*Supplementary Online Materials*

**The origins of cognitive flexibility in chimpanzees**

Averill Cantwell, Joshua W. Buckholtz, Rebeca Atencia, & Alexandra G Rosati

## Study 1: Development of shifting abilities

*Participants*

We tested a total of 82 semi-free-ranging chimpanzees from the Tchimpounga Chimpanzee Sanctuary in the Republic of Congo. As this population is primarily wild-born and exact birthdates are generally unknown, ages are estimated using initial assessments by sanctuary veterinarians at arrival (typically when the infants are between 1 and 3 years old); when possible, estimates are adjusted based on longitudinal measurements of weight and dental emergence data from those individuals, based on known patterns of ape development and validated through checks of sanctuary-born individuals with known birthdates (for more details on the age estimation method, see Rosati & Hare, 2012; Rosati 2019; Wobber et al., 2010; Wobber et al., 2014). These age estimates have accorded well with results from teeth and weight estimates in prior work examining cognitive and physiological development in this population (Wobber et al., 2010; Wobber et al., 2013). In analyses, we specifically compared individuals across three age cohorts based on chimpanzee developmental milestones (Goodall, 1983; Kawanaka, 1989): a *juvenile* cohort of chimpanzees up to 15 years (youngest age in this sample: 7 years); a *young adult* cohort of chimpanzees up to 20 years; and an *adult* cohort of chimpanzees 20 years and up (oldest age in this sample: 33 years). The breakdown of individuals is reported in Table S1.

| | Juveniles | Young adults | Adults |
|---|---|---|---|
| Females | 8 | 10 | 19 |
| Males | 13 | 15 | 17 |
| Total | 21 | 25 | 36 |

**Table S1: Chimpanzee characteristics in Study 1.** Individuals by age and sex.

*Main analyses*

As reported in the main text, our main analyses in this study examined overall performance in learning versus reversal, and the effects of age cohort and sex on trial-by-trial improvement in the first reversal phase. The supplement reports the model parameters for the best-fitting models described in the main text: the model comparing performance in the learning and first reversal phases (Table S2), and the model assessing trial-by-trial improvement in the first reversal phase (Table S3). In addition, we show trial-by-trial plots of performance in the first reversal phase (Figure S1) to complement the model effects shown in Figure 2 in the main text.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.07 | 0.005 | 15.44 | **<0.001** |
| Sex (reference: females) | -0.17 | 0.22 | -0.78 | 0.44 |
| Cohort (Linear) | 0.02 | 0.19 | 0.08 | 0.93 |
| Cohort (Quadratic) | -0.12 | 0.20 | -0.60 | 0.55 |
| Trial Type (reference: learning) | -1.74 | 0.10 | -17.81 | **<0.001** |

**Table S2: Predictors of overall performance in Study 1's learning versus first reversal phase.** The base model included *trial number*, *sex*, and *cohort* (as an ordered factor); *trial type* (learning or reversal) was then added to a second model, which improved fit. Reference values for model 2 predictors indicated in table.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Learning Count | -0.003 | 0.03 | -0.12 | 0.90 |
| Trial Number | 0.19 | 0.02 | 11.90 | **<0.001** |
| Sex (reference: females) | 0.63 | 0.44 | 1.44 | 0.15 |
| Cohort (Linear) | -0.42 | 0.38 | -1.11 | 0.27 |
| Cohort (Quadratic) | -0.13 | 0.37 | -0.36 | 0.72 |
| Trial Number*Cohort (Linear) | 0.05 | 0.01 | 4.25 | **<0.001** |
| Trial Number*Cohort (Quadratic) | 0.02 | 0.01 | 1.56 | 0.12 |
| Trial Number*Sex | -0.08 | 0.02 | -4.71 | **<0.001** |

**Table S3. Predictors of trial-by-trial performance in Study 1's first reversal phase.** The base model included trial number, sex, cohort, and learning count to index learning phase performance. Interactions between *trial number X cohort*, *trial X sex*, and the three-way interaction between *trial number X cohort X sex* were then added. The best-fitting model shown here included both two-way interactions. Reference values for predictors indicated in table.

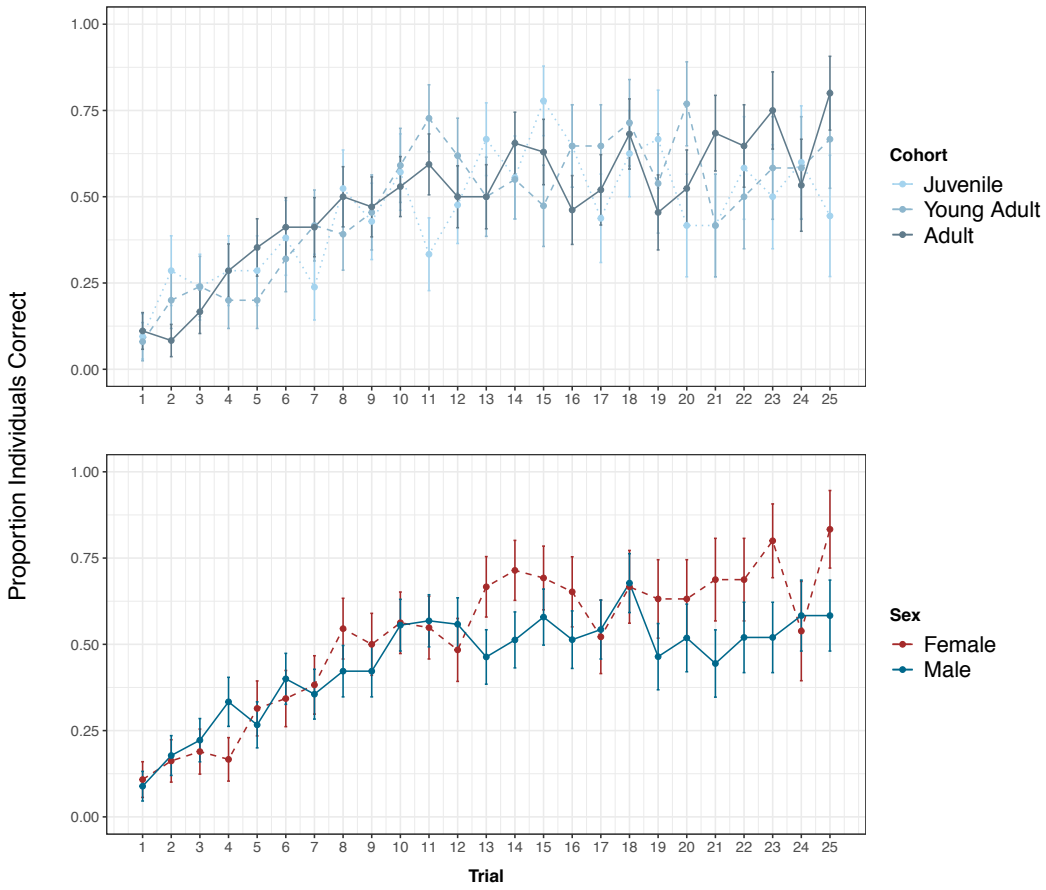


**Figure S1. Trial-by-trial improvement in Study 1's first reversal phase, split by age cohort and sex.** Raw data is plotted over the first 25 trials (more than 50% of chimpanzees met the passing criterion by this point).

*Supplemental analyses: second reversal phase*

In addition to the primary analyses focusing on only the learning and first reversal phase, we also examined performance in the second reversal phase. Importantly, chimpanzees did not progress to the second reversal if they reached the 75-trial limit in the first reversal, or if they took more than 40 trials to pass the first reversal. As a result, a large proportion of juvenile and young adult chimpanzees (19% of each of these cohorts, compared to 9% of the adult cohort) and males (21%, compared to 6% of females) did not progress to the second reversal. As reported in the main text, a comparison of overall performance across learning, the first reversal, and the second reversal revealed better performance in the learning phase compared to either reversal phase (see Table S4 for parameters).

We also examined trial-by-trial improvement in the second reversal. However, as many individuals did not reach this phase because of the pre-set trial limitations, and those who were excluded were predominantly younger and male, we focused primarily on the first reversal phase for this study. Here, we used the same analysis approach used to assess the first reversal; models also included *number of first reversal trials* to account for performance in that phase. In a second model, inclusion of *trial number X cohort* improved fit [LRT: $\chi^2 = 8.39$, $df = 2$, $p = 0.01$; model 1

AIC: 1263.0, model 2 AIC: 1258.6]; post-hoc comparisons indicated that young adults improved faster than adults (p = 0.02) and trended to outperform juveniles (p = 0.055). Adding *trial number X sex* or the three-way interaction did not further improve fit [model 3 LRT: $\chi^2 = 1.10$, $df = 1$, $p = 0.29$; AIC: 1259.6; model 4 LRT: $\chi^2 = 8.41$, $df = 5$, $p = 0.13$; AIC: 1260.4]. The second model received 47% of the weight in an AIC comparison; the third model received 29% ($\Delta$AIC = 0.9 between model 2 and model 3; see Table S5 for parameters from model 2). This indicates that the young adults that were able to make it to the second reversal were quicker to shift than adults.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.12 | 0.01 | 20.08 | **<0.001** |
| Sex (reference: females) | -0.09 | 0.18 | -0.49 | 0.62 |
| Cohort (Linear) | -0.06 | 0.15 | -0.42 | 0.67 |
| Cohort (Quadratic) | -0.25 | 0.17 | -1.53 | 0.13 |
| Trial Type R1 (reference: learning) | -1.55 | 0.11 | -14.55 | **<0.001** |
| Trial Type R2 (reference: learning) | -1.45 | 0.11 | -13.43 | **<0.001** |

**Table S4: Predictors of overall performance in the learning, first reversal phase, and second reversal phases in Study 1.** The base model included *trial number, sex*, and *cohort* (ordered factor); *trial type* (learning, reversal 1, reversal 2) was then added to test its importance. The best fitting-model shown here included trial type. Reference values for predictors indicated in table.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.25 | 0.02 | 14.25 | **<0.001** |
| Learning Count | -0.05 | 0.04 | -1.36 | 0.17 |
| Reversal 1 Count | -0.01 | 0.03 | -0.52 | 0.60 |
| Sex (reference: females) | -0.79 | 0.45 | -1.74 | 0.08 |
| Cohort (Linear) | 0.43 | 0.45 | 0.96 | 0.33 |
| Cohort (Quadratic) | 0.41 | 0.52 | 0.80 | 0.42 |
| Trial Number*Cohort (Linear) | -0.01 | 0.02 | -0.43 | 0.66 |
| Trial Number*Cohort (Quadratic) | -0.09 | 0.03 | -2.67 | **0.008** |

**Table S5. Predictors of trial-by-trial performance in second reversal phase of Study 1.** The base model included *trial number, sex, cohort, learning count*, and *first reversal count*. Interactions between *trial number X cohort*, *trial X sex*, and the three-way interaction were then added to test their importance. The best-fitting model included only the cohort interaction. Reference values indicated in table.

## Study 2: Spatial vs. perceptual reversal learning

*Participants*

We tested a subset of individuals who previously participated in Study 1 and had successfully completed both reversal phases in that study. The breakdown of the individuals included by these criteria are reported in Table S6.

| | Juveniles | Young adults | Adults |
|---|---|---|---|
| Females | 4 | 1 | 7 |
| Males | 5 | 4 | 3 |
| Total | 9 | 5 | 10 |

**Table S6: Chimpanzee characteristics in Study 2.** Individuals by age and sex.

*Main analyses*

As reported in the main text, our main analyses in this study compared performance between the spatial and perceptual conditions. The supplement reports the model parameters for the best-fitting models described in the main text, specifically those comparing performance between the spatial and perceptual conditions in learning and reversal (Table S7) and comparing trial-by-trial performance in the reversal phase (Table S8). In addition, we show trial-by-trial plots of performance in the reversal phase (Figure S2) to complement the model effects shown in Figure 3 in the main text.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.05 | 0.00 | 9.83 | **<0.001** |
| Sex (reference: females) | 0.01 | 0.16 | 0.06 | 0.95 |
| Cohort (Linear) | -0.20 | 0.12 | -1.60 | 0.11 |
| Cohort (Quadratic) | -0.14 | 0.16 | -0.84 | 0.40 |
| Trial Type (reference: learning) | -2.13 | 0.14 | -15.04 | **<0.001** |
| Condition (reference: perceptual) | 0.60 | 0.17 | 3.42 | **<0.001** |
| Trial Type*Condition | 0.68 | 0.21 | 3.27 | **0.001** |

**Table S7. Predictors of learning and reversal performance between the spatial and perceptual conditions in Study 2.** The base model included *trial number, sex*, and *cohort* (as an ordered factor); *trial type* (learning or reversal), *condition* (spatial or perceptual) and their interaction were then added to subsequent models. The best fitting-model shown here included these predictors. Reference values for predictors indicated in the table.

*Supplemental analyses: age and sex effects*

As an additional check of this data, we further tested for developmental or sex effects in the reversal phases of each condition, using the same basic approach as in Study 1. We used GLMMs to build a base model that included *subject* as a random effect, *trial number*, *sex*, and *cohort*, as well as *number of learning trials* to reach criterion. Adding *trial number X cohort* in a second model did not improve model fit in either condition [*spatial* LRT: $\chi^2 = 1.84$, $df = 2$, $p =$

0.40; model 1 AIC: 773.4, model 2 AIC: 775.6; *perceptual* LRT: $\chi^2 = 1.56$, $df = 2$, $p = 0.46$; model 1 AIC: 1083.6, model 2 AIC: 1086.1]. Adding *trial number X sex* also did not improve model fit in either condition [*spatial* LRT: $\chi^2 = 0.92$, $df = 1$, $p = 0.34$; model 3 AIC: 774.5; *perceptual* LRT: $\chi^2 = 0.24$, $df = 1$, $p = 0.62$; model 3 AIC: 1085.4; see Tables S9 and S10 for parameters from the best-fitting models]. Overall, we did not find developmental changes or a sex difference in this sample, likely due to the fact that the specific individuals tested in the study were selected based on their strong performance in Study 1.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Learning Count | -0.02 | 0.01 | -2.78 | **0.005** |
| Trial Number | 0.06 | 0.01 | 10.59 | **<0.001** |
| Sex (reference: females) | -0.02 | 0.26 | -0.10 | 0.92 |
| Cohort (Linear) | -0.21 | 0.20 | -1.09 | 0.27 |
| Cohort (Quadratic) | -0.0008 | 0.25 | -0.003 | 0.99 |
| Condition (reference: perceptual) | 1.32 | 0.13 | 10.10 | **<0.001** |

**Table S8. Predictors of trial-by-trial performance in reversal phase in Study 2.** The base model included *trial number, sex, cohort* (as an ordered factor), and *learning count* to index learning phase performance. *Condition* (spatial versus perceptual) and the interaction between *trial number X condition* were then added to test their importance. The best-fitting model shown here included only the main effect of condition. Reference values for predictors indicated in the table.

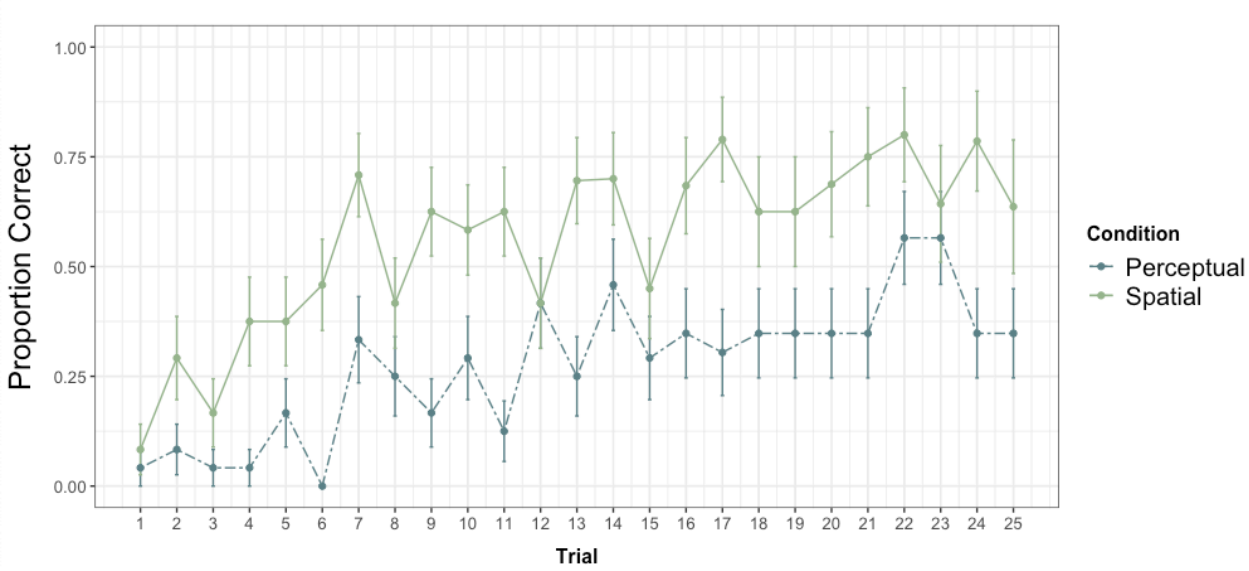


**Figure S2. Trial-by-trial improvement in the reversal phase in Study 2, split by condition.** Raw data is plotted over the first 25 trials (more than 50% of chimpanzees in the *spatial condition* met the passing criterion by this point).

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.09 | 0.01 | 7.70 | **<0.001** |
| Sex (reference: females) | -0.24 | 0.42 | -0.57 | 0.57 |
| Cohort (Linear) | -0.18 | 0.31 | -0.59 | 0.56 |
| Cohort (Quadratic) | 0.10 | 0.40 | 0.26 | 0.80 |
| Learning Count | -0.003 | 0.03 | -0.09 | 0.93 |

**Table S9. Predictors of trial-by-trial reversal performance in the spatial cue condition in Study 2.** The base model included *trial number, sex, cohort* (as an ordered factor), and *learning count*. Interactions between *trial number X cohort* and *trial number X sex* were then added. The best-fitting (base) model is shown here. Reference values for predictors indicated in the table.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.06 | 0.01 | 8.12 | **<0.001** |
| Sex (reference: females) | -0.04 | 0.33 | -0.12 | 0.91 |
| Cohort (Linear) | -0.30 | 0.25 | -1.20 | 0.23 |
| Cohort (Quadratic) | -0.25 | 0.32 | -0.78 | 0.43 |
| Learning Count | -0.02 | 0.02 | -1.25 | 0.21 |

**Table S10. Predictors of trial-by-trial reversal performance in the perceptual cue condition in Study 2.** The base model included *trial number, sex, cohort,* and *learning count*. Interactions between *trial number X cohort* and *trial number X sex* were then added to test their importance. The best-fitting (base) model is shown here. Reference values for predictors indicated in the table.

**Study 3: Probabilistic reversal learning**

*Participants*

This study comprised 40 individuals from the same populations as studies 1 and 2. This data was collected approximately 7 years prior to the other two studies; 22 individuals later participated in Study 1; 9 of these also participated in Study 2. The breakdown of the individuals included is reported in Table S11.

| | Juveniles | Young adults | Adults |
|---|---|---|---|
| Females | 11 | 3 | 5 |
| Males | 7 | 5 | 9 |
| Total | 18 | 8 | 14 |

**Table S11: Chimpanzee characteristics in Study 3.** Individuals by age and sex.

*Main analyses*

As reported in the main text, our main analyses in this study examined overall performance between the learning and reversal phases, and age cohort and sex effects on trial-by-trial

improvement in the reversal phase. The supplement reports the model parameters for the best-fitting models described in the main text: the model comparing performance across the learning and reversal phase (Table S12), and the model examining trial-by-trial performance in the reversal phase (Table S13). In addition, we show trial-by-trial plots of performance in the reversal phase (Figure S3) to complement the model effects shown in Figure 4 in the main text.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.08 | 0.01 | 9.19 | **<0.005** |
| Sex (reference: females) | -0.25 | 0.40 | -0.62 | 0.53 |
| Cohort (Linear) | -0.20 | 0.32 | -0.64 | 0.52 |
| Cohort (Quadratic) | 0.01 | 0.40 | -0.02 | 0.98 |
| Trial Type (reference: learning) | -3.05 | 0.17 | -17.91 | **<0.001** |

**Table S12: Predictors of overall performance in Study 3's learning versus reversal phase.** The base model included *trial number, sex*, and *cohort* (as an ordered factor); *trial type* (learning or reversal) was then added to a second model to test its importance. The best fitting-model shown here included trial type. Reference values for predictors indicated in the table.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.51 | 0.07 | 7.45 | **<0.001** |
| Learning Count | 0.18 | 0.13 | 1.37 | 0.17 |
| Sex (reference: females) | 3.24 | 2.27 | 1.42 | 0.15 |
| Cohort (Linear) | -2.14 | 1.74 | -1.23 | 0.22 |
| Cohort (Quadratic) | -0.90 | 2.19 | -0.41 | 0.68 |
| Trial Number*Cohort (Linear) | 0.09 | 0.04 | 2.13 | **0.03** |
| Trial Number*Cohort (Quadratic) | 0.12 | 0.04 | 2.74 | **<0.01** |
| Trial Number*Sex | -0.29 | 0.07 | -3.99 | **<0.001** |

**Table S13. Predictors of trial-by-trial performance in Study 3's reversal phase.** The base model included *trial number, sex, cohort*, and *learning count* to index learning phase performance. Interactions between *trial number X cohort*, *trial X sex*, and the three-way interaction between *trial number X cohort X sex* were then added to test their importance. The best-fitting model shown here included only the two-way interactions. Reference values for predictors indicated in the table.

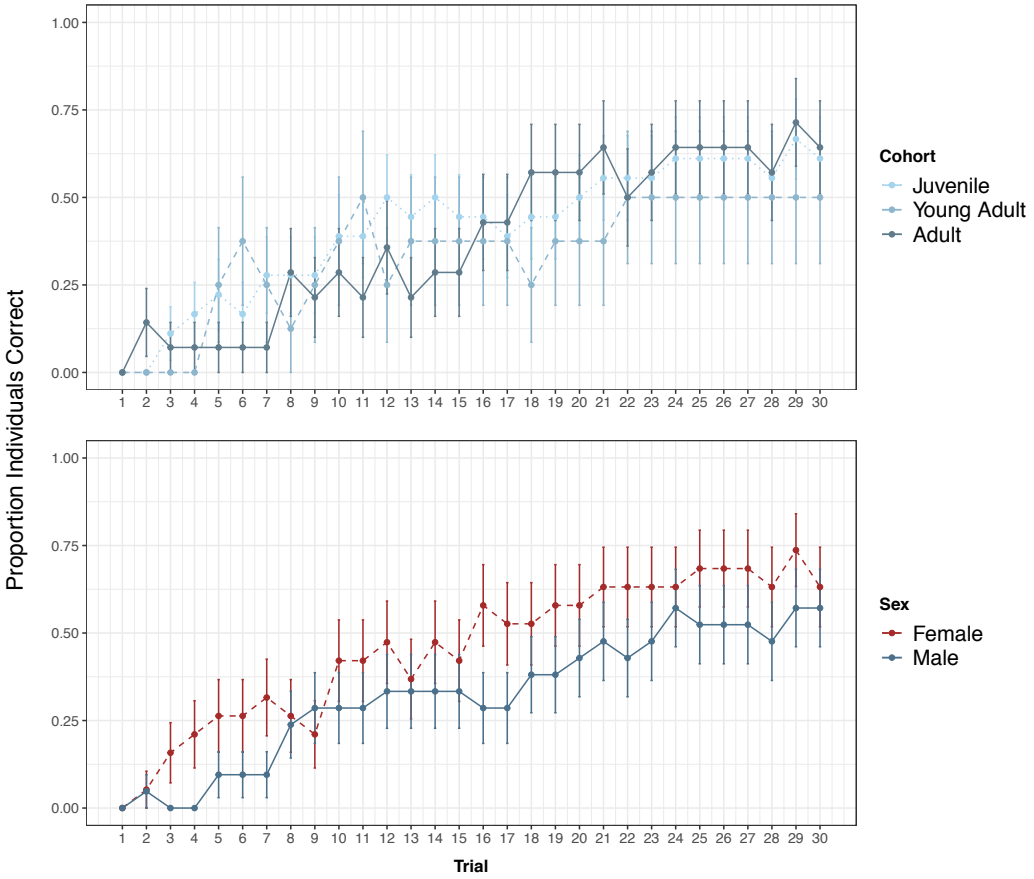


**Figure S3. Trial-by-trial improvement in the reversal phase in Study 3, split by age cohort and sex.** Raw data is plotted over the total 30 trials.

*Supplemental analyses: response to positive versus negative feedback*

We also examined the effects of receiving positive or negative feedback on the chimpanzees' subsequent choice. To analyze switching behavior, we constructed a base model predicting the response of *switching* with *subject* as a random effect, *trial number*, *cohort, sex,* and *trial type* (learning or reversal). Inclusion of *prior-trial feedback* (positive or negative outcome) improved model fit [LRT: $\chi^2$=60.36, *df*=1, *p*<0.0001; model 1 AIC=990.3, model 2 AIC=932.0]: chimpanzees were more likely to switch side after receiving negative feedback, and overall less likely to switch during the reversal phase. Adding *previous choice* (correct or incorrect) did not further improve model fit [LRT: $\chi^2$=0.002, *df*=1, *p*=0.96; model 3 AIC=934.0]. Adding a *trial type X feedback* interaction did improve model fit [LRT: $\chi^2$=10.28, *df*=1, *p*=0.001; model 4 AIC=925.7]: post-hoc tests showed that chimpanzees were more likely to switch based on negative feedback in learning compared to reversal (*p*<.0001). We then added a *trial type X previous choice* interaction to test if the effect of correct or incorrect previous choice differed in learning and reversal trials. This also improved model fit [LRT: $\chi^2$=15.57, *df*=1, *p*<0.0001; model 5 AIC=912.2]: post-hoc comparisons showed that chimpanzees were more likely to switch after the incorrect choice in the learning phase compared to the reversal phase (p<0.0001; see Table S14 for model parameters). In this model, the *trial type X feedback* interaction was no longer

significant. Finally, the three-way interaction between *trial type, feedback,* and *previous choice* did not improve fit [LRT: $\chi^2$=1.81, *df*=2, *p*=0.40; AIC=914.4]. An AIC comparison weighted model 5 highly (75%, ΔAIC=2.2). Overall, chimpanzees switched less after the incorrect choice in the reversal phase compared to the learning phase, reflecting the high rates of perseveration we observed.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | -0.01 | 0.01 | -0.87 | 0.39 |
| Sex (reference: females) | 0.12 | 0.27 | 0.43 | 0.67 |
| Cohort (Linear) | -0.09 | 0.21 | -0.42 | 0.68 |
| Cohort (Quadratic) | 0.10 | 0.27 | 0.37 | 0.71 |
| Prev. Choice (reference: correct) | 0.87 | 0.33 | 2.67 | **<0.01** |
| Trial Type (reference: learning) | 0.29 | 0.42 | 0.70 | 0.49 |
| Prev. Feedback (reference: negative) | -1.72 | 0.35 | -4.87 | **<0.001** |
| Trial Type*Prev. Feedback | 0.38 | 0.49 | 0.77 | 0.44 |
| Trial Type*Prev. Choice | -1.90 | 0.47 | -4.00 | **<0.001** |

**Table S14. Predictors of switching behavior in Study 3.** The base model included *trial number, sex, cohort*, and *trial type* (learning or reversal). *Prior trial feedback, prior trial choice*, and interactions between *trial type X prior trial choice* and *trial type X feedback* were then added to test their importance. The best-fitting model is shown here included *previous choice, previous outcome, trial type X previous feedback,* and *trial type X previous* choice. Reference values for predictors indicated in the table.

*Supplementary analyses: responses across all reversal trials*

The analyses reported in the main text use the number of trials that successful chimpanzees required to meet the 10 out of 12 passing criterion, mirroring the methods and analyses implemented in Study 1 and 2. However, in this particular study all chimpanzees actually completed 30 trials regardless of performance, so as an additional check we then also ran some the same analyses reported above using all data to check if this impacted results.

First, we analyzed trial-by-trial performance in the reversal phase. A base model included *subject* as a random effect, *trial number*, *sex*, *cohort*, and *number of learning trials.* Adding the interaction *trial number X cohort* trended to improve model fit [LRT: $\chi^2$=4.86, *df* =2, *p*=0.09; model 1 AIC: 728.1, model 2 AIC: 727.3]. Adding the interaction *trial number X sex* improved fit [LRT: $\chi^2$=6.02, *df*=1, *p*=0.01; AIC: 723.3; see Table S15 for parameters from this model]. Post-hoc comparisons showed that females were faster to improve compared to males, and that adults were faster than juveniles (p<0.05 for all significant comparisons); adults also trended to shift faster than young adults (p=0.08). Adding the three-way interaction between *trial number, cohort,* and *sex* did not improve model fit [LRT: $\chi^2$=8.58, *df*=4, *p*=0.07; AIC: 722.9]. By AIC comparison, however, this final model received slightly more weight (50%) compared to the previous model (41%, ΔAIC = 0.4). Post-hoc comparisons of this three-way interaction indicated that young adult males were specifically slower to improve in comparison to adult females (p<0.05). Overall our findings using all 30 trials for each chimpanzee generally align with the main analyses looking

only at the number of trials to successfully meet passing criterion or reach the maximum number of trials, which found similar developmental and sex patterns.

| **Predictor** | **Estimate** | **S.E.** | **Z** | ***p* value** |
|---|---|---|---|---|
| Trial Number | 0.31 | 0.03 | 9.19 | **<0.001** |
| Learning Count | 0.14 | 0.10 | 1.50 | 0.13 |
| Sex (reference: females) | 0.51 | 1.54 | 0.33 | 0.74 |
| Cohort (Linear) | -1.59 | 1.24 | -1.28 | 0.20 |
| Cohort (Quadratic) | -0.17 | 1.53 | -0.11 | 0.91 |
| Trial Number*Cohort (Linear) | 0.08 | 0.03 | 2.55 | **0.01** |
| Trial Number*Cohort (Quadratic) | 0.05 | 0.04 | 1.25 | 0.21 |
| Trial Number*Sex | -0.10 | 0.04 | -2.44 | **0.01** |

**Table S15. Predictors of trial-by-trial performance in Study 3's reversal phase using all 30 trials.** The base model included *trial number, sex, cohort*, and *learning count* to index learning phase performance. Interactions between *trial number X cohort, trial X sex*, and the three-way interaction were then added to test their importance. The best-fitting model shown here included only the two-way interactions. Reference values for predictors indicated in the table.

Second, we analyzed switching behavior using all 30 trials. We constructed a base model predicting the response of *switching* with *subject* as a random effect, *trial number*, *cohort, sex*, and *trial type* (learning or reversal). Inclusion of *prior-trial feedback* (positive or negative outcome) improved model fit [LRT: $\chi^2$=67.9, *df*=1, *p*<0.0001; model 1 AIC: 1108.3, model 2 AIC: 1042.4]: chimpanzees were more likely to switch side after receiving negative feedback and overall less likely to switch in the reversal phase. Adding *previous choice* (correct or incorrect option) did not further improve model fit [LRT: $\chi^2$=2.32, *df*=1, *p*=0.13; model 3 AIC: 1042.1]. We then added a *trial type X feedback* interaction, to test whether the effects of positive and negative feedback on switching behavior depended on whether chimpanzees were in the learning or reversal phase. This improved model fit [LRT: $\chi^2$=10.16, *df*=1, *p*=0.001; model 4 AIC: 1034.0]; post-hoc tests showed that chimpanzees were more likely to switch based on negative feedback in learning compared to reversal (p<.0001). We then added the *trial type X previous choice* interaction to test if an effect of correct or incorrect prior choice differed in learning and reversal trials. This also improved model fit [LRT: $\chi^2$=5.25, *df*=1, *p*=0.02; model 5 AIC: 1030.8]: post-hoc comparisons showed that chimpanzees were more likely to switch after the incorrect choice in the learning phase compared to the reversal phase (p<0.001, see Table S16 for model parameters). Finally, we tested the three-way interaction between *trial type, feedback,* and *previous choice,* which did not further improve fit [LRT: $\chi^2$=2.12, *df*=2, *p*=0.35; model 6 AIC: 1032.7]. An AIC comparison weighted the model with the separate interactions (model 5) highly (63%, ΔAIC = 1.9). Overall, chimpanzees were less likely to change their response after negative feedback or incorrect choices in the reversal phase compared to the learning phase. This aligns with the main results looking only at performance until meeting the passing criterion or reaching the maximum in the reversal phase.

| Predictor | Estimate | S.E. | Z | *p* value |
|---|---|---|---|---|
| Trial Number | -0.00 | 0.01 | -0.40 | 0.69 |
| Sex (reference: females) | 0.09 | 0.31 | 0.28 | 0.78 |
| Cohort (Linear) | -0.10 | 0.24 | -0.41 | 0.68 |
| Cohort (Quadratic) | 0.21 | 0.30 | 0.68 | 0.49 |
| Prev. Choice (reference: correct) | 0.87 | 0.33 | 2.63 | **<0.01** |
| Trial Type (reference: learning) | -0.41 | 0.39 | -1.04 | 0.30 |
| Prev. Feedback (reference: negative) | -1.77 | 0.36 | -4.95 | **<0.001** |
| Trial Type*Prev. Feedback | 0.75 | 0.45 | 1.64 | 0.10 |
| Trial Type*Prev. Choice | -1.05 | 0.45 | -2.32 | **0.02** |

**Table S16. Predictors of switching behavior in Study 3 across all 30 trials.** The base model included *trial number, sex, cohort*, and *trial type* (learning or reversal). Prior trial feedback, prior trial choice, and interactions between *trial type X prior trial choice* and *trial type X feedback* were then added to test their importance. The best-fitting model is shown here included *previous choice, previous outcome*, *trial type X previous feedback* and *trial type X previous choice*. Reference values for predictors indicated in the table.

**References**

Goodall, J. (1963). Population dynamics during a 15 year period in one community of free-living Chimpanzees in the Gombe National Park, Tanzania. *Zeritschrift fuer tierpsychologie, 61*(1), 1-60.

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.

Rosati, A. G., & Hare, B. (2012). Chimpanzees and bonobos exhibit divergent spatial memory development. *Developmental science*, *15*(6), 840-853.

Rosati, A. G. (2019). Heterochrony in chimpanzee and bonobo spatial memory development. *American journal of physical anthropology*, *169*(2), 302-321.

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., Lipson, S., Wrangham, R., & Ellison, P. (2013). Different ontogenetic patterns of testosterone production reflect divergent male reproductive strategies in chimpanzees and bonobos. *Physiology & behavior*, *116*, 44-53.

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.