

# Bayesian reinforcement learning models reveal how great-tailed grackles improve their behavioral flexibility in serial reversal learning experiments.

Lukas D<sup>1\*</sup>    McCune KB<sup>2</sup>    Blaisdell AP<sup>3</sup>    Johnson-Ulrich Z<sup>2</sup>  
MacPherson M<sup>2</sup>    Seitz B<sup>3</sup>    Sevchik A<sup>4</sup>    Logan CJ<sup>1,2</sup>

2024-07-22

**Affiliations:** 1) Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, 2) University of California Santa Barbara, USA, 3) University of California Los Angeles, USA, 4) Arizona State University, Tempe, AZ USA. \*Corresponding author: [dieter\\_lukas@eva.mpg.de](mailto:dieter_lukas@eva.mpg.de)

**This is the third of three post-study manuscripts from the preregistration that was pre-study peer reviewed and received an In Principle Recommendation on 26 Mar 2019 by:**

Aurélie Coulon (2019) Can context changes improve behavioral flexibility? Towards a better understanding of species adaptability to environmental changes. *Peer Community in Ecology*, 100019. [10.24072/pci.ecology.100019](https://doi.org/10.24072/pci.ecology.100019). Reviewers: Maxime Dahirel and Andrea Griffin

**Preregistration:** [html](#), [pdf](#), [rmd](#)

**Post-study manuscript** (submitted to PCI Ecology for post-study peer review on 3 Jan 2022, revised and resubmitted Feb 2024 and July 2024): [preprint at EcoEvoRxiv](#), [rmd with code at github](#)

## Abstract

Environments can change suddenly and unpredictably and animals might benefit from being able to flexibly adapt their behavior through learning new associations. Serial (repeated) reversal learning experiments have long been used to investigate differences in behavioral flexibility among individuals and species. In these experiments, individuals initially learn that a reward is associated with a specific cue before the reward is reversed back and forth between cues, forcing individuals to reverse their learned associations. Cues are reliably associated with a reward, but the association between the reward and the cue frequently changes. Here, we apply and expand newly developed Bayesian reinforcement learning models to gain additional insights into how individuals might dynamically modulate their behavioral flexibility if they experience serial reversals. We derive mathematical predictions that, during serial reversal learning experiments, individuals will gain the most rewards if they 1) increase their *rate of updating associations* between cues and the reward to quickly change to a new option after a reversal, and 2) decrease their *sensitivity* to their learned association to explore the alternative option after a reversal. We reanalyzed reversal learning data from 19 wild-caught great-tailed grackles (*Quiscalus mexicanus*), eight of whom participated in serial reversal learning experiment, and found that these predictions were supported. Their estimated association-updating rate was more than twice as high at the end of the serial reversal learning experiment than at the beginning, and their estimated sensitivities to their learned associations declined by about a third. The changes in behavioral flexibility that grackles showed in their experience of the serial reversals also influenced their behavior in a subsequent experiment, where individuals with more extreme rates or sensitivities solved more options on a multi-option puzzle box. Our findings offer new insights into how individuals react to uncertainty and changes in their

39 environment, in particular, showing how they can modulate their behavioral flexibility in response to their  
40 past experiences.

