

The origins of cognitive flexibility in chimpanzees

Averill Cantwell^{1*}, Joshua Buckholtz³, Rebeca Atencia⁴, & Alexandra G Rosati^{1,2*}

¹Department of Psychology

²Department of Anthropology
University of Michigan
Ann Arbor, MI USA

³Department of Psychology
Harvard University

⁴Jane Goodall Institute Congo
Pointe Noire, Republic of Congo

*averill@umich.edu or *rosati@umich.edu

Conflict of Interest Statement:

The authors have no conflict of interest to declare.

Data Availability Statement:

All data from this study will be made available in Dryad Digital Repository upon publication.

Acknowledgements:

We thank Megan Mulhinch and Lauren DiNicola for their assistance with data coding, and Consulting for Statistics, Computing and Analytics Research (CSCAR) at the University of Michigan for statistical advice. We also thank the Congolese Ministry of Research and the Jane Goodall Institute for support of this work. At the Tchimpounga Chimpanzee Sanctuary, we thank Sofia Fernandez Navarro and the chimpanzee caretakers and staff for their assistance and support. This work was supported by a fellowship to AC supported by NICHD Training Grant 5T32HD007109-40; and NSF grants #1926653 and #1944881, NIH grant R01AG049395, and Sloan Foundation Fellowship #FG-2019-12054 to AGR.

Research Highlights:

- Comparisons of nonhuman primate cognitive development can provide insights into the evolutionary origins of human cognition, including both similarities and differences
- We examined chimpanzee cognitive flexibility across multiple contexts using reversal learning tasks, testing the importance of spatial versus perceptual cues and fixed versus probabilistic payoffs
- We found an extended developmental trajectory for shifting abilities in chimpanzees, with abilities developing into adulthood; additionally, female chimpanzees shifted responses more quickly than males
- Chimpanzees also showed some constraints, preferentially using spatial rather than perceptual cues to solve problems, and perseverating at higher rates when response feedback was probabilistic

Abstract

Cognitive flexibility is a core component of executive function, a suite of cognitive capacities that enables individuals to update their behavior in dynamic environments. Human executive functions are proposed to be enhanced compared to other species, but this inference is based primarily on neuroanatomical studies. To address this, we examined the nature and origins of a core component of executive function—cognitive flexibility—in chimpanzees, our closest living relatives. Across three studies, we examined different components of cognitive flexibility using reversal learning tasks where individuals first learned one contingency and then had to shift responses when contingencies flipped. In Study 1, we tested $n=82$ chimpanzees ranging from juvenility to adulthood on a spatial reversal task, to characterize the development of basic shifting skills. In Study 2, we tested how $n=24$ chimpanzees use spatial versus arbitrary perceptual information to shift, a proposed difference between human and nonhuman cognition. In Study 3, we tested $n=40$ chimpanzees on a probabilistic reversal task. We found an extended developmental trajectory for basic shifting and shifting in response to probabilistic feedback—chimpanzees did not reach mature performance until late in

ontogeny. Additionally, females were faster to shift than males were. We also found that chimpanzees were much more successful when using spatial versus perceptual cues, and highly perseverative when faced with probabilistic versus consistent outcomes. These results identify both core features of chimpanzee cognitive flexibility that are shared with humans, as well as constraints on chimpanzee cognitive flexibility that may represent evolutionary changes in human cognitive development.

Keywords: Comparative development, primates, cognitive control, cognitive evolution

Introduction

The ability to flexibly adapt to changes in the environment—a key feature of human cognition—is enabled by a suite of cognitive processes collectively referred to as executive function. Cognitive flexibility is a core executive function that allows individuals to adjust behavior in response to changes in the environment, such that it aligns with current goals. This can include shifting between different responses when faced with changes in feedback (e.g., reversal learning), shifting between different rule sets within the same task (e.g., set shifting), or shifting between distinct cognitive operations (e.g., task switching). In humans, executive functions develop slowly throughout our long childhood and adolescence, and are linked to important developmental outcomes including mature theory of mind (Hughes, 1998; Hughes & Ensor, 2007; Kloo et al., 2020), mathematical reasoning capacities (Bull & Scerif, 2001), and educational attainment (Willoughby et al., 2012). Executive functions generally, and cognitive flexibility specifically, are also important mechanisms supporting novel problem-solving, creativity, and fluid intelligence (Decker et al., 2007; Kafadar & Orhan, 2015), highlighting how these capacities are central to our conception of ‘intelligent’ behavior in humans.

Enhanced executive function is also thought to be a key evolutionary change in human cognition (Laland & Seed, 2021; Rosati, 2017; Sherwood et al., 2008; Smaers et al., 2017). Evidence in support of this idea comes from comparative anatomical studies of the prefrontal cortex, a neurobiological locus for executive functions (Miller, 2000; Miller & Cohen, 2001). The prefrontal cortex has undergone a number of evolutionary changes in humans: the human prefrontal cortex is absolutely larger and exhibits differences in organization, connectivity, and development relative to other primates, including great apes (Rilling & Insel, 1999; Schenker et al., 2005; Semendeferi et al., 2002; 2011; Sherwood et al., 2008; Smaers et al., 2017). For example, human prefrontal cortex development exhibits a protracted period of myelination compared to chimpanzees (Miller et al., 2012), and a greater degree of gyrification than expected for a primate of our size (Rilling & Insel, 1999). Given the link between the prefrontal cortex and executive function, some have hypothesized that these neuroanatomical changes drove a concurrent enhancement in humans' capacity for cognitive flexibility (Sherwood et al., 2008; Smaers et al., 2017; Teffer & Semendeferi, 2012). Inferences from comparative neuroanatomy, however, are a proxy for understanding specific cognitive traits. Strong inferences about changes in human cognition require direct evidence about differences in cognition and behavior between species (Healy & Rowe, 2006; Logan et al., 2018).

Direct comparison of executive functions between humans and other primates has yielded mixed results. In part, this may be because this approach has typically compared young human children to mixed-age samples of animals. For example, in a study comparing the performance of children and the four other great ape species in an inhibitory control task, apes showed near-ceiling levels of success similar to children (Barth & Call, 2006). Young children and apes also performed similarly on a detour-reaching task, which requires the participant to inhibit a reaching response directly towards a visible reward in favor of an

indirect path (Amici et al., 2008; Vlamings et al., 2010). In tests of working memory, another core component of executive function, chimpanzees do not appear to differ significantly from school-aged children (Völter et al., 2019), or in some cases even adults (Inoue & Matsuzawa, 2007; but see Cook & Wilson, 2010). Yet, in other contexts children do outperform apes, such as when demands on planning or flexibility increase (Dunbar et al., 2005; Herrmann et al., 2015). For example, chimpanzees and 3-year-old children were less successful than 6-year-old children in a trap task, where they first learned to pull a food reward directly towards them, and then had to shift their response after a hole opened along the path they previously used (Herrmann et al., 2015). In addition, there may be important differences in the cognitive strategies that animals and humans use to solve a given task. For example, children may act more efficiently even when animals are also successful (Völter & Call, 2014; Völter et al., 2019). These results highlight the importance of assessing performance across different contexts, as well as the need to include nonhumans at comparable ages and life history stages as in human studies (Gómez, 2005; Matsuzawa, 2007; Rosati et al., 2014).

Here we tested the roots of human cognitive flexibility through comparative studies of cognitive development in our closest living relative, chimpanzees. Chimpanzees are an important species for contextualizing human cognition due to their close phylogenetic relationship, complex behavior, and relatively slow development. First, chimpanzees exhibit a suite of complex technical and social behaviors that may be underpinned by executive functions. Wild chimpanzees routinely use and manufacture various types of tools, including probes, sponges, and hammers (Goodall, 1964; Nishida, 1968; Sanz & Morgan, 2010)—skills that can require years of experience to perform competently (Biro et al., 2003; Lonsdorf, 2005, 2006). Chimpanzees also exhibit a fission-fusion social system, in which community members split into parties of fluctuating membership over time, which is thought to demand complex cognitive abilities to track a dynamic social landscape (Amici et al., 2008; Aureli et

al., 2008). Additionally, chimpanzees have a relatively slow life history pattern with a long juvenile period, more like that seen in humans (Emery Thompson et al., 2007; Nishida et al., 2003; Pusey, 1990; Walker et al., 2018). A long juvenile period has been proposed to facilitate the emergence of complex cognition and behavior as it allows for an extended period to develop complex skills (Bjorklund & Green, 1992; Bruner, 1972; Kaplan et al., 2000). Indeed, mature executive functions in humans emerge over this long developmental period (Anderson, 2002; Diamond, 2013; Zelazo et al., 2003). Chimpanzees often outperform monkey and lemur species lacking these complex ecological, social, and life history features on executive function tasks (Amici et al., 2008; Deaner et al., 2006; MacLean et al., 2014; Rumbaugh, 1997). However, the extent to which these features shared by humans and chimpanzees translate into shared patterns of executive function abilities is unclear. Finally, studies of cognitive development in nonhumans can disentangle the influence of processes like language and schooling that shape human executive function development (Kuhn et al., 2014; McCrea et al., 1999). In particular, studies of animals provide a complimentary test of developmental change in the absence of these human-specific features. Examining the development of executive function in chimpanzees therefore provides important comparative evidence for understanding the biological bases of human development.

In this set of studies, our first goal was to characterize the developmental trajectory of cognitive flexibility across a large sample of chimpanzees ranging from juvenility to adulthood. Prior comparative work has focused on problem-solving contexts that likely tap into cognitive flexibility, but also a number of other cognitive abilities (e.g., Hare et al., 2001; Horner and Whiten, 2005; Seed et al., 2009). Here, we implemented a reversal learning paradigm to focus on core capacities for shifting. This paradigm is widely used in studies of both human and nonhuman cognition and examines responses to changing stimulus-reward contingencies—a key measure of cognitive flexibility across species (see Izquierdo et al.,

2017 for a review). Apes have outperformed other primate species in reversal learning tasks, possibly by using rule-based rather than simpler associative learning strategies (Rumbaugh & Gill, 1973; reviewed in Rumbaugh, 1997). Yet, there is little data on how this skill develops in apes, and some work suggests that young chimpanzees exhibit little change in performance across early development (Wobber et al., 2010). Chimpanzees do show some human-like declines in cognitive flexibility during aging (Lacreuse et al., 2018; Manrique & Call, 2015). However, to date, no study has tracked the developmental trajectory of chimpanzee reversal learning from juvenility to adulthood.

