembled baboons', in
 464 that it was far greater than that of humans. However, as evidenced by their increased
 465 inter-individual variation in strategy selection as well as their use of the SS, chimpanzees'
 466 response style was still influenced by their learned strategy to some extent. These results
 467 contradict some existing reports of inflexibility in chimpanzees; however, we suggest that
 468 this is due to methodological differences.

469

470 Implications for Optional-Switch Cognitive Flexibility in Chimpanzees

471 First, a portion of the social learning studies reporting inflexibility in apes may have

472 inadvertently extinguished the alternative strategy during training by ‘locking’ it or
473 rendering it ineffective (Davis et al., 2018, 2016; Manrique & Call, 2015; Price et al.,
474 2009). This is important because if subjects attempt to use the alternative during training
475 (the reason it is locked) and are met with failure, subsequent avoidance of the alternative
476 solution could just be a product of their previously unrewarding experience with that
477 strategy (i.e. it did not work), resulting in their using it less than naïve controls. By
478 design, the LS-DS task is meant to promote shortcut use in PROBE trials. The DS is
479 highly salient and does not rely on working memory; once the triangle appears it remains
480 on the screen alongside three white squares, until it is selected (either before or after
481 Squares 1 and 2). Furthermore, the triangle is very familiar and highly associated with
482 reward; every time it is selected, throughout Training 3 and the experimental trials (both
483 BASE and PROBE), subjects receive a reward. Clearly, subjects of all species that used
484 the shortcut the first time it was available (baboons $N = 15/15$, chimpanzees $N = 6/10$,
485 humans $N = 5/104$) directly associated the triangle with the reward.⁵ Given the task
486 design, it is not surprising when subjects, of any species, prefer the DS. What is
487 remarkable is humans’ robust proclivity for the relatively inefficient LS.

488 Second, as mentioned previously, many social learning studies rely on a model to
489 demonstrate the more efficient alternative (Davis et al., 2018, 2016; Dean, Kendal,
490 Schapiro, Thierry, & Laland, 2012; Gruber et al., 2011; Hanus, Mendes, Tennie, & Call,
491 2011; Hrubesch et al., 2009; Marshall-Pescini & Whiten, 2008; Price et al., 2009),

⁵ As mentioned, half of the chimpanzee subjects avoided selecting the triangle when it first appeared in Training 3 (one subject even ran away from the touch screen the very first time it appeared), only selecting it when a modified version of the task removed all other options. This further suggests that chimpanzees’ flexibility on the LS-DS task was reliant upon their familiarity with the triangle. Davis et al. (2018) reported similarly enhanced behavioral flexibility when an alternative strategy for opening a puzzle box was simple and involved a familiar motor action.

492 introducing the possibility that perseveration may not be derived from cognitive
493 inflexibility but from a failure to extract the relevant information from (or even attend to)
494 the demonstration. In the LS-DS task, subjects are not shown how to enact the shortcut. It
495 is an already familiar step in their previously learned solution and is therefore not
496 confounded by social learning ability.

497 Finally, some tasks that describe conservative behavior in chimpanzees state that
498 the behavior is clearly within the species' repertoire (Marshall-Pescini & Whiten, 2008)
499 and thus, would be readily utilized if subjects were not influenced by a more familiar
500 solution. However, what may be present in the species' repertoire is not necessarily
501 present within the individual's and inexperience with specific manipulations, especially
502 those involving fine motor control or tool-use, should not be discounted (Dean et al.,
503 2012; Gruber et al., 2011). The time spent learning how to perform the seemingly more
504 efficient alternative may easily render it less so. By comparison, the LS, DS, and SS
505 strategies enlist motorically identical actions. Our findings demonstrate that, under
506 certain conditions, chimpanzees are capable of flexibly switching between abstract
507 response strategies. However, the extent to which this carries over into more naturalistic
508 contexts should be elucidated in future endeavors.