## 41 Introduction

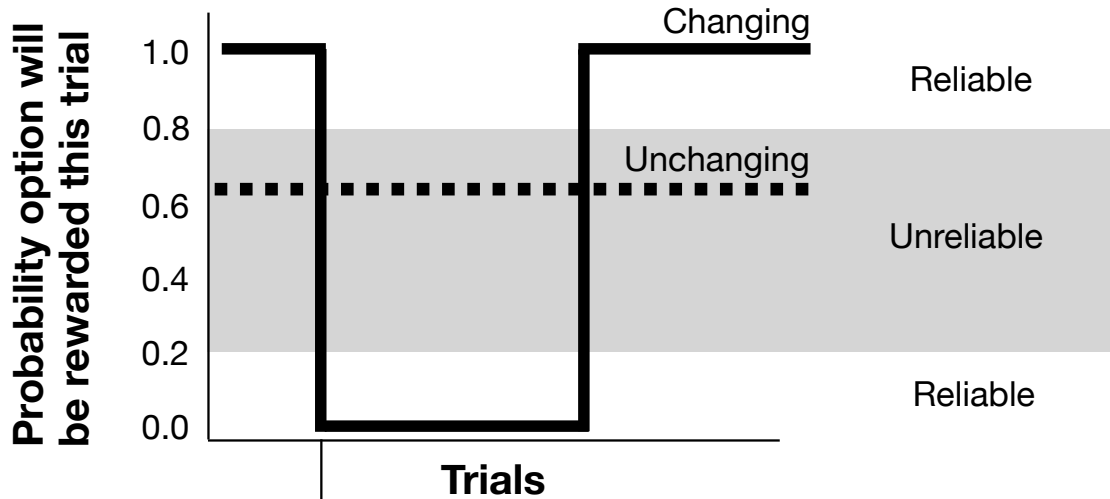
42 Most animals live in environments that undergo changes that can affect key components of their lives, such  
43 as where to find food or which areas are safe. Accordingly, individuals that cannot react to these changes  
44 should have reduced survival and/or reproductive success (Boyce et al., 2006; Starrfelt & Kokko, 2012). One  
45 of the ways animals react to changes is through behavioral flexibility, the ability to change behavior when  
46 circumstances change (Shettleworth, 2010). The level of behavioral flexibility present in a given species is  
47 often assumed to have been shaped by selection, with past levels of change in the environment determining  
48 how well species might be able to cope with more rapidly changing (Sih, 2013) or novel environments (Sol  
49 et al., 2002). However, in another conception, behavioral flexibility is itself plastic (Wright et al., 2010).  
50 Behavioral flexibility arises because individuals update their information about the environment through  
51 personal experience and make that information available to other cognitive processes (Mikhalevich et al.,  
52 2017). Such modulation of behavioral flexibility is presumably relevant if the rate and extent of environmental  
53 change is variable and unpredictable (Donaldson-Matasci et al., 2013; Tello-Ramos et al., 2019). We are still  
54 limited in our understanding of when and how individuals might react to their experiences of environmental  
55 change.

56 Evidence that animals can change their behavioral flexibility based on their recent experience comes from  
57 serial reversal learning experiments. Serial reversal learning experiments have long been used to understand  
58 how individuals keep track of biologically important associations in changing environments (Dufort et al.,  
59 1954; Mackintosh et al., 1968; Bitterman, 1975). In these experiments, individuals are presented with multi-  
60 ple options associated with cues, such as different colors or locations, that differ in their reward. Individuals  
61 can repeatedly choose among the options to learn the associations between rewards and cues. After they show  
62 a clear preference for the most rewarded option, the rewards are reversed across cues, and individuals are  
63 observed to see how quickly they learn the changed associations. When they have reversed their preference,  
64 the reward is changed back to the other option, until the individual reverses their preference again, and these  
65 reversals continue in a process called serial reversals. Their performance during the reversal task is taken as  
66 a measure of their behavioral flexibility, with the more flexible individuals being those that need fewer trials  
67 to consistently choose the rewarded option after a reversal (Bond et al., 2007). While the primary focus  
68 of these serial reversal learning experiments has been to measure differences in behavioral flexibility across  
69 individuals and species (Lea et al., 2020), several of these experiments show that behavioral flexibility is not  
70 a fixed trait, but that individuals can improve their performance if they experience repeated reversals (Bond  
71 et al., 2007; Liu et al., 2016; Cauchoix et al., 2017). Here, we investigate how individuals might change  
72 their behavioral flexibility during serial reversal learning experiments to better understand what cognitive  
73 processes could lead to the observed differences and adjustments in behavioral flexibility (Izquierdo et al.,  
74 2017; Danwitz et al., 2022).