Our second goal was to test whether chimpanzees can flexibly shift responses using different information types. Adult humans are proficient at using many different types of information to guide flexible responses (Laland & Seed, 2021; Penn et al., 2008), but children and animals show some constraints in the kinds of information they use to solve problems. In particular, great apes and very young children appear to prefer to solve problems using spatial information, and struggle to use perceptual cues that are not directly relevant to the task. For example, when searching for a hidden reward based either on the location where they had previously seen it placed, or the visual features of the container, all great ape species and 1-year-old children preferentially choose the same location rather than the same container (Haun et al., 2006). In contrast, older children preferentially choose the same container. Chimpanzees also performed better in a working memory task when they were able to use spatial information versus feature-only cues (Völter et al., 2019). This aligns with a broader line of work suggesting that acquisition of language and cultural information allows children to integrate spatial and perceptual information in increasingly flexible ways (Haun et al., 2006; Hermer-Vazquez et al., 1999, 2001). Animals do routinely learn associations between arbitrary perceptual features and outcomes with enough experience (e.g., Rumbaugh, 1997), yet it appears that many animals will preferentially use spatial or causal information to solve problems, and may struggle to use more arbitrary perceptual features in similar contexts without extensive exposure (Penn et al., 2008; Seed et al., 2011). However, there has not been a direct comparison of how information type enables or constrains cognitive flexibility in nonhumans.

Our final goal was to assess how chimpanzees respond to a more complex task involving probabilistic feedback. In the real world, response-outcome contingencies are rarely fully known or consistent. Rather, individuals must flexibly respond to stochastic situations. As such, probabilistic reversal learning tasks—where both the ‘correct’ and ‘incorrect’ responses provide variable feedback—are commonly used to assess cognitive flexibility in human adults (Cools et al., 2002; Izquierdo et al., 2017). While reversal tasks with fixed outcomes can be navigated using a win-stay, lose-shift strategy, probabilistic tasks require maintaining a particular choice strategy despite some negative feedback (the “good” option will still result in losses sometimes), and then shifting responses as evidence accumulates that the previous response is no longer optimal. This typically results in perseverative errors, where subjects continue to choose the previously correct response post-reversal. Still, humans are typically able to shift responses within a few trials and maintain this response without additional errors (Cools et al., 2002; Izquierdo et al., 2017). Chimpanzees, however, may be especially susceptible to perseveration. Several studies have found that chimpanzees are fairly conservative in problem-solving tasks: once they have learned one solution, they may not readily change their response—even if it is less optimal than a new solution (Hrubesch et al., 2009; Marshall-Pescini & Whiten, 2008; Van Leeuwen & Call, 2017). Enhanced flexibility in humans may be underpinned by increased skill in using new information to shift response-outcome representations, but to date no work has examined how chimpanzees use probabilistic feedback to shift responses.

Across three studies, we therefore examined multiple aspects of chimpanzee cognitive flexibility and assessed their developmental trajectories. To do so, we tested a large sample of semi-free-ranging chimpanzees living in a naturalistic context. We used a reversal learning paradigm where individuals first learned one stimulus-reward contingency and then contingencies reversed, and implemented versions of this paradigm with varying levels of difficulty. As reversal learning tasks have been used with nonhumans, human children, and human adults, this allowed us to test for a skill that is broadly applicable; in addition, the complexity of related tasks (such as set-shifting) precluded testing larger developmental samples of apes in this context. We designed tasks that involved minimal training to capture more spontaneous responses to reversal problems, and to ensure that a broad age

range could participate. Study 1 examined chimpanzees ranging from juvenility to adulthood on a simple spatial reversal task, where one of two possible spatial locations contained a food reward. Study 2 contrasted shifting using spatial versus arbitrary perceptual cues to test whether information type constrains chimpanzee cognitive flexibility. Finally, Study 3 examined the development of shifting in response to probabilistic feedback.

Study 1: Development of shifting abilities

This study used a serial spatial reversal learning task to assess developmental change in chimpanzee cognition. Each chimpanzee completed one session where a spatial cue (left or right side) predicted reward location, and then the predictive side switched up to two subsequent times (see Figure 1a-b).

a. Ethics statement

These behavioral studies at Tchimpounga Chimpanzee Sanctuary were approved by the Ministry of Scientific Research and Technological Innovation in Republic of Congo and the Jane Goodall Institute. They had Institutional Animal Care and Use Committee approval from the University of Michigan (#8102) and Harvard University (#14-07-206-1).

b. Subjects

We tested 82 semi-free-ranging chimpanzees from the Tchimpounga Chimpanzee Sanctuary in the Republic of Congo (45 males, 37 females; range 7-33 years, mean 19.5; see Table S1 for breakdown). This sample size exceeds prior studies of chimpanzee shifting abilities; to examine individual differences and developmental change, our goal was to test as many chimpanzees from this population as were available for testing and willing to participate in the study. Tchimpounga is an accredited member of the Pan-African Sanctuary Alliance (PASA), and animal care complied with PASA standards. Apes in African sanctuaries are typically wild-born and arrive at the sanctuary between 1-3 years of age. These chimpanzees spend most of their time in large forest enclosures in

species-appropriate social groups; prior work shows typical cognition, behavior, and physiology in this population (Cole et al., 2020; Rosati et al., 2013; Wobber & Hare, 2011). Chimpanzees had ad libitum access to water, were never food deprived for testing, and were tested in familiar night dormitories. All sessions were voluntary; if the chimpanzee stopped participating, the session ended.

c. Procedure

Chimpanzees were tested in their indoor dormitories. The experimenter and the chimpanzee sat across from each other at a table with a sliding top (80cm x 40cm x 50cm), separated by the bars of the chimpanzee's room. The experimenter could set up the two containers on the table and then push the tabletop within the chimpanzee's reach, so they could make a choice by touching or pointing at one of the containers (Figure 1a). The experimenter looked down at the midline of the table to avoid potential social cuing while the chimpanzee chose, and gave them the food once the chimpanzee indicated their choice. The session consisted of four phases: a *warm-up* to familiarize the basic setup; an initial *learning phase* where they learned a spatial contingency; a *first reversal phase* where the contingency was switched; and a *second reversal phase* where the contingency was switched again. Analyses focused on 82 individuals who passed at least the learning phase. Some individuals reached pre-set trial limits (described below) or stopped participating during the study; these individuals were included in analyses for which they contributed relevant data.

First, each chimpanzee completed four *warm-up trials* where food (a banana slice or peanut depending on the chimpanzee's preferences) was placed directly on the left or right side of the table to ensure that all were able to make clear choices. They then completed a series of *learning trials* where two identical containers were placed on the left and right sides of the table, and they had to learn which side (counterbalanced across chimpanzees) was consistently baited. Here, the experimenter visibly placed a reward in the center of the table; blocked their view with an occluder (76cm x 50cm); and then baited and fake-baited the two containers using consistent motions from right to left to prevent any cues as to the correct container. Chimpanzees had to choose the correct side on 10 out of their last 12 trials to proceed. We implemented a 50-trial learning trial maximum set to ensure they

stayed motivated to participate across the session. One chimpanzee did not reach the learning criterion within 50 trials on his first attempt, but was later successfully retested and included in the sample; another chimpanzee did not reach criterion within 50 trials but was unavailable to be retested so was not included in the final sample.

Next, in the *first reversal* phase, the rewarded side was switched (e.g., if the reward was on the right side in learning trials, it was now on the left) and trials proceeded with the same procedure as learning trials. Again, chimpanzees needed to choose correctly on 10/12 trials. We implemented a pre-set maximum of 75 total trials (learning plus reversal trials) to ensure that chimpanzees stayed motivated to participate. In the first reversal, five reached the 75-trial maximum, and an additional seven stopped participating. These individuals were included in the main analyses of learning and the first reversal as GLMMs are able to account for unequal repeats; additional checks removing these individuals showed comparable results to the full sample. Additionally, if chimpanzees took more than 40 trials to reach 10/12 correct we had a pre-set rule that they did not proceed to the second reversal. Six chimpanzees did not proceed based on this rule. In total, 64 chimpanzees proceeded to the *second reversal*, where the rewarded side switched again. During this phase, nine reached the 75-trial maximum and two more stopped participating.

d. Coding and data analysis

The chimpanzee's choices were coded live by the experimenter. All sessions were videotaped, and a coder blind to the hypotheses of the study coded 20% of sessions from video with high reliability ($K=.998$, $n=908$ trials). We analyzed data in Rv4.1.2 (R Core Team, 2021). To analyze trial-by-trial binary choices, we implemented generalized linear mixed models (GLMMs) with a binomial structure using the *glmer* function from the *lme4* package (Bates, 2010). All models included random *subject* effects to account for repeated measurements (Baayen, 2008). To assess the effects of developmental stage on performance, we split chimpanzees into three cohorts based on life history characteristics (Goodall, 1983; Kawanaka, 1989): a *juvenile* cohort of chimpanzees up to 15 years; a *young adult* cohort up to 20 years; and an *adult* cohort 20 years and up. We implemented cohort as an ordered factor in analyses, which can assess both linear and non-linear age effects. We compared

model fit using likelihood-ratio tests (LRTs), and additionally report Akaike Information Criterion (AIC), where lower AIC scores indicate better fit (Bolker et al., 2008; Burnham & Anderson, 2002). For post-hoc comparisons, we used the *emmeans* package with a Tukey correction (Lenth et al., 2018). Graphs showing predicted effects and 95% confidence intervals from these models were calculated using the effects package (Fox, 2003).

e. Results and discussion

We first confirmed that reversal was in fact more difficult for the chimpanzees than the initial learning phase, as expected in a reversal learning task (see Figure 2a). To do so, we compared chimpanzees' overall performance between learning and reversal phases. Learning the initial rule was indeed easier than reversing it: chimpanzees took mean=14.9 ± SE=0.8 trials to pass the learning phase, compared with 26.5±1.5 trials to reach criterion (or trial limits) in the first reversal phase (excluding individuals who stopped participating). The 70 individuals who successfully passed the first reversal within trial limits did so in 25.0±1.3 trials. In the second reversal, chimpanzees took 19.5±1.0 trials, with the subset of 53 individuals who passed within trial limits doing so in 19.2±1.0 trials. We also assessed two types of errors: *perseverative errors*, consecutive incorrect choices after reversal, and the percent of *regressive errors*, or incorrect choices after at least one correct choice. Chimpanzees perseverated for an average of 5.8±0.7 trials after the first reversal. After making a correct choice, they made regressive errors on 30.2±2.5% of their remaining trials (calculated for chimpanzees who completed at least 10 trials after perseveration—e.g., at least 9 trials after the first correct choice).

We used GLMMs to compare performance in the learning and first reversal phases to confirm that the reversal phase was more difficult than initial learning. The base model included *subject* as a random effect, *trial number* (within each trial type) as a continuous predictor, and *sex* and *cohort* as fixed effects. We then added *trial type* (learning or reversal), which improved fit [LRT: $\chi^2=366.24$, $df=1$, $p<0.0001$; AIC 1=4185.6, AIC 2=3821.4; see Table S2 for model parameters]: as expected, chimpanzees were more successful in the learning phase.

Our main analyses then focused on chimpanzees' capacities to shift responses after the rewarded location flipped. To examine this, we analyzed trial-by-trial performance in the reversal phase. We constructed a base model of first reversal performance that again included *subject* as a random effect, *reversal phase trial number*, *sex*, and *cohort*, and *count of total learning trials* as a covariate (to account for any individual variation in experience needed to acquire the initial association). To test whether age cohorts differed in how quickly they shifted their responses over reversal trials, we then added the interaction between *trial number X cohort*. This improved model fit with a linear effect of cohort [LRT: $\chi^2=35.10$, $df=2$, $p<0.0001$; AIC 1=2317.8, AIC 2=2286.7]: post-hoc tests showed that adults improved more quickly than both juveniles and young adults ($p<0.0001$ for significant comparisons). We then assessed if performance varied by sex, adding the interaction between *trial number X sex*. This improved model fit [LRT: $\chi^2=24.25$, $df=1$, $p<0.0001$; AIC=2264.5]: post-hoc tests indicated females were faster to shift than males ($p<0.0001$). Finally, we added the three-way interaction of *trial number X sex X cohort*, which did not improve model fit [LRT: $\chi^2=5.28$, $df=4$, $p=0.26$; AIC=2267.3], indicating independent effects of sex and cohort. Model 3 received 80% of the weight in an AIC comparison of all models ($\Delta AIC 2.8$), showing that adults were faster to shift compared to juveniles and young adults, and females faster to shift compared to males (Figure 2b-2c; see Table S3).

In a second series of models, we used the same basic approach to examine performance in the second reversal phase, in the subset of subjects who reached that phase. These chimpanzees exhibited worse performance on both reversal phases compared to the learning phase. Trial-by-trial reversal analyses showed that, in this subsample, young adults were faster to shift than adults (see SOM for full reporting). However, as the chimpanzees that did not reach the second reversal phase (due to the trial limit criteria described above) were disproportionately young and male, we focused on the full sample in the first reversal phase for our primary developmental comparisons.

Overall, these results revealed that chimpanzees exhibit an extended period of developmental change as well as sex differences in cognitive flexibility. First, juvenile and young adult chimpanzees were slower to shift their responses compared to adults in the first reversal phase, indicating that this

skill continues to develop through young adulthood. We also found that female chimpanzees were quicker to shift their responses than males. Importantly, these analyses accounted for initial learning performance acquiring the spatial contingency, so these results reflect differences in chimpanzees' capacities to shift responses.

Study 2: Shifting across different information types

This study compared chimpanzees' ability to shift using spatial versus perceptual information. Each chimpanzee completed two conditions where both perceptual (the color and shape of the container) and spatial cues (the location of the container) varied independently, but only one predicted the presence of reward (see Figure 1c-1d). Based on past work demonstrating that apes preferentially use spatial information (Haun et al., 2006; Völter et al., 2019), we predicted chimpanzees would be more successful using spatial versus perceptual cues.

a. Subjects

We tested 24 chimpanzees (12 males, 12 females; range 7-32 years, mean 18.6; see Table S6) in a within-subjects design. Here we aimed to test individuals who had successfully completed all three phases in Study 1, to ensure they would be able to complete this more difficult task. Four additional chimpanzees began the study but either did not meet pre-set criteria for performance or stopped participating as described below.

b. Sessions

Each chimpanzee completed both the spatial and perceptual condition in counterbalanced order. The basic methods were similar to those in Study 1, except here the two containers differed in color and shape, and their spatial location quasi-randomly varied over trials (Figure 1c). In the *spatial condition*, one side (left or right) was always baited, regardless of which container (cup or box, varying in color) was placed there. Chimpanzees had to use spatial information as in Study 1, but here the containers also provided irrelevant perceptual cues. In the *perceptual condition*, however, the perceptual features of the container predicted the reward, regardless of which side that container was

placed on. This required chimpanzees to focus on perceptual information while ignoring irrelevant spatial information. These two conditions thereby provided a matched comparison where both kinds of cues were always present, but only one was relevant to acquiring the reward in each condition.

Each chimpanzee completed both conditions, with a break of at least one day in between to reduce carryover effects. Each condition consisted of a single session with a *learning phase* (where the reward was always on one side or under one container, initial side and container counterbalanced), followed by a *reversal phase* (where the opposite side or container predicted reward). Different pairs of containers were used to reduce the possibility of transfer between conditions (a dark blue cup and green box were always used in the first session, a red cup and light blue box in the second, counterbalancing which containers were used in each condition across participants).

c. Procedure

The procedure was largely the same as in Study 1. Chimpanzees first completed four *warm-up* trials (identical to those in Study 1) and then additionally completed four *visible baiting trials* (identical to subsequent learning trials, except without the occluder, in order to facilitate learning of the correct cue-reward pairing). In *learning trials*, as in Study 1, chimpanzees needed to choose correctly on 10/12 trials within a 50-trial maximum. Three chimpanzees included in the final sample did not meet this criterion during the perceptual condition and were retested after a delay. Three additional chimpanzees were excluded from the study because they either did not meet the learning criterion and were not available to be retested, or because they stopped participating. After passing the learning phase, contingencies reversed so that the other side or container predicted the reward. Again, chimpanzees needed to choose correctly on 10/12 trials. Here, we implemented a maximum of 40 reversal trials to ensure motivation. One chimpanzee stopped participating during the reversal phase and could not be retested, so was excluded from the final sample.

d. Coding and data analysis

The chimpanzee's choice was coded live by the experimenter. All sessions were videotaped, and a coder blind to the hypotheses of the study coded 20% of sessions from video with perfect

reliability ($K=1$, $n=547$ trials). Statistical analyses followed the same general approach described for Study 1.

e. Results and discussion

We first compared chimpanzees' overall performance between learning and reversal phases. Chimpanzees took an average of 14.3 ± 1.3 trials to pass the learning phase in the spatial condition, similar to learning performance in Study 1. In contrast, they took 19.3 ± 2.0 trials in the learning phase of the perceptual condition. There was an even starker contrast between conditions in the reversal phase. In the spatial condition, chimpanzees completed an average of 25.3 ± 2.0 reversal trials (to reach the passing criterion or the 40-trial maximum); 20/24 chimpanzees successfully passed (doing so in an average of 22.3 trials). In the perceptual condition, chimpanzees completed an average of 38.4 ± 1.1 reversal trials; only 3/24 switched successfully (these few took 27.0 trials to do so; see Figure 3a). We also compared the types of errors that chimpanzees made, using the same metrics as in Study 1. Chimpanzees made an average of 4.3 ± 1.0 perseverative errors in the spatial condition, similar to Study 1, but 8.1 ± 1.2 errors in the perceptual condition. After at least one correct reversal choice, chimpanzees in the spatial condition made regressive errors on $27.1 \pm 3.6\%$ of their remaining trials, again similar to Study 1. In contrast, chimpanzees made regressive errors on $59.0 \pm 3.5\%$ of their remaining trials in the perceptual condition.

Our main question concerned how chimpanzees shifted responses using spatial versus perceptual information. As in Study 1, we first compared performance across the learning and reversal phases. To do so, we constructed a base model of correct choices including *subject* as a random effect, *trial number* (within trial type), *cohort*, and *sex*. Adding *trial type* (learning versus reversal) to a second model improved fit [LRT: $\chi^2=294.44$, $df=1$, $p<0.0001$; AIC 1=3223.8; AIC 2=2931.4]: as expected, performance was worse in the reversal phase. Adding *condition* (spatial or perceptual) further improved fit [LRT: $\chi^2=122.12$, $df=1$, $p<0.0001$; AIC=2811.3]: performance was better in the spatial condition overall. Finally, we added the interaction between *condition X trial type* to assess how learning versus reversal performance varied by condition. This improved model fit [LRT:

$\chi^2=10.55$, $df=1$, $p=0.005$; $AIC=2802.7$]: post-hoc tests showed a wider gap between learning and reversal performance in the perceptual condition compared to the spatial ($p=0.01$; see Table S7). This final model was strongly weighted (99%, ΔAIC 8.5) in an AIC comparison. Chimpanzees were challenged by the perceptual condition in both learning and reversal, but could learn the initial perceptual association—however, they especially struggled to shift responses based on perceptual cues.