509

510 **Species Differences in Cognitive Set on the LS-DS Task**

511 Why are humans but not baboons nor (to a large extent) chimpanzees, affected by
512 cognitive set in the LS-DS task? Thousands more training trials were needed for
513 chimpanzees (range: 2,784-36,966) and baboons (range: 5,043-20,060) to learn the LS,
514 compared with humans (range: 24-152). We suggest that LS-use may be aided by

515 humans' enhanced ability to learn rules, either through verbal encoding (Ghirlanda, Lind,
516 & Enquist, 2017), heightened working memory (Fagot & De Lillo, 2011; but see Inoue &
517 Matsuzawa, 2007), or chunking strategies (Gobet et al., 2001; Kolodny, Edelman, &
518 Lotem, 2015). In fact, many of the human subjects were able to learn the rule after only 8
519 trials – a quarter of the total possible configurations – illustrating that, once learned,
520 humans are capable of accurately applying the LS to novel situations (Pope et al., 2015).
521 On the other hand, chimpanzees and baboons required many encounters with the same
522 trial configurations before they began to accurately respond. When examined,
523 chimpanzees' and baboons' errors seem to have been driven, to a large extent, by
524 erroneously reversing the order of the demonstrated squares (i.e. selecting Square2 first).
525 This reversal error is in line with recent findings which suggest that humans exhibit a
526 pronounced advantage when it comes to sequential encoding (Ghirlanda et al., 2017).
527 Indeed, a less firmly encoded LS could conceivably facilitate baboons' and chimpanzees'
528 ability to replace it.

529 It is also possible that increased DS-use might result from failure to inhibit the
530 triangle response, which during the final training level is most closely associated with
531 reward. Indeed, baboons' and chimpanzees' propensity for reversal errors during the
532 training phase seems to support this notion. However, if failure to inhibit were driving
533 shortcut-use in PROBE trials, we would expect similar inhibitory failures to trigger
534 erroneous BASE responses. Yet there were no significant differences between species'
535 Response 1 (Square 1) accuracies in BASE trials (Figure S4). Thus, although it is
536 plausible that baboons' and chimpanzees' training phase reversal errors were a function
537 of inhibitory failure, it does not appear to have driven species differences in strategy-use

538 during the experimental phase.

539 The current study found that chimpanzees, unlike humans (Pope et al., 2018), did
540 not exhibit switch costs on the LS-DS task. As far as we are aware, this is the first study
541 to investigate deficits associated with switching between abstract rules in chimpanzees.
542 Stoet & Snyder (2003) posited that switch costs are a product of firmly encoded rules,
543 such that greater cognitive effort is required to suppress the previous strategy and activate
544 the current. Additionally, less entrenched rules have been hypothesized to amplify
545 distractibility (Stoet & Snyder, 2008), which for the LS-DS task might further promote
546 the use of the shortcut.

547 That said, chimpanzees, but not baboons, utilized the SS, suggesting that
548 chimpanzees may have been influenced by their familiarity with the LS, at least to some
549 extent. We posit that differences in how baboons and chimpanzees process sequences
550 might have affected their relative abilities to separate the shortcut from the sequence as a
551 whole. When identifying components of a sequence, baboons have been shown to find
552 and select each component in turn, yet chimpanzees and humans appeared to identify the
553 entire sequence first, only then proceeding to make the appropriate selections (Kawai &
554 Matsuzawa, 200AD; Ohshiba, 1997). Chimpanzees may exhibit intermediate
555 susceptibility to cognitive set on the LS-DS task because, unlike humans, they are not
556 verbally encoding their solution strategy and, unlike baboons, they are processing the
557 sequence more holistically (however, see Beran, Pate, Washburn, & Rumbaugh, 2004).

558 Our findings are also in line with the overimitation literature which shows that,
559 compared to humans, chimpanzees are much better able to cut out irrelevant portions of a
560 demonstrated box-opening sequence (Horner & Whiten, 2005; McGuigan, Makinson, &

561 Whiten, 2011; McGuigan, Whiten, Flynn, & Horner, 2007). Future efforts aimed at
562 disentangling these influences would do well to compare cognitive set between monkeys
563 and apes using a non-sequential task.