75 We recently found that great-tailed grackles (*Quiscalus mexicanus*; hereafter grackles) can be trained to  
76 improve how quickly they learn to change associations in a serial reversal learning experiment (Logan et al.,  
77 2023a). After training birds to search for food in a yellow tube, the reversal learning experiment consisted  
78 of presenting birds with a light gray and a dark gray tube, only one of which contained a reward. After  
79 individuals chose one of the tubes, thus experiencing whether this color was rewarded or not, the experiment  
80 was reset, with the reward being in the same colored tube as before. Once an individual chose the rewarded  
81 color more than expected by chance (passing criterion of choosing correctly in at least 17 out of the last 20  
82 trials, which represents a significant association according to the chi-square test), the reward was switched to  
83 the other color. Again, individuals made choices until they chose the now rewarded tube above the passing  
84 criterion. For one set of individuals, the trained group, we repeated the reversal of rewards from one color  
85 to the other until the birds reached the serial reversal passing criterion of forming a preference in 50 trials  
86 or less in two consecutive reversals. The median number of trials birds in this trained group needed to reach  
87 the passing criterion during their first reversal was 75, which improved to 40 trials in their final reversal.  
88 Importantly, we found that, in comparison to a control group who only experienced a single reversal, trained

89 grackles who experienced serial reversals also showed increased behavioral flexibility and innovativeness in  
90 other contexts. In particular, trained grackles performed better on multi-option puzzle boxes than control  
91 grackles, being faster to switch to a new access option on a box if the previous option was closed, and they  
92 solved more of the available access options (Logan et al., 2023a). This indicates that individuals did not  
93 just learn an abstract rule about the serial reversal learning experiment, but rather changed their overall  
94 behavioral flexibility in response to their experience. To understand these changes in behavioral flexibility,  
95 we need approaches that can reflect how individuals might update their cognitive processes based on their  
96 experience.

97 Previous analyses of serial reversal learning experiments were limited in understanding the potential changes  
98 in behavioral flexibility because they focused on summaries of the choices that individuals make (e.g. Bond  
99 et al., 2007). These approaches are more descriptive, making it difficult to link flexibility differences to  
100 specific processes and to predict how variation in behavior might transfer to other tasks. While there have  
101 been attempts to identify potential rules that individuals might learn during serial reversal learning (Spence,  
102 1936; Warren, 1965a; Warren, 1965b; Minh Le et al., 2023), these rules were often about abstract switches  
103 to extreme behaviors (e.g. win-stay / lose-shift) and therefore could not account for the full variation of  
104 behavior. A number of theoretical models have recently been developed that appear to reflect the potential  
105 cognitive processes individuals seem to rely on when making choices in reversal learning experiments (for a  
106 recent review see, for example, Frömer & Nassar, 2023). These theoretical models deconstruct the behavior  
107 of individuals in a reversal learning task into two primary parameters (Camerer & Hua Ho, 1999; Chow et al.,  
108 2015; Izquierdo et al., 2017; Bartolo & Averbeck, 2020). Importantly, in the Bayesian reinforcement learning  
109 models there are now also statistical approaches to infer these underlying parameters from the behavior of  
110 individuals (Camerer & Hua Ho, 1999; Lloyd & Leslie, 2013). The first process reflects the *rate of updating*  
111 *associations* (which we refer to hereafter as  $\phi$ , the Greek letter phi), or how quickly individuals learn about  
112 the associations between the cues and potential rewards (or dangers). In the reinforcement learning models,  
113 this rate is reflected by the Rescorla-Wagner rule (Rescorla & Wagner, 1972). The rate weights the most  
114 recent information proportionally to the previously accumulated information for that cue (as a proportion,  
115 the rate can range between 0 and 1; for details on the calculations see the section on the reinforcement  
116 learning model in the Methods). Individuals are expected to show different rates in different environments,  
117 particularly in response to the reliability of the cues (Figure 1). Lower updating rates are expected when  
118 associations are not perfect such that a single absence of a reward might be an error rather than indicating  
119 a new association. Higher updating rates are expected when associations are reliable such that individuals  
120 should update their associations quickly when they encounter new information (Dunlap & Stephens, 2009;  
121 Breen & Deffner, 2023). The second process, the *sensitivity to their learned associations* (which we hereafter  
122 refer to as  $\lambda$ , the Greek letter lambda) reflects how individuals, when presented with a set of cues, might  
123 decide between these alternative options based on their learned associations of the cues. In the reinforcement  
124 learning model, the sensitivity to learned associations modifies the relative difference in learned rewards to  
125 generate the probabilities of choosing either option (Daw et al., 2006; Agrawal & Goyal, 2012; Danwitz et  
126 al., 2022). A value of zero means individuals do not pay attention to their learned associations, but choose  
127 randomly, whereas increasingly larger values mean that individuals show biases in choice as soon as there  
128 are small differences in their learned associations. Individuals with larger sensitivities will quickly prefer  
129 the option that previously gave them the highest reward (or the lowest danger), while individuals with  
130 lower sensitivities will continue to explore alternative options. Sensitivities are expected to reflect the rate  
131 of change in the environment (Figure 1), with larger sensitivities occurring when environments are static  
132 such that individuals start to exploit any differences they recognise as soon as possible. Lower sensitivities  
133 are expected when changes are frequent, such that individuals continue to explore alternative options when  
134 conditions change (Daw et al., 2006; Breen & Deffner, 2023).