We next examined trial-by-trial performance across the reversal phase, as in Study 1. We constructed a base model that included *subject* as a random effect, *trial number*, *sex*, and *cohort*, and *number of learning trials* (for that condition). We then added *condition*, which improved model fit [LRT: $\chi^2=110.38$, $df=1$, $p<0.0001$; AIC 1=1984.2, AIC 2=1875.8]: performance was worse in the perceptual condition. We next added the interaction between *trial number* \times *condition* to test whether the rate of improvement across trials differed between the conditions. This did not improve model fit [LRT: $\chi^2=2.55$, $df=1$, $p=0.11$, $AIC=1875.3$] (though this model received 57% of the weight by AIC comparison [ΔAIC 0.5 between model 2 and 3]). This indicates that chimpanzees performed worse in the perceptual condition overall but had similar rates of improvement over trials within both conditions (Figure 3b). This result may be due in part to a rapid jump in performance in the spatial reversal: in trial 1 of either condition, 1-2 chimpanzees spontaneously chose correctly (this should be low as they have no indicator of reversal); by trial 2, 29% of chimpanzees chose correctly in the spatial condition, while only 8% did so in the perceptual condition (see Figure S2). We additionally checked for developmental or sex effects as in Study 1, but found no significant effects (see SOM), which was not unexpected given that this sample was selected based on success in Study 1.

One possible explanation for the difference in performance between conditions is that by selecting chimpanzees who were able to successfully complete the spatial reversal task in Study 1, we tested chimpanzees who were more inclined to use spatial information at the expense of other information types. To address this, we examined whether there was any correlation between performance across the spatial and perceptual conditions. There was no relationship between spatial

and perceptual performance for either the learning ($r=-0.14$, $p=0.51$, n.s.) or reversal phase ($r=0.23$, $p=0.29$, n.s.). Given this, it is unlikely that success in the spatial context constrained performance in the perceptual context.

Overall, chimpanzees were overwhelmingly more successful at shifting based on spatial information. This is in line with past work indicating that apes may preferentially rely on spatial cues (Haun et al., 2006; Völter et al., 2019). Our work demonstrates that chimpanzees were able to successfully acquire rules based on perceptual features in the learning phase, but especially struggled to shift responses based on perceptual information. This relative difficulty may place a cognitive constraint on opportunities for behavioral flexibility in chimpanzees.

Study 3: Shifting in response to probabilistic feedback

Our final study examined chimpanzees' ability to shift based on probabilistic feedback (see Figure 1e). Here, the 'correct' option (predicted by congruent spatial and perceptual information) provided an 80% chance of a preferred food reward and 20% chance of a non-preferred reward; the 'incorrect' option provided the opposite (80% non-preferred, 20% preferred). After learning to choose the correct option, the contingencies were reversed. We predicted that this problem would be more difficult than a task with fixed payoffs, and that chimpanzees would exhibit greater perseveration when the contingencies reversed.

a. Subjects

We tested 40 chimpanzees (21 males, 19 females; range 6-25 years, mean 14.9; see Table S11) from the same population as studies 1 and 2. This study was conducted seven years prior, as one session in a battery of decision-making tasks (see Rosati et al., 2018). 22 of these chimpanzees later participated in Study 1, and nine also participated in Study 2—direct comparison of these individuals' performance is difficult, however, due to the time gap between studies.

b. Sessions and test phases

Each chimpanzee completed a single reversal learning session. On an earlier day they had completed a food preference test where they made choices between pairs of high-value (banana), intermediate-value (peanut), and low-value (cucumber) rewards. All chimpanzees in this study chose banana over cucumber (the two foods used in the task) 100% of the time.

The general procedure followed the prior studies, except in two respects. First, the rewards varied: the correct option provided a high-value reward 80% of the time and a low-value reward 20% of the time, while the incorrect option provided the reverse. Second, the correct response was predicted by both spatial and perceptual information. The containers differed in appearance (yellow cup and white bowl), but here they had fixed side assignments such that spatial and perceptual cues were congruent and both predictive. This setup was used as we anticipated this would be a challenging task. The rewarded option was counterbalanced across subjects.

As in the prior studies, chimpanzees first completed *learning trials* where they had to choose correctly on 10/12 trials to proceed. Here, there was a longer 60-trial maximum as this was expected to be difficult. Three chimpanzees did not pass on their first attempt, but were retested successfully after a delay. After passing the learning phase, the reward contingencies reversed; the previously-correct side now provided high-value rewards on only 20% of trials, and the previously-incorrect side now did so on 80% of trials. All chimpanzees completed 30 reversal trials.

e. Coding and data analysis

The chimpanzee's choice was coded live by the experimenter, and a second coder blind to the hypotheses of the study coded 20% of sessions from video ($K=1$, $n=346$). Statistical analyses followed the same general procedure as in Study 1. To ensure that results were comparable to the preceding studies, our primary analyses looked at performance until a chimpanzee reached 10/12 correct or reached the 30-trial maximum, as studies 1 and 2 only tested individuals until they reached the passing criterion or trial limits. Since all chimpanzees completed 30 reversal trials regardless of when they reached 10/12 correct, additional checks confirmed that using all trials did not impact primary results (see SOM).

f. Results and discussion

On average, chimpanzees required 15.2 ± 1.1 trials to pass the learning phase, similar to learning performance in Study 1 and the spatial condition of Study 2. In the reversal phase, chimpanzees took an average of 24.9 ± 1.0 trials to either pass or reach the 30-trial limit (see Figure 4a). Only 21/40 chimpanzees successfully shifted within those 30 trials, doing so in 20.2 ± 1.2 trials. Perseverative errors in this task, as expected, were quite high compared to the prior studies: 15.2 ± 1.8 trials. In fact, 10/40 individuals continued to choose the incorrect side for all 30 reversal trials. If a chimpanzee made at least one correct choice, they regressed back to the incorrect choice on $21.5 \pm 5.6\%$ of their remaining trials.

To compare learning and reversal, we created a base model as in the prior studies, with *subject* as a random effect, *cohort*, *sex*, and *trial number* (within trial type). Adding *trial type* in a second model improved fit [LRT: $\chi^2=473.58$, $df=1$, $p<0.0001$; AIC 1=2108.4, AIC 2=1636.8; see Table S12]; as expected, performance was worse in the reversal phase. This model received 100% of the weight in an AIC comparison (ΔAIC 471.6). We then examined whether rates of improvement over trials in the reversal phase varied by sex and cohort. As in the prior studies, the base model included *subject* as a random effect, *trial number*, *sex*, *cohort*, and *number of learning trials*. Adding the interaction between *trial number* \times *cohort* improved model fit [LRT: $\chi^2=7.94$, $df=2$, $p=0.02$; AIC 1: 632.7, AIC 2: 628.9]: post-hoc comparisons indicated that adults were faster to shift compared to young adults ($p=0.01$). Adding the interaction between *trial number* \times *sex* further improved fit [LRT: $\chi^2=21.95$, $df=1$, $p<0.0001$; AIC=609.0]: post-hoc comparisons showed that females were faster to shift compared to males ($p=0.0001$). This model also revealed linear and quadratic effects of cohort, such that adults were faster to shift compared to young adults ($p<0.005$) and trended to shift faster than juveniles ($p=0.08$) (Figure 3c-3d; see Table S13). Adding the three-way interaction *trial number* \times *cohort* \times *sex* did not further improve model fit [LRT: $\chi^2=2.04$, $df=4$, $p=0.73$; AIC=615.1], indicating independent effects of cohort and sex.

We also looked at how chimpanzees responded to the positive and negative feedback received from each choice, by examining when they switched their next response (excluding the first trial from each phase). In the learning phase, chimpanzees exhibited very low rates of switching after positive feedback—3-4% of the time whether feedback was from the incorrect or correct choice. In response to negative feedback, they switched at higher rates as expected: 19% of the time when that feedback came from the correct choice, and 59% of the time when it came from incorrect choice. This indicates that they were able to learn to maintain the response that offered the higher proportion of positive feedback, despite receiving positive and negative feedback from both options. In the reversal phase, in contrast, they switched more in response to negative feedback from the newly-correct option (32%) versus the incorrect option (16%) (Figure 4b). In line with these numbers, analyses found that chimpanzees were more likely to switch based on negative feedback, and more likely to switch after choosing the incorrect option, in learning compared to reversal (see SOM for details). This indicates high rates of perseveration post-reversal on the response that previously provided a high proportion of the high-value reward.

Overall, chimpanzees struggled to shift based on probabilistic feedback, as evidenced by their high rates of perseverative errors compared to the other studies. Despite the difference in task difficulty, we also found a period of extended development as well as sex differences that mirror those seen in Study 1, where we used a simpler spatial task with a larger sample of chimpanzees. In particular, young adults were slower to shift compared to adults in the reversal phase; juveniles also trended to show worse performance than adults. In addition, males were slower to shift compared to females. This suggests that the developmental and sex differences seen in Study 1 may reflect a more general pattern that impacts cognitive flexibility across contexts.

General Discussion

Across three studies, we examined the development and nature of chimpanzee cognitive flexibility. We first characterized the development of this skill using a simple spatial task, and found

an extended developmental trajectory as well as a sex difference where females were faster to shift responses than males. We then compared shifting in response to spatial versus perceptual cues, and found that chimpanzees are much more successful at shifting based on spatial information even though they could initially learn to use both cues. Finally, we examined shifting in response to probabilistic feedback and found that chimpanzees showed high levels of perseveration but also largely parallel developmental and sex differences in this context. Overall, these results highlight some commonalities with human performance in the extended development of these skills, but also support the hypothesis that there are core features of human cognitive flexibility that are distinct from those seen in other apes.

In humans, executive functions like cognitive flexibility can take many years to reach mature levels of performance (Anderson, 2002; Diamond, 2002; Zelazo et al., 2003). For example, in the Wisconsin Card Sorting Task, assessing rule-based set shifting, children do not perform comparably to adults until 10-11 years old (Diamond, 2002). However, simple response shifting can develop earlier, as young as 3-4 years old (Anderson, 2002). In Study 1, we found that chimpanzees up to age 20 were slower to shift responses compared to adults. While matched comparisons with children will be crucial to directly compare developmental trajectories, these results support the idea that the development of shifting abilities may actually be slower in chimpanzees than in humans. This contrasts with several proposals that nonhumans like chimpanzees should exhibit faster cognitive developmental trajectories compared to humans, in line with their more rapid physical maturation (Bjorklund & Green, 1992; Charnov & Berrigan, 1993; Kaplan et al., 2000). These results also show that language or formal schooling are not strictly necessary for developmental change in executive functions, however differences in developmental pace between humans and nonhuman apes may reflect facultative effects of these human-specific processes.

These findings align with accumulating evidence that basic cognitive abilities shared with other animals can develop more rapidly in humans than in nonhumans (Herrmann et al., 2007; Langer, 2006; Wobber et al., 2014). For example, human children outpaced chimpanzees in basic physical and social reasoning abilities by age 4 (Wobber et al., 2014), and gaze-following abilities

that emerge in the first year of life in humans may not emerge until 2-3 years in chimpanzees (Moll & Tomasello, 2004; Tomasello et al., 2005). This suggests that humans' long juvenile period is not necessarily linked to overall slower development. Rather, humans may show relatively accelerated development of abilities that are shared with other primates, and then go on to develop even more complex skills scaffolded by earlier-emerging capacities. For example, slower development of basic response shifting in chimpanzees may limit set-shifting abilities, which emerge later in humans (Anderson, 2002; Diamond, 2013). Thus, differences between nonhuman and human cognitive development may lie not just in the *pace* of development of one particular skill, but in the *patterning* of skills (Rosati et al., 2014, 2016; Wobber et al., 2014), as the earlier acquisition of core cognitive capacities may facilitate the development of more elaborate abilities. In these studies, we used a cross-sectional approach, as this allowed us to assess a large sample from across the age range, and avoided potential training effects from repeated testing. However, future work incorporating a longitudinal approach would be useful to confirm the developmental results found here (e.g., as in Wobber et al., 2014), and would be crucial to examine causal patterning between different components of executive function across development.

We also found that females were quicker to shift their responses compared to males in both the spatial and probabilistic reversal contexts. Humans do not appear to exhibit major gender differences in executive functions, though girls may perform better on some tasks of attention and inhibitory control (reviewed in Grissom & Reyes, 2019). There is also very little current evidence for sex differences in chimpanzee cognition, whether in executive function skills or other cognitive domains. For example, in a sample of 100 chimpanzees tested on a large battery of social and physical cognitive tasks, males outperformed females in spatial memory and object knowledge tasks but no other sex differences were detected (Herrmann et al., 2007). There is also little evidence for sex differences in primate decision-making more generally (De Petrillo & Rosati, 2021). Interestingly, however, some reversal learning studies in non-primate species found that females outperformed males, potentially linked to the effects of testosterone (Guillamón et al., 1986; Lucon-Xiccato & Bisazza, 2014; Neese & Schantz, 2012; Rogers, 1974). Importantly, chimpanzees do exhibit sex

differences in behaviors that may recruit executive functions. For example, wild female chimpanzees use tools at a higher rate or with greater efficiency than males (Boesch & Boesch, 1981; Gruber et al., 2010; Lonsdorf et al., 2004; Lonsdorf, 2005; McGrew, 1979), and develop these skills earlier (Lonsdorf et al., 2004; Lonsdorf, 2005). While there is currently no data linking variation in executive functions to real-world behaviors in chimpanzees, one interesting possibility is that sex differences in executive function play a role in this variation in wild behavior. Future work on executive function should consider sex differences in these skills in nonhuman primates, as well as the potential for gender differences in the development of human executive function.

Finally, we found that chimpanzee cognitive flexibility may be constrained in some important ways relative to humans. For example, chimpanzees struggled to shift responses based on arbitrary perceptual features, even when they could successfully learn the initial perceptual cue. This aligns with a growing body of evidence that chimpanzees preferentially track information and make choices based on concrete cues that causally predict rewards, like spatial location, but are less successful at reasoning using more abstract features, like arbitrary color or shape cues (Haun et al., 2006; Penn et al., 2008; Seed et al., 2011). While chimpanzees are certainly able to shift based on arbitrary features with enough experience (e.g., as reviewed in Rumbaugh, 1997), their difficulty in spontaneously using these cues may constrain cognitive flexibility relative to humans. These differences may partially reflect how language development reshapes relevant cognitive representations. For example, in prior work apes and 1-year-old children made choices between objects based on a spatial strategy, but older children, who had acquired language, used a strategy focused on perceptual cues (Haun et al., 2006). Similarly, some proposals argue that language specifically allows children to integrate spatial and perceptual information in other contexts (e.g., Hermer-Vazquez et al., 1999, 2001). Future work could examine if younger versus older children show a distinction between spatial and perceptual information in cognitive flexibility tasks like those used here. Another important step would be to further define the types of information that chimpanzees flexibly use—for example, chimpanzees may be more successful when reasoning about a cue's functional properties, or using natural categories relevant to real-world decisions (Herrmann et al., 2008; Seed et al., 2011).

We similarly found that probabilistic feedback may constrain chimpanzee shifting abilities. Chimpanzees were able to effectively use probabilistic feedback in the learning phase, but their ability to rapidly adjust based on changes in probabilistic information was more limited, and they were highly perseverative despite accumulating negative feedback. While there is no exact comparison between our task and computer-based tasks used with adult humans, in one example human adults averaged 2.6 perseverative errors during a probabilistic reversal task (Cools et al., 2002). Here, chimpanzees averaged 15.2 perseverative errors, with 25% of chimpanzees continuing to choose the previously correct option throughout the entire reversal phase. This generally aligns with the proposal that chimpanzees may be relatively conservative problem-solvers (Call, 2015). For example, a number of studies on social learning find that chimpanzees will maintain a previously-learned solution even when more effective options become apparent (Hrubesch et al., 2009; Marshall-Pescini & Whiten, 2008). Minimal motivation to shift from an initial solution, despite lower levels of positive feedback, may serve as one constraint on innovation or social learning in chimpanzees. However, although chimpanzees as a group struggled in our task, there was significant inter-individual variation: some individuals shifted after only 2-3 perseverative errors, a pattern similar to human performance. Future work should examine the mechanisms underlying these individuals' successful use of probabilistic feedback.

Together, these studies suggest that there have been evolutionary changes in human cognitive flexibility. Overall, chimpanzees' shifting skills develop at a slower pace than that observed in humans. Chimpanzees also exhibit important constraints on cognitive flexibility, both by information type, and in response to probabilistic feedback. This generally aligns with prior inferences from neuroanatomy about the cognitive consequences of humans' large and complex prefrontal cortices relative to nonhuman apes (Sherwood et al., 2008; Smaers et al., 2017; Teffer & Semendeferi, 2012). Importantly, executive function is an umbrella term comprising multiple regulatory skills that shape behavior across contexts. While we have examined the scope and limitations of chimpanzee cognitive flexibility, a crucial question concerns how these potential differences in cognition translate into flexible, goal-oriented behaviors, in conjunction with other executive function skills. Linking

variation in cognition to variation in naturalistic, real-world problem-solving will be the next step in understanding the evolution of human executive functions.

References

- Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility and inhibitory control in primates. *Current Biology*, *18*, 1415–1419.
- Anderson, P. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology*, *8*(2), 71-82.
- Aureli, F., Schaffner, C. M., Boesch, C., Bearder, S. K., Call, J., Chapman, C. A., Connor, R., Di Fiore, A., Dunbar, R. I., Henzi, S. P., Holekamp, K., Korstjens, A. H., Layton, R., Lee, P., Lehmann, J., Manson, J. H., Ramos-Fernandez, G., Strier, K. B., & Schaik, C. P. V. (2008). Fission-fusion dynamics: new research frameworks. *Current Anthropology*, *49*(4), 627-654.
- Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics*. Cambridge University Press, Cambridge, MA.
- Barth, J., & Call, J. (2006). Tracking the displacement of objects: a series of tasks with great apes (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, and *Pongo pygmaeus*) and young children (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(3), 239.
- Bates, D. The LME4 package: linear mixed-effects models using S4 classes. (2010).
- Boesch, C., & Boesch, H. (1981). Sex differences in the use of natural hammers by wild chimpanzees: A preliminary report. *Journal of Human Evolution*, *10*(7), 585-593.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, *24*(3), 127-135.
- Biro, D., Inoue-Nakamura, N., Tonooka, R., Yamakoshi, G., Sousa, C., & Matsuzawa, T. (2003). Cultural innovation and transmission of tool use in wild chimpanzees: Evidence from field experiments. *Animal Cognition*, *6*(4), 213–223.
- Bjorklund, D., & Green, B. (1992). The adaptive nature of cognitive immaturity. *American Psychologist*, *47*, 46-54.

- Bruner, J. S. (1972). Nature and uses of immaturity. *American Psychologist*, 27(8), 687–708.
- Bull, R., & Scerif, G. (2001). Executive Functioning as a Predictor of Children's Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology*, 19(3), 273–293.
- Burnham, K. P., & Anderson, D.R. (2002). Model selection and multi-modal inference: a practical information-theoretic approach. New York, NY: Springer-Verlag.
- Call, J. (2015). Conservatism versus innovation: The great ape story. In: *Animal creativity and innovation*. Academic Press, pp. 397-418.
- Charnov, E. L., & Berrigan, D. (1993). Why do female primates have such long lifespans and so few babies? Or life in the slow lane. *Evolutionary Anthropology: Issues, News, and Reviews*, 1(6), 191-194.
- Cole, M. F., Cantwell, A., Rukundo, J., Ajarova, L., Fernandez-Navarro, S., Atencia, R., & Rosati, A. G. (2020). Healthy cardiovascular biomarkers across the lifespan in wild-born chimpanzees (Pan troglodytes). *Philosophical Transactions of the Royal Society B*, 375(1811), 20190609.
- Cook, P., & Wilson, M. (2010). Do young chimpanzees have extraordinary working memory? *Psychonomic Bulletin & Review*, 17(4), 599–600.
- Cools R., Clark L., Owen A.M., & Robbins T.W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, 22, 4563-4567.
- Deaner, R., van Schaik, C., Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology*, 4, 149–196.
- Decker, S., Hill, S., & Dean, R. (2007). Evidence of construct similarity in Executive Functions and Fluid Reasoning abilities. *The International journal of neuroscience*, 117, 735-48.
- De Petrillo, F., & Rosati, A. G. (2021). Variation in primate decision-making under uncertainty and the roots of human economic behaviour. *Philosophical Transactions of the Royal Society B*, 376(1819), 20190671.
- Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. *Principles of frontal lobe function*, 466, 503.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168.
- Dunbar, R. I. M., McAdam, M. R., & O'Connell, S. (2005). Mental rehearsal in great apes (Pan troglodytes and Pongo pygmaeus) and children. *Behavioural processes*, 69(3), 323-330.
- Emery Thompson, M., Jones, J.H., Pusey, A.E., Brewer-Marsden, S., Goodall, J., Marsden, D.,

- Matsuzawa, T., Nishida, T., Reynolds, V., Sugiyama, Y., & Wrangham, R.W. (2007). Aging and fertility patterns in wild chimpanzees provide insights into the evolution of menopause. *Current Biology*, *17*, 2150–2156.
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of statistical software*, *8*(15), 1-27.
- Gómez, J. C. (2005). Species comparative studies and cognitive development. *Trends in cognitive sciences*, *9*(3), 118-125.
- Goodall, J. (1983). Population dynamics during a 15 year period in one community of free-living chimpanzees in the Gombe National Park, Tanzania. *Zeitschrift fuer tierpsychologie*, *61*(1), 1-60.
- Goodall, J. (1964). Tool-Using and Aimed Throwing in a Community of Free-Living Chimpanzees. *Nature*, *201*, 1264–1266.
- Grissom, N. M., & Reyes, T. M. (2019). Let's call the whole thing off: evaluating gender and sex differences in executive function. *Neuropsychopharmacology*, *44*(1), 86-96.
- Gruber, T., Clay, Z., & Zuberbühler, K. (2010). A comparison of bonobo and chimpanzee tool use: evidence for a female bias in the Pan lineage. *Animal Behaviour*, *80*(6), 1023-1033.
- Guillamón, A., Valencia, A., Calés, J., & Segovia, S. (1986). Effects of early postnatal gonadal steroids on the successive conditional discrimination reversal learning in the rat. *Physiology & behavior*, *38*(6), 845-849.
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal behaviour*, *61*(1), 139-151.
- Haun, D. B. M., Call, J., Janzen, G., & Levinson, S. C. (2006). Evolutionary Psychology of Spatial Representations in the Hominidae. *Current Biology*, *16*(17), 1736–1740.
- Healy, S.D., & Rowe, C. (2007). A critique of comparative studies of brain size. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1609), 453-464.
- Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive psychology*, *39*(1), 3-36.
- Hermer-Vazquez, L., Moffet, A., & Munkholm, P. (2001). Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition*, *79*(3), 263-299.
- Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., Tomasello, M. (2007). Humans have

- evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science*, 317, 1360–66
- Herrmann, E., Wobber, V., Call, J. (2008). Great apes' understanding of tool functionality after limited experience. *Journal of Comparative Psychology*, 122(2), 220-230.
- Herrmann, E., Misch, A., Hernandez-Lloreda, V., Tomasello, M., (2015). Uniquely human self control begins at school age. *Developmental Science*, 18(6), 979-993.
- Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal cognition*, 8(3), 164-181.
- Hrubesch, C., Preuschoft, S., & van Schaik, C. (2009). Skill mastery inhibits adoption of observed alternative solutions among chimpanzees (*Pan troglodytes*). *Animal Cognition*, 12, 209-216.
- Hughes, C., (1998). Executive function in preschoolers: links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16, 233–253.
- Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, 43(6), 1447-1459.
- Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, 17, 1004–1005.
- Izquierdo, A., Brigman, J. L., Radke, A. K., Rudebeck, P. H., & Holmes, A. (2017). The neural basis of reversal learning: An updated perspective. *Neuroscience*, 345, 12-26.
- Kafadar, H., & Orhan, İ. B. (2015). The Relationship between Wisconsin Card Sorting Test and Raven Standard Progressive Matrices: A Latent Variable Analysis. *International Online Journal of Educational Sciences*, 8(1).
- Kaplan, H., Hill, K., Lancaster, J., & Hurtado, M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology*, 9, 156-185.
- Kawanaka, K. (1989). Age differences in social interactions of young males in a chimpanzee unit-group at the Mahale Mountains National Park, Tanzania. *Primates*, 30(3), 285-305.
- Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development*, 44(2), 107–115.

- Kuhn, L. J., Willoughby, M. T., Wilbourn, M. P., Vernon-Feagans, L., Blair, C. B., & Family Life Project Key Investigators. (2014). Early communicative gestures prospectively predict language development and executive function in early childhood. *Child development, 85*(5), 1898-1914.
- Lacreuse, A., Parr, L., Chennareddi, L., & Herndon, J. (2018). Age-related decline in cognitive flexibility in female chimpanzees. *Neurobiology of Aging, 72*.
- Laland, K., & Seed, A. (2021). Understanding human cognitive uniqueness. *Annual Review of Psychology, 72*, 689-716.
- Langer, J. (2006). The Heterochronic Evolution of Primate Cognitive Development. *Biological Theory, 1*, 41-43.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. Package 'emmeans': <https://cran.r-project.org/web/packages/emmeans/index.html> (2018)
- Logan, C. J., Avin, S., Boogert, N., Buskell, A., Cross, F. R., Currie, A., Jelbert, S. A., Lukas, D., Mares, R., Navarrete, A. F., Shigeno, S., & Montgomery, S. H. (2018). Beyond brain size: Uncovering the neural correlates of behavioral and cognitive specialization. *Comparative cognition and behavior reviews*.
- Lonsdorf, E. V., Eberly, L. E., & Pusey, A. E. (2004). Sex differences in learning in chimpanzees. *Nature, 428*(6984), 715-716.
- Lonsdorf, E. V. (2005). Sex differences in the development of termite-fishing skills in the wild chimpanzees, *Pan troglodytes schweinfurthii*, of Gombe National Park, Tanzania. *Animal Behaviour, 70*(3), 673-683.
- Lonsdorf, E. V. (2006). What is the role of mothers in the acquisition of termite-fishing behaviors in wild chimpanzees (*Pan troglodytes schweinfurthii*)? *Animal Cognition, 9*(1), 36-46.
- Lucon-Xiccato, T., & Bisazza, A. (2014). Discrimination reversal learning reveals greater female behavioural flexibility in guppies. *Biology Letters, 10*(6), 20140206.
- MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., Aureli, F., Baker, J. M., Bania, A. E., Barnard, A. M., Boogert, N. J., Brannon, E. M., Bray, E. E., Bray, J., Brent, L. J., Burkart, J. M., Call, J., Cantlon, J. F., Cheke, L. G., ... Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America, 111*, E2140-E2148.
- Manrique, H.M., & Call, J. (2015). Age-dependent cognitive inflexibility in great apes. *Animal Behaviour, 102*, 1-6.

- Marshall-Pescini, S., & Whiten, A. (2008). Chimpanzees (*Pan troglodytes*) and the question of cumulative culture: an experimental approach. *Animal Cognition*, *11*, 449-456.
- Matsuzawa, T. (2007). Comparative cognitive development. *Developmental science*, *10*(1), 97-103.
- McCrea, S. M., Mueller, J. H., & Parrila, R. K. (1999). Quantitative analyses of schooling effects on executive function in young children. *Child Neuropsychology*, *5*(4), 242-250.
- McGrew, W. S. (1979). Evolutionary implications of sex differences in chimpanzee predation and tool use. *The great apes*.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature reviews neuroscience*, *1*(1), 59-65.
- Miller E.K., & Cohen J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*:167–202.
- Miller, D.J., Duka, T., Stimpson, C. D., Schapiro, S. J., Baze, W. B., McArthur, M. J., Fobbs, A. J., Sousa, A. M. M., Sestan, N., Wildman, D. E., Lipovich, L., Kuzawa, C. W., Hof, P. R., Sherwood, C. C. (2012). Prolonged myelination in human neocortical evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 16480–16485.
- Moll, H., & Tomasello, M. (2004). 12-and 18-month-old infants follow gaze to spaces behind barriers. *Developmental science*, *7*(1), F1-F9.
- Neese, S. L., & Schantz, S. L. (2012). Testosterone impairs the acquisition of an operant delayed alternation task in male rats. *Hormones and behavior*, *61*(1), 57-66.
- Nishida, T. (1968). The social group of wild chimpanzees in the Mahali Mountains. *Primates*, *9*, 167-224.
- Nishida, T., Corp, N., Hamai, M., Hasegawa, T., Hiraiwa-Hasegawa, M., Hosaka, K., Hunt, K.D., Itoh, N., Kawanaka, K., Matsumoto-Oda, A., Mitani, J.C., Nakamura, M., Norikoshi, K., Sakamaki, T., Turner, L., Uehara, S., & Zamma, K. (2003). Demography, female life history, and reproductive profiles among the chimpanzees of Mahale. *American Journal of Primatology*, *59*, 99-121.
- Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2), 109–130.
- Pusey, A. E. (1990). Behavioural changes at adolescence in chimpanzees. *Behaviour*, *115*(3-4),

- R Development Core Team. A language and environment for statistical computing. (Vienna, Austria) (2019).
- Rilling, J. K., & Insel, T. R. (1999). The primate neocortex in comparative perspective using magnetic resonance imaging. *Journal of Human Evolution*, 37(2), 191–223.
- Rogers, L. J. (1974). Persistence and search influenced by natural levels of androgens in young and adult chickens. *Physiology & behavior*, 12(2), 197-204.
- Rosati, A. G., Herrmann, E., Kaminski, J., Krupenye, C., Melis, A. P., Schroepfer, K., Tan, J., Warneken, F., Wobber, V., & Hare, B. (2013). Assessing the psychological health of captive and wild apes: A response to Ferdowsian et al. (2011).
- Rosati, A. G., Wobber, V., Hughes, K., & Santos, L. R. (2014). Comparative developmental psychology: how is human cognitive development unique?. *Evolutionary Psychology*, 12(2), 147470491401200211.
- Rosati, A. G., Arre, A. M., Platt, M. L., & Santos, L. R. (2016). Rhesus monkeys show human-like changes in gaze following across the lifespan. *Proceedings of the Royal Society B: Biological Sciences*, 283(1830), 20160376.
- Rosati, A. G. (2017). The evolution of primate executive function: from response control to strategic decision-making. In: *Evolution of Nervous Systems, Second Edition*, Volume 3 (J. Kaas & L. Krubitzer, eds.). Amsterdam: Elsevier, pp. 423-437.
- Rosati, A. G., Di Nicola, L. M., & Buckholtz, J. W. (2018). Chimpanzee Cooperation is Fast and Independent From Self-Control. *Psychological Science*, 29(11), 1832-1845.
- Rumbaugh, D. M., & Gill, T. V. (1973). The learning skills of great apes. *Journal of Human Evolution*, 2(3), 171-179.
- Rumbaugh, D. M. (1997). Competence, cortex, and primate models: A comparative primate perspective. In: *Development of the prefrontal cortex: Evolution, neurobiology, and behavior*, (N. A. Krasnegor, G. R. Lyon, & P. S. Goldman-Rakic, eds.), pp. 117–139.
- Sanz, C. M., & Morgan, D. B. (2010). The Complexity of Chimpanzee Tool-Use Behaviors. In: *The mind of the chimpanzee: ecological and empirical perspectives* (E. V. Lonsdorf, S. R. Ross & T. Matsuzawa, eds.). Chicago, IL: University of Chicago Press, pp 127-140.
- Schenker, N. M., Desgouttes, A. M., & Semendeferi, K. (2005). Neural connectivity and cortical substrates of cognition in hominoids. *Journal of Human Evolution*, 49, 547–569.
- Seed, A. M., Call, J., Emery, N. J., & Clayton, N. S. (2009). Chimpanzees solve the trap problem

when the confound of tool-use is removed. *Journal of Experimental Psychology: Animal Behavior Processes*, 35(1), 23.

- Seed, A., Hanus, D., & Call, J. (2011). Causal knowledge in corvids, primates, and children. In: *Tool use and causal cognition* (T. McCormack, C. Hoerl, & Butterfill, S., eds). Oxford Press, pp 89-108.
- Semendeferi, K., Lu, A., Schenker, N., & Damasio, H. (2002). Humans and great apes share a large frontal cortex. *Nature Neuroscience*, 5, 272–276.
- Semendeferi, K., Teffer, K., Buxhoeveden, D., Park, M. S., Bludau, S., & Amunts, K. (2011). Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cerebral Cortex*, 21, 1485–1497.
- Sherwood, C. C., Subiaul, F., & Zawidzki, T. W. (2008). A natural history of the human mind: tracing evolutionary changes in brain and cognition. *Journal of anatomy*, 212(4), 426–454.
- Smaers J. B., Gómez-Robles A., Parks A. N., & Sherwood C. C. (2017). Exceptional evolutionary expansion of prefrontal cortex in great apes and humans. *Current Biology*, 27, 714–720.
- Teffer, K., & Semendeferi, K. (2012). Human prefrontal cortex: evolution, development, and pathology. In *Progress in brain research* (Vol. 195, pp. 191-218). Elsevier.
- Tomasello, M., Carpenter, M., & Hobson, R. P. (2005). The emergence of social cognition in three young chimpanzees. *Monographs of the Society for Research in Child development*, i-152.
- Van Leeuwen, E. J., & Call, J. (2017). Conservatism and “copy-if-better” in chimpanzees (Pan troglodytes). *Animal cognition*, 20(3), 575.
- Vlamings, P. H. J. M., Hare, B., & Call, J. (2010). Reaching around barriers: The performance of the great apes and 3–5-year-old children. *Animal Cognition*, 13(2), 273–285.
- Völter, C. J., & Call, J. (2014). Younger apes and human children plan their moves in a maze task. *Cognition*, 130(2), 186–203.
- Völter, C. J., Mundry, R., Call, J., & Seed, A. M. (2019). Chimpanzees flexibly update working memory contents and show susceptibility to distraction in the self-ordered search task. *Proceedings of the Royal Society B: Biological Sciences*, 286(1907), 20190715.
- Walker, K. K., Walker, C. S., Goodall, J., & Pusey, A. E. (2018). Maturation is prolonged and variable in female chimpanzees. *Journal of Human Evolution*, 114, 131–140.
- Willoughby, M. T., Wirth, R. J., Blair, C. B., & Greenberg, M. (2012). The measurement of

executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, 24(1), 226–239

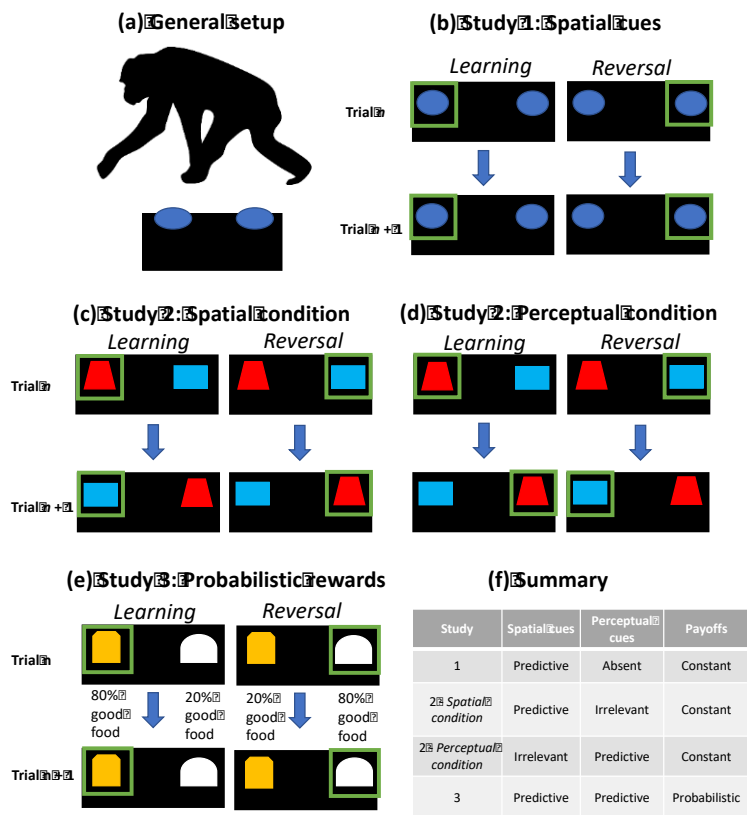
Wobber, V., Wrangham, R., & Hare, B. (2010) Bonobos exhibit delayed development of social behavior and cognition relative to chimpanzees. *Current Biology*, 20(3), 226-230.

Wobber, V., & Hare, B. (2011). Psychological health of orphan bonobos and chimpanzees in African sanctuaries. *PLOS ONE*, 6(6), e17147.

Wobber, V., Herrmann, E., Hare, B., Wrangham, R., & Tomasello, M. (2014). Differences in the early cognitive development of children and great apes. *Developmental Psychobiology*, 56, 547–573.

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., Chiang, J. K., Hongwanishkul, D., Schuster, B. V., Sutherland, A., & Carlson, S. M. (2003). The Development of Executive Function in Early Childhood. *Monographs of the Society for Research in Child Development*, 68(3).

Author Manuscript



Figures

Figure 1. Setup for studies. (a) Across reversal tasks, chimpanzee could find food rewards hidden under one of two containers. (b) In Study 1, the containers were visually identical, and the correct response depended on their spatial location (left or right side). (c) In the *spatial condition* of Study 2, location again predicted rewards, but the containers also visually differed and quasi-randomly appeared on either side; this feature information was irrelevant in this condition. (d) In the *perceptual condition* of Study 2, these perceptual cues instead predicted reward, and the spatial location of the correct container was then irrelevant. (e) In Study 3, spatial and perceptual cues both congruently predicted the correct response, but options provided probabilistic good versus bad payoffs. (f) Summary of cue types and payoffs across all studies and conditions.

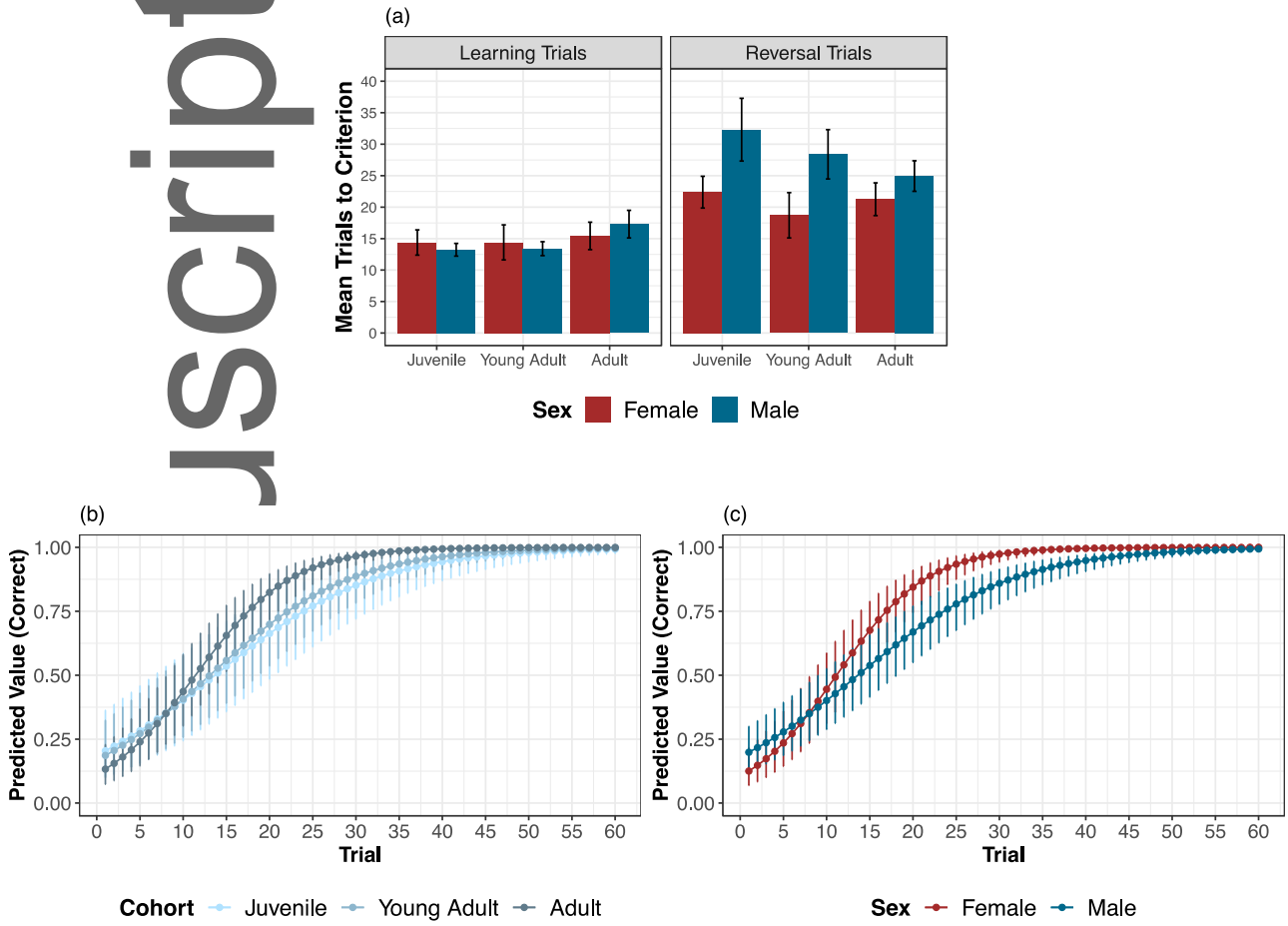


Figure 2. Performance in Study 1 by age cohort and sex. (a) Mean number of trials to reach the 10 out of 12 passing criterion in the learning and first reversal phases (or trial limits in the reversal phase); note that larger numbers indicate worse performance. Error bars indicate SE. (b) Estimates of trial-by-trial performance in the reversal phase by age cohort, and (c) by sex, from the best-fit model accounting for both. Error bars indicate 95% confidence interval estimates.

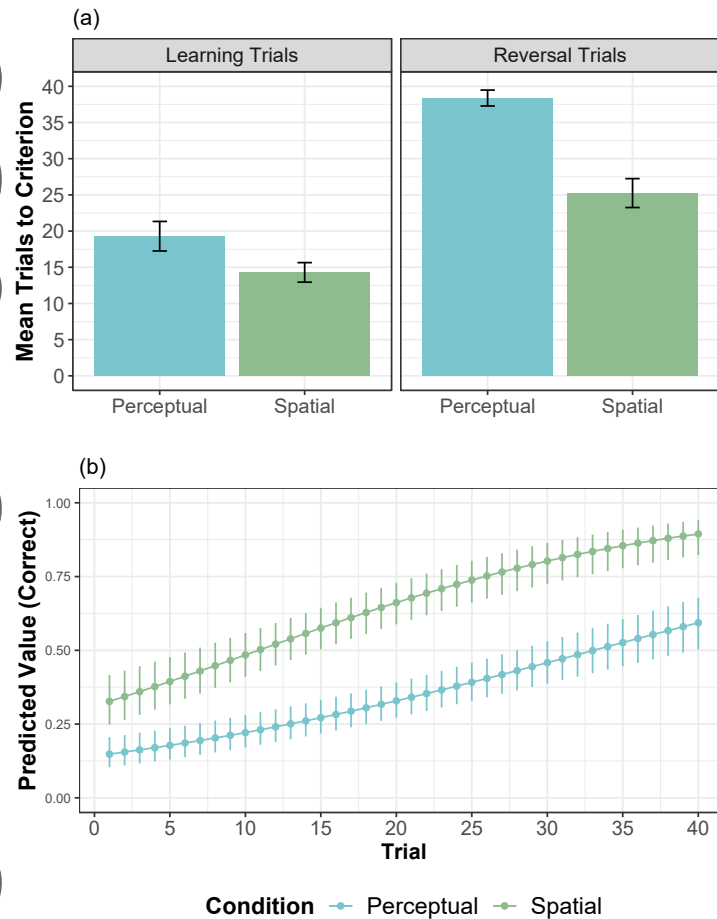


Figure 3. Performance in Study 2 by condition. (a) Mean number of trials to reach the 10 out of 12 passing criterion or trial limits in the learning and reversal phases by condition (perceptual and spatial). Note that larger numbers indicate worse performance, and the maximum number of reversal trials was 40 in this study. Error bars indicate SE. (b) Estimates of trial-by-trial performance in the reversal phase by condition. Error bars indicate 95% confidence interval estimates.

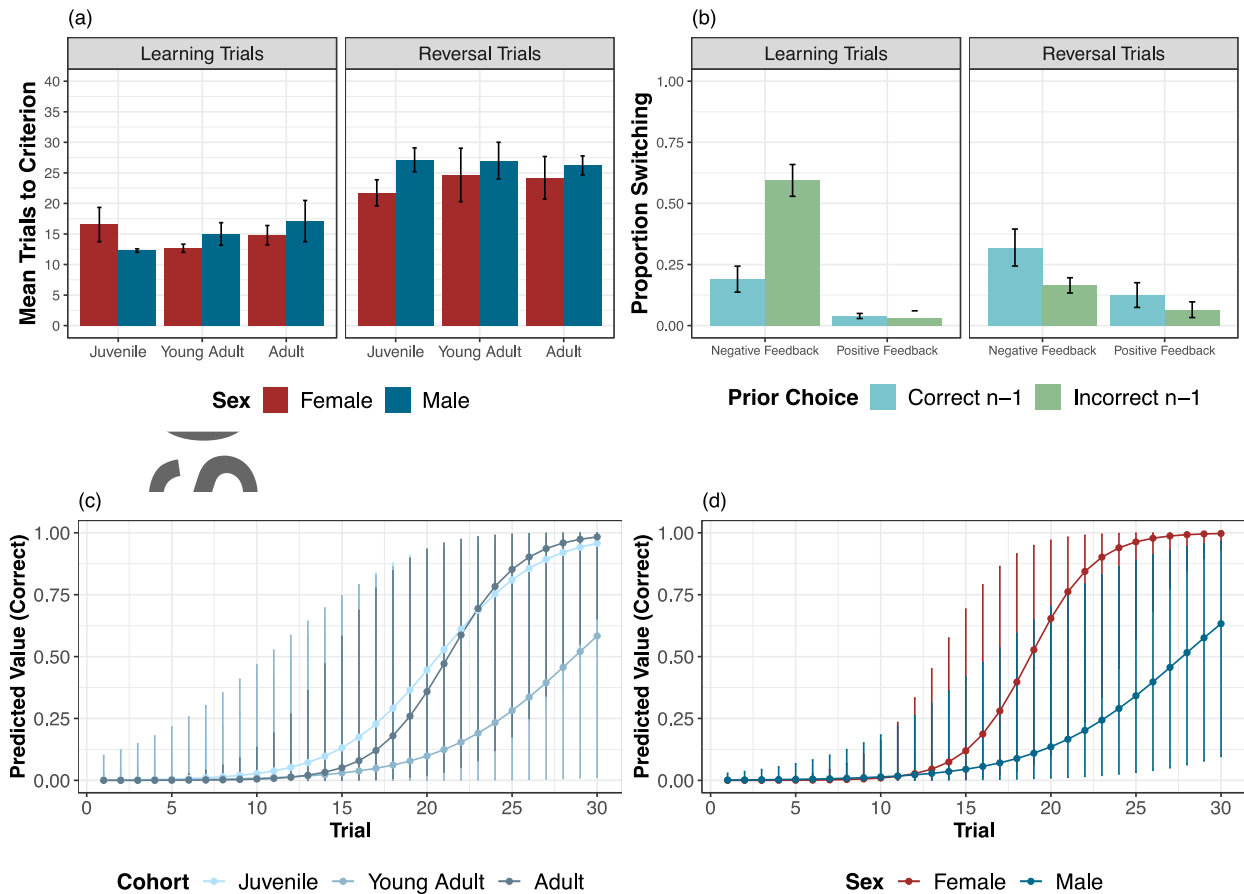


Figure 4. Performance in Study 3. (a) Mean number of trials to reach the 10 out of 12 passing criterion or trial limits in the learning and reversal phases by age cohort and sex. Note that larger numbers indicate worse performance, and the maximum possible number of reversal trials in this study was 30. Error bars indicate SE. (b) Proportion of next-trial switching after negative and positive feedback; error bars indicate SE. (c) Estimates of trial-by-trial performance in the reversal phase by age cohort and (d) by sex, from the best-fit model. Error bars indicate 95% confidence interval estimates.