564 We suggest that the distinctive response patterns observed between baboons,
565 chimpanzees, and human adults on the LS-DS task may be explained by a trade-off
566 between how quickly abstract rules can be learned and the degree to which they can be
567 flexibly used, a process potentially governed by differences in strategy-encoding. That
568 said, we consistently find that some humans both adopt the LS quickly *and* use the DS
569 (Pope et al., 2015). For example, we recently found population-level differences in
570 shortcut-use between Western humans (6% used the DS in more than 5% of trials) and
571 the seminomadic Himba (35% used the DS in more than 5% of trials; Pope et al., 2018)
572 suggesting that, at least within humans, propensity toward cognitive set on the LS-DS
573 task may vary by culture. Yet, although 100% of the tested chimpanzees and baboons
574 used the DS in more than 5% of trials, the highest proportion of humans to be classified
575 as DSers so far is 69% (Pope, Washburn, & Hopkins, submitted), which occurred after
576 subjects watched a video demonstrating the shortcut.⁶

577

578 **Concluding Remarks**

579 Adaptive behavior requires a balance between flexible and perseverative response
580 strategies. On one hand, if a familiar strategy is too concrete it may be unresponsive to
581 changing contextual cues, leading to an inefficient or even maladaptive response. On the

⁶ Note: 31% of human adults that watched a video of the shortcut and still did not use it in subsequent trials.

582 other hand, if strategies are too malleable, then responses may be crippled by irrelevant or
583 even erroneous information (Cools, 2008; Hommel & Colzato, 2017; Roberts, 2008).
584 Cognitive flexibility is a multifaceted construct that should be measured in both forced-
585 switch and optional-switch contexts. Although humans outperform non-human primates
586 in forced-switch contexts, the current study found that, within an optional-switch context,
587 chimpanzees exhibited an intermediary susceptibility to cognitive set compared with
588 baboons, who easily adopted the shortcut, and humans, who became stuck on the familiar
589 but relatively inefficient learned strategy. Future investigations into how the various
590 aspects of cognitive flexibility vary within individuals and between primate species must
591 necessarily include optional-switch metrics.

592

593 **ACKNOWLEDGMENTS**

594 Sincerest thanks to David Washburn for his thoughtful comments and guidance in
595 the compilation of this manuscript. Additionally, S. Pope is grateful for the funding
596 received in support of this project from the Kenneth W. and Georganna F. Honeycutt
597 Fellowship and Second Century Initiative Primate Social Cognition, Evolution, and
598 Behavior fellowship. W. Hopkins received support for this project from National
599 Institutes of Health grants NS-42867, NS-73134, HD-60563.

600

- 601 Adamson, R. E. (1952). Functional fixedness as related to problem solving: A repetition
602 of three experiments. *Journal of Experimental Psychology*, 44(4), 288–291.
603 Retrieved from internal-pdf://189.27.88.186/Adamson_1952.pdf
- 604 Aftanas, M. S., & Koppenaal, R. J. (1962). Effects of instructional problems and jar
605 position variation on the water-jar einstellung test. *Psychological Reports*, 10, 359–
606 362. Retrieved from internal-pdf://161.87.246.188/Aftanaas_1962.pdf internal-
607 pdf://3797244139/PsycINFO_Export.ris
- 608 Arrington, C. M., & Logan, G. D. (2004). The cost of voluntary task switch.
609 *Psychological Science*, 15(9), 610–615. Retrieved from internal-
610 pdf://113.184.42.236/Arrington_2004.pdf
- 611 Beran, M. J., Pate, J. L., Washburn, D. A., & Rumbaugh, D. M. (2004). Sequential
612 responding and planning in chimpanzees (*Pan troglodytes*) and rhesus macaques
613 (*Macaca mulatta*). *J Exp Psychol Anim Behav Process*, 30(3), 203–212.
614 <https://doi.org/10.1037/0097-7403.30.3.203>
- 615 Berg, E. A. (1948). A Simple Objective Technique for Measuring Flexibility in Thinking.
616 *The Journal of General Psychology*, 39(1), 15–22.
617 <https://doi.org/10.1080/00221309.1948.9918159>
- 618 Bilalić, M., Mcleod, P., & Gobet, F. (2008). Inflexibility of experts—Reality or myth?
619 Quantifying the Einstellung effect in chess masters. *Cognitive Psychology*, 56(2),
620 73–102.
- 621 Bonté, E., Flemming, T., & Fagot, J. (2011). Executive control of perceptual features and
622 abstract relations by baboons (*Papio papio*). *Behavioural Brain Research*, 222(1),
623 176–182. <https://doi.org/10.1016/j.bbr.2011.03.034>
- 624 Brass, M., Derrfuss, J., & von Cramon, D. Y. (2007). The role of the Posterior
625 frontolateral cortex in task-related control. In *Neuroscience of Rule-Guided*
626 *Behavior* (pp. 177–195). New York: Oxford University Press.
- 627 Brosnan, S. F., & Hopper, L. M. (2014). Psychological limits on animal innovation.
628 *Animal Behaviour*, 92, 325–332. <https://doi.org/10.1016/j.anbehav.2014.02.026>
- 629 Brown, V. J., & Tait, D. S. (2016). Attentional Set-Shifting Across Species BT -
630 Translational Neuropsychopharmacology. In T. W. Robbins & B. J. Sahakian (Eds.)
631 (pp. 363–395). Cham: Springer International Publishing.
632 https://doi.org/10.1007/7854_2015_5002
- 633 Bürkner, P.-C. (2017). An R Package for Bayesian Multilevel Models Using Stan.
634 *Journal of Statistical Software*, 80(1), 1–28.
- 635 Buttelmann, F., & Karbach, J. (2017). Development and Plasticity of Cognitive
636 Flexibility in Early and Middle Childhood. *Front Psychol*, 8, 1040.
637 <https://doi.org/10.3389/fpsyg.2017.01040>

- 638 Caselli, L., & Chelazzi, L. (2011). Does the macaque monkey provide a good model for
639 studying human executive control? A comparative behavioral study of task
640 switching. *PLoS One*, 6(6), e21489. <https://doi.org/10.1371/journal.pone.0021489>
- 641 Chrysikou, E. G., & Weisberg, R. W. (2005). Following the wrong footsteps: fixation
642 effects of pictorial examples in a design problem-solving task. *J Exp Psychol Learn*
643 *Mem Cogn*, 31(5), 1134–1148. <https://doi.org/10.1037/0278-7393.31.5.1134>
- 644 Cools, R. (2008). Dopaminergic Modulation of flexible cognitive control: the role of the
645 striatum. In S. A. Bunge & J. D. Wallis (Eds.), *Neuroscience of Rule-Guided*
646 *Behaviors* (pp. 313–334). New York: Oxford University Press.
- 647 Crooks, N. M., & McNeil, N. M. (2009). Increased practice with “set” problems hinders
648 performance on the water jar task. *Proceedings of the 31st Annual Conference of the*
649 *Cognitive Science Society, Cognitive*, 643–648. Retrieved from internal-
650 pdf://205.50.171.134/Crooks_2009.pdf
- 651 Davis, S. J. (2017). *The Context of Behavioural Flexibility in Chimpanzees (Pan*
652 *troglodytes): Implications for the Evolution of Cumulative Culture*. University of St.
653 Andrews.
- 654 Davis, S. J., Schapiro, S. J., Lambeth, S. P., Wood, L. A., & Whiten, A. (2018).
655 Behavioral conservatism is linked to complexity of behavior in chimpanzees (Pan
656 troglodytes): implications for cognition and cumulative culture. *Journal of*
657 *Comparative Psychology*.
- 658 Davis, S. J., Vale, G. L., Schapiro, S. J., Lambeth, S. P., & Whiten, A. (2016).
659 Foundations of cumulative culture in apes: improved foraging efficiency through
660 relinquishing and combining witnessed behaviours in chimpanzees (Pan
661 troglodytes). *Sci Rep*, 6, 35953. <https://doi.org/10.1038/srep35953>
- 662 De Magalhaes, J. P., & Costa, J. (2009). A database of vertebrate longevity records and
663 their relation to other life-history traits. *Journal of Evolutionary Biology*, 22(8),
664 1770–1774.
- 665 Deak, G. O., & Wiseheart, M. (2015). Cognitive flexibility in young children: General or
666 task-specific capacity? *J Exp Child Psychol*, 138, 31–53.
667 <https://doi.org/10.1016/j.jecp.2015.04.003>
- 668 Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012).
669 Identification of the social and cognitive processes underlying human cumulative
670 culture. *Science*, 335(6072), 1114–1118. <https://doi.org/10.1126/science.1213969>
- 671 Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the Dimensional Change Card
672 Sort: Implications for developmental theories and the measurement of executive
673 function in children. *Dev Rev*, 38, 241–268. <https://doi.org/10.1016/j.dr.2015.09.001>
- 674 Duncker, K., & Lees, L. S. (1945). On problem-solving. *Psychological Monographs*,

- 675 58(5), i-113. <https://doi.org/10.1037/h0093599>
- 676 Fagot, J., & De Lillo, C. (2011). A comparative study of working memory: immediate
677 serial spatial recall in baboons (*Papio papio*) and humans. *Neuropsychologia*,
678 49(14), 3870–3880. <https://doi.org/10.1016/j.neuropsychologia.2011.10.003>
- 679 Ghirlanda, S., Lind, J., & Enquist, M. (2017). Memory for stimulus sequences: a divide
680 between humans and other animals? *R Soc Open Sci*, 4(6), 161011.
681 <https://doi.org/10.1098/rsos.161011>
- 682 Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M.
683 (2001). Chunking mechanisms in human learning. *Trends Cogn Sci*, 5(6), 236–243.
684 Retrieved from internal-pdf://0034740865/Gobet_2001.pdf
- 685 Gruber, T., Muller, M. N., Reynolds, V., Wrangham, R., & Zuberbuhler, K. (2011).
686 Community-specific evaluation of tool affordances in wild chimpanzees. *Sci Rep*, 1,
687 128. <https://doi.org/10.1038/srep00128>
- 688 Hanus, D., Mendes, N., Tennie, C., & Call, J. (2011). Comparing the performances of
689 apes (*Gorilla gorilla*, *Pan troglodytes*, *Pongo pygmaeus*) and human children (*Homo*
690 *sapiens*) in the floating peanut task. *PLoS One*, 6(6), e19555.
691 <https://doi.org/10.1371/journal.pone.0019555>
- 692 Harlow, H. (1949). The formation of learning sets. *Psychological Review*, 56(1), 51–65.
693 Retrieved from internal-pdf://145.144.205.228/Harlow_1949.pdf
- 694 Harrison, R. A., & Whiten, A. (2018). Chimpanzees (*Pan troglodytes*) display limited
695 behavioural flexibility when faced with a changing foraging task requiring tool use.
696 *PeerJ*, 6, e4366. <https://doi.org/10.7717/peerj.4366>
- 697 Hommel, B., & Colzato, L. S. (2017). The social transmission of metacontrol policies:
698 Mechanisms underlying the interpersonal transfer of persistence and flexibility.
699 *Neurosci Biobehav Rev*. <https://doi.org/10.1016/j.neubiorev.2017.01.009>
- 700 Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching
701 in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Anim Cogn*, 8(3),
702 164–181. <https://doi.org/10.1007/s10071-004-0239-6>
- 703 Hrubesch, C., Preuschoft, S., & van Schaik, C. (2009). Skill mastery inhibits adoption of
704 observed alternative solutions among chimpanzees (*Pan troglodytes*). *Anim Cogn*,
705 12(2), 209–216. <https://doi.org/10.1007/s10071-008-0183-y>
- 706 Huguet, P., Barbet, I., Belletier, C., Monteil, J. M., & Fagot, J. (2014). Cognitive control
707 under social influence in baboons. *J Exp Psychol Gen*, 143(6), 2067–2073.
708 <https://doi.org/10.1037/xge0000026>
- 709 Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Curr*
710 *Biol*, 17(23), R1004-5. <https://doi.org/10.1016/j.cub.2007.10.027>

- 711 Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in*
712 *Psychology*, 30(2), 190–200. <https://doi.org/10.1016/j.newideapsych.2011.11.001>
- 713 Ionescu, T. (2017). The variability-stability-flexibility pattern: A possible key to
714 understanding the flexibility of the human mind. *Review of General Psychology*,
715 21(2), 123–131. <https://doi.org/10.1037/gpr0000110>
- 716 Jacobson, S. L., & Hopper, L. M. (2019). Hardly habitual: chimpanzees and gorillas
717 show flexibility in their motor responses when presented with a causally-clear task.
718 *PeerJ*, 7, e6195. <https://doi.org/10.7717/peerj.6195>
- 719 Kawai, N., & Matsuzawa, T. (200AD). Cognition: numerical memory span in a
720 chimpanzee. *Nature*, 403(6765), 39–40. Retrieved from internal-
721 pdf://189.27.88.170/Kawai_2000.pdf
- 722 Kolodny, O., Edelman, S., & Lotem, A. (2015). Evolved to adapt: A computational
723 approach to animal innovation and creativity. *Current Zoology*, 61(2), 350–367.
724 Retrieved from internal-pdf://189.27.88.172/Kolodny_2014.pdf
- 725 Lacreuse, A., Parr, L., Chennareddi, L., & Herndon, J. G. (2018). Age-related decline in
726 cognitive flexibility in female chimpanzees. *Neurobiology of Aging*, 72(August),
727 83–88. <https://doi.org/10.1016/j.neurobiolaging.2018.08.018>
- 728 Lehner, S. R., Burkart, J. M., & Schaik, C. P. (2011). Can captive orangutans (*Pongo*
729 *pygmaeus abelii*) be coaxed into cumulative build-up of techniques? *J Comp*
730 *Psychol*, 125(4), 446–455. <https://doi.org/10.1037/a0024413>
- 731 Lemaire, P., Luwel, K., & Brun, F. (2017). Does the Number of Available Strategies
732 Change How Children Perform Cognitive Tasks? Insights from Arithmetic. *Journal*
733 *of Educational and Developmental Psychology*, 7(2).
734 <https://doi.org/10.5539/jedp.v7n2p43>
- 735 Luchins, A. S. (1942). Mechanization of problem solving: The effect of Einstellung.
736 *Psychological Monographs*, 54(6), 1–95. Retrieved from internal-
737 pdf://112.204.51.47/Luchins_1942.pdf internal-
738 pdf://3797244140/162756_Janitzky_Optogenetic_silencing_of_locus.enw
- 739 Luchins, A. S., & Luchins, E. H. (1950). New experimental attempts at preventing
740 mechanization in problem solving. *The Journal of General Psychology*, 42, 279–
741 297. Retrieved from internal-pdf://6.199.226.106/LuchinsLuchins_1950.pdf
- 742 Luwel, K., Schillemans, V., Onghena, P., & Verschaffel, L. (2009). Does switching
743 between strategies within the same task involve a cost? *Br J Psychol*, 100(Pt 4),
744 753–771. <https://doi.org/10.1348/000712609X402801>
- 745 Manrique, H. M., & Call, J. (2015). Age-dependent cognitive inflexibility in great apes.
746 *Animal Behaviour*, 102, 1–6. Retrieved from internal-
747 pdf://59.214.21.234/Manrique_2015.pdf

- 748 Manrique, H. M., Völter, C. J., & Call, J. (2013). Repeated innovation in great apes.
749 *Animal Behaviour*, 85(1), 195–202. <https://doi.org/10.1016/j.anbehav.2012.10.026>
- 750 Marshall-Pescini, S., & Whiten, A. (2008). Chimpanzees (*Pan troglodytes*) and the
751 question of cumulative culture: an experimental approach. *Anim Cogn*, 11(3), 449–
752 456. <https://doi.org/10.1007/s10071-007-0135-y>
- 753 McElreath, R. (2015). *Statistical rethinking: a Bayesian course with examples in R and*
754 *Stan*. London: Chapman and Hall/CRC.
- 755 McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-
756 copying: adults imitate causally irrelevant aspects of tool use with higher fidelity
757 than young children. *Br J Psychol*, 102(1), 1–18.
758 <https://doi.org/10.1348/000712610X493115>
- 759 McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque
760 versus causally transparent tool use by 3- and 5-year-old children. *Cognitive*
761 *Development*, 22(3), 353–364. <https://doi.org/10.1016/j.cogdev.2007.01.001>
- 762 Meiran, N. (1996). Reconfiguration of processing mode prior to task performance.
763 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6),
764 1423. Retrieved from internal-pdf://0309198986/Meiran_1996.pdf
- 765 Moore, T. L., Killiany, R. J., Herndon, J. G., Rosene, D. L., & Moss, M. B. (2005). A
766 non-human primate test of abstraction and set shifting: An automated adaptation of
767 the Wisconsin Card Sorting Test. *Journal of Neuroscience Methods*, 146(2), 165–
768 173. <https://doi.org/10.1016/j.jneumeth.2005.02.005>
- 769 Ohshiba, N. (1997). Memorization of serial items by japanese monkeys, a chimpanzee,
770 and humans. *Japanese Psychological Research*, 39(3), 236–252. Retrieved from
771 internal-pdf://2.45.22.121/Ohshiba_1997.pdf
- 772 Pope, S. M., Fagot, J., Meguerditchian, A., Washburn, D. A., & Hopkins, W. D. (2018).
773 Enhanced Cognitive Flexibility in the Seminomadic Himba. *Journal of Cross-*
774 *Cultural Psychology*, 0022022118806581.
775 <https://doi.org/10.1177/0022022118806581>
- 776 Pope, S. M., Meguerditchian, A., Hopkins, W. D., & Fagot, J. (2015). Baboons (*Papio*
777 *papio*), but not humans, break cognitive set in a visuomotor task. *Anim Cogn*, 18(6),
778 1339–1346. <https://doi.org/10.1007/s10071-015-0904-y>
- 779 Pope, S. M., Washburn, D. A., & Hopkins, W. D. (n.d.). *Breaking Cognitive Set: Seeing*
780 *an Alternative Strategy is Not the Same as Looking for One*. Submitted.
- 781 Price, E. E., Lambeth, S. P., Schapiro, S. J., & Whiten, A. (2009). A potent effect of
782 observational learning on chimpanzee tool construction. *Proc Biol Sci*, 276(1671),
783 3377–3383. <https://doi.org/10.1098/rspb.2009.0640>

- 784 Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card
785 sorting test: A meta-analytic review. *Psychology and Aging, 19*(3), 482–494.
786 <https://doi.org/10.1037/0882-7974.19.3.482>
- 787 Roberts, A. C. (2008). Dopaminergic and Serotonergic modulation of two distinct forms
788 of flexible cognitive control: Attentional set-shifting and Reversal learning. In S. A.
789 Bunge & J. D. Wallis (Eds.), *Neuroscience of Rule-Guided Behavior* (pp. 283–312).
790 New York: Oxford University Press.
- 791 Roberts, A. C., Robbins, T. W., & Everitt, B. J. (1988). The effects of intradimensional
792 and extradimensional shifts on visual discrimination learning in humans and non-
793 human primates. *The Quarterly Journal of Experimental Psychology, 40*(4), 321–
794 341.
- 795 Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between cognitive
796 tasks. *Journal of Experimental Psychology: General, 124*(2), 207. Retrieved from
797 internal-pdf://74.149.84.253/Rogers_1995.pdf
- 798 Rumbaugh, D. M., Savage-Rumbaugh, S., & Washburn, D. A. (1996). Toward a new
799 outlook on primate learning and behavior: complex learning and emergent processes
800 in comparative perspective. *Japanese Psychological Research, 38*(3), 113–125.
801 Retrieved from internal-pdf://196.11.173.245/Rumbaugh_1996.pdf
- 802 Ruscio, A. M., & Amabile, T. M. (1999). Effects of instructional style on problem-
803 solving creativity. *Creativity Research Journal, 12*(4), 251–266. Retrieved from
804 internal-pdf://83.159.127.18/Ruscio_1999.pdf
- 805 Stan Development Team. (2018). RStan: the R interface to Stan. R package.
- 806 Stoet, G., & Snyder, J. (2008). Task-Switching in Human and Nonhuman Primates:
807 Understanding rule encoding and control from behavior to single neurons. In S. A.
808 Bunge & J. D. Wallis (Eds.), *The Neuroscience of Rule-Guided Behavior* (pp. 227–
809 253). New York, New York: Oxford University Press.
- 810 Stoet, G., & Snyder, L. H. (2003). Executive control and task-switching in monkeys.
811 *Neuropsychologia, 41*(10), 1357–1364. [https://doi.org/10.1016/s0028-](https://doi.org/10.1016/s0028-3932(03)00048-4)
812 [3932\(03\)00048-4](https://doi.org/10.1016/s0028-3932(03)00048-4)
- 813 Sullivan, E. V., Mathalon, D. H., Zipursky, R. B., Kersteen-Tucker, Z., Knight, R. T., &
814 Pfefferbaum, A. (1993). Factors of the Wisconsin Card Sorting Test as measures of
815 frontal-lobe function in schizophrenia and in chronic alcoholism. *Psychiatry*
816 *Research, 46*(2), 175–199.
- 817 Sweller, J., Mawer, R. F., & Howe, W. (1982). Consequences of history-cued and means-
818 end strategies in problem solving. *The American Journal of Psychology, 95*(3), 455–
819 483. Retrieved from internal-pdf://101.210.161.55/Sweller_1982b.pdf
- 820 Van Leeuwen, E. J., Cronin, K. A., Schutte, S., Call, J., & Haun, D. B. (2013).

- 821 Chimpanzees (*Pan troglodytes*) flexibly adjust their behaviour in order to maximize
822 payoffs, not to conform to majorities. *PLoS One*, 8(11), e80945.
823 <https://doi.org/10.1371/journal.pone.0080945>
- 824 Yamamoto, S., Humle, T., & Tanaka, M. (2013). Basis for cumulative cultural evolution
825 in chimpanzees: social learning of a more efficient tool-use technique. *PLoS One*,
826 8(1), e55768. <https://doi.org/10.1371/journal.pone.0055768>
- 827 Zelazo, P. D. (2008). The Development of Rule Use in Children. In S. A. Bunge & J. D.
828 Wallis (Eds.), *Neuroscience of Rule-Guided Behavior* (pp. 441–455). New York:
829 Oxford University Press.
- 830 Zelazo, P. D., Muller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ... Carlson,
831 S. (2003). The development of executive function in early childhood. *Monographs*
832 *of the Society for Research in Child Development*, i-151. Retrieved from internal-
833 pdf://0937558954/Zelazo_2003.pdf
- 834