What to do in the trials after X?

- Reliable: switch to alternate (large  $\phi$ )
- Changing: sometimes explore (small  $\lambda$ )
- ■ ■ ■ Unreliable: stay with current (small  $\phi$ )
- ■ ■ ■ Unchanging: always exploit (large  $\lambda$ )

135

136 **Figure 1.** Individuals are expected to update their associations and make decisions differently depending on  
 137 the environment they experience. In serial reversal learning experiments, associations are reliable, such that  
 138 if an option is associated with a reward, it is rewarded during every trial (white background). However, the  
 139 associations between options and the rewards change across trials (solid line). In these reliable, but changing  
 140 conditions, individuals are expected to gain the most rewards if they update their associations quickly (large  
 141  $\phi$ ) to switch away from an option if it is no longer being rewarded, but to have small sensitivities to their  
 142 learned associations to continue to explore all options to check if associations have changed again (small  
 143  $\lambda$ ). In contrast, in unchanging and unreliable conditions, the probability that an option is rewarded stays  
 144 constant across trials (dotted lines), but is closer to 50% (gray background). In these conditions, individuals  
 145 are expected to gain the most rewards if they build their associations by averaging information across many  
 146 trials (small  $\phi$ ), and have high sensitivities to learned associations to exploit the option with the highest  
 147 association (large  $\lambda$ ). Grackle picture credit (CC BY 4.0): Dieter Lukas.

148 Here, we applied and modified the Bayesian reinforcement learning models to data from our grackle research  
 149 on behavioral flexibility to assess if and how the cognitive processes might have changed as individuals  
 150 experienced the serial reversal learning experiment. We previously found that the model can predict the  
 151 performance of grackles in a reversal learning task with a single reversal of a color preference (Blaisdell et al.,  
 152 2021). Grackles experiencing the serial reversal learning experiment are expected to infer that associations  
 153 can frequently change but that, before and after a change, cues reliably indicate whether a reward is present  
 154 or not. Based on the theoretical models, we predict that individuals increase their association-updating rate  
 155 because cues are highly reliable, such that they can change their associations as soon as there is a change  
 156 in the reward (Dunlap & Stephens, 2009; Breen & Deffner, 2023). In addition, we predict that individuals  
 157 reduce their sensitivity to the learned associations, because the option that is rewarded reverses frequently,  
 158 requiring individuals to explore alternative options (Neftci & Averbeck, 2019; Leimar et al., 2024). Given  
 159 that reversals in the associations are not very frequent, we also expect some variation in individuals in  
 160 whether they switch to the newly rewarded option because they find the reward quickly through continued

161 exploration (somewhat lower  $\lambda$  and higher  $\phi$ ) or because they quickly move away from the option that is no  
162 longer rewarded (somewhat higher  $\lambda$  and lower  $\phi$ ). To assess these predictions, we addressed the following six  
163 research questions. With the first research question, we determined the feasibility and validity of our approach  
164 using simulations. As far as we were aware, Bayesian reinforcement learning models had not been used to  
165 investigate temporal changes in behavior. We therefore used simulations as a proof-of-concept assessment  
166 to show their sensitivity and ability to answer our questions. With the second research question, we derive  
167 mathematically specific predictions about the role of  $\phi$  and  $\lambda$  in the serial reversal learning experiment. With  
168 the other four questions, we analyzed the grackle data to determine how the association-updating rate and  
169 the sensitivity to learned associations reflect the variation and changes in behavioral flexibility in grackles.

170 **1) Are the Bayesian reinforcement learning models sufficiently sensitive to detect changes that**  
171 **occur across the limited number of serial reversals that individuals participated in?**

172 We used agent-based simulations to answer this question, where simulated individuals made choices based  
173 on assigned  $\phi$  and  $\lambda$  values. We determined how to apply the Bayesian reinforcement learning models to  
174 recover the assigned values from the choices in each trial. Previous applications of the Bayesian reinforcement  
175 learning models always combined the full sample of observations, so it is not clear whether these models are  
176 sufficiently sensitive to detect the changes over time that we are interested in. Two problems arise when  
177 trying to infer the underlying processes from a limited number of trials. The stochasticity in which option  
178 an individual chooses based on a given set of associations introduces differences in the set of choices across  
179 trials even among individuals with the same  $\phi$  and  $\lambda$  values. On the flip-side, because of the probabilistic  
180 decisions, a given series of specific choices during a short number of trials can occur even if individuals have  
181 different  $\phi$  and  $\lambda$  values. We varied the number of trials we analyzed to determine how many trials per  
182 individual are necessary to recover the assigned  $\phi$  and  $\lambda$  values in light of this noise.

183 **2) Is a high rate of association-updating ( $\phi$ ) and a low sensitivity to learned associations ( $\lambda$ )**  
184 **best to reduce errors in the serial reversal learning experiment?**

185 We used analytical approaches to systematically vary  $\phi$  and  $\lambda$  to determine how the interaction of the two  
186 processes shapes the behavior of individuals throughout the serial reversal learning experiment. Previous  
187 studies made general predictions about the role of  $\phi$  and  $\lambda$  in different environments (Dunlap & Stephens,  
188 2009; Breen & Deffner, 2023). We assessed here whether, under the specific conditions in the serial reversal  
189 experiments, where information is reliable and changes occur frequently, the best approach for individuals  
190 is to show high  $\phi$  and low  $\lambda$ .

191 **3) Which of the two parameters  $\phi$  or  $\lambda$  explains more of the variation in the serial reversal**  
192 **learning experiment performance of the tested grackles?**

193 Across both the trained (experienced serial reversals) and control (experienced a single reversal) grackles,  
194 we assessed whether variation in the number of trials an individual needed to reach the criterion in a given  
195 reversal is better explained by their inferred association-updating rate or by their inferred sensitivity to  
196 learned associations.

197 **4) Do the grackles who improved their performance through the serial reversal learning ex-**  
198 **periment show the predicted changes in  $\phi$  and  $\lambda$ ?**

199 If individuals learn the contingencies of the serial reversal experiment, they should reduce their sensitivity to  
200 learned associations  $\lambda$  to explore the alternative option when rewards change, and increase their association-  
201 updating rate  $\phi$  to quickly exploit the new reliably rewarded option.

202 **5) Are some individuals better than others at adapting to the serial reversals?**

203 In previous work, we found that there are individual differences that persist throughout the experiment,  
204 with individuals who required fewer trials to solve the initial reversal also requiring fewer trials in the final  
205 reversal after their training (McCune et al., 2023). We could expect that these individual differences are  
206 guided by consistency in how individuals solve the reversal learning paradigm, meaning they are reflected  
207 in individual consistency in  $\phi$  and  $\lambda$  that persist through the serial reversals. In addition, it is not clear  
208 whether some grackles change their behavior more than others. For example, it could be that individuals  
209 who have a higher association-updating rate  $\phi$  at the beginning of the experiment might also be better able  
210 to quickly change their behavior to match the particular conditions of the serial reversal learning experiment.  
211 Therefore, we also analyzed whether the  $\phi$  and  $\lambda$  values of individuals at the beginning predict how much

212 they changed throughout the serial reversal learning experiment.

213 **6) Can the  $\phi$  or  $\lambda$  from the performance of the grackles during their final reversal predict**  
214 **variation in the performance on the multi-option puzzle boxes?**

215 Grackles would be expected to solve more options on the multi-option puzzle boxes if they quickly update  
216 their previously learned associations when a previous option becomes unavailable (high  $\phi$ ). Given that,  
217 in the puzzle box experiment, individuals only receive a reward at any given option a few times, instead  
218 of repeatedly as in the reversal learning task, we predict that those individuals who are less sensitive to  
219 previously learned associations and instead continue to explore alternative options (low  $\lambda$ ) can also gain  
220 more rewards.

## 221 Materials and Methods

### 222 Data

223 For question 1, we re-analyzed data we previously simulated for power analyses to estimate sample sizes for  
224 population comparisons (Logan et al., 2023c). In brief, we simulated choices in an initial association learning  
225 and single reversal experiment for a total of 640 individuals. The  $\phi$  and  $\lambda$  values for each individual were  
226 drawn from a distribution representing one of 32 populations, with different mean  $\phi$  (8 different means) and  
227 mean  $\lambda$  (4 different values) values for each population (32 populations is the combination of each  $\phi$  and  $\lambda$ ).  
228 We simulated 20 individuals in each of the 32 populations. The range for the  $\phi$  and  $\lambda$  values assigned to  
229 the artificial individuals in the simulations were based on the previous analysis of single reversal data from  
230 grackles in a different population (Santa Barbara, California, USA) (Blaisdell et al., 2021) to reflect the  
231 likely expected behavior. Based on their assigned  $\phi$  and  $\lambda$  values, each individual was simulated to pass first  
232 through the initial association learning phase and, after they reached criterion (chose the correct option 17  
233 out of the last 20 times), the rewarded option switched and simulated individuals went through the reversal  
234 learning phase until they again reached criterion. Each choice that each individual made was simulated  
235 consecutively. Choices during trials were based on the associations that individuals formed between each  
236 option and the reward based on their experience. The first choice a simulated individual made in the initial  
237 association learning was random because we assumed individuals had no information about the rewards and  
238 therefore set the initial attractions to both options to be equally low. Based on their choices, individuals  
239 updated their internal associations with the two options based on their individual learning rate. We excluded  
240 simulated individuals from further analyses if they did not reach criterion either during the initial association  
241 or the reversal within 300 trials, the maximum that was also set for the experiments with the grackles. For  
242 each simulated individual, we recorded their assigned  $\phi$  and  $\lambda$  values, as well as the series of choices they  
243 made during the initial association and the first reversal. For a given  $\phi$  and  $\lambda$ , the stochasticity in which  
244 option a simulated individual chooses based on their attractions, plus the experience of either receiving a  
245 reward or not during previous choices, can lead to differences in the actual choices individuals make. The  
246 aim was to see what sample is needed to correctly infer the assigned  $\phi$  and  $\lambda$  given the noise in the choice  
247 data. We also used the simulated data for question 3, to compare the influence of  $\phi$  and  $\lambda$  on the behavior  
248 of the simulated individuals with that of the grackles.

249 To address question 2, we used an analytical approach and did not analyze any data.

250 For the empirical questions 3-6, we re-analyzed data on the performance of grackles in serial reversal learning  
251 and multi-option puzzle box experiments (Logan et al., 2023a). The data collection was based on our  
252 preregistration that received in principle acceptance at PCI Ecology (Coulon, 2023). All of the analyses  
253 reported here were not part of the original preregistration. The data we use here were published as part of  
254 the earlier article and are available at the Knowledge Network for Biocomplexity's data repository (Logan  
255 et al., 2023b).

256 In brief, grackles were caught in the wild in Tempe, Arizona, USA for individual identification (colored  
257 leg bands in unique combinations), and brought temporarily into aviaries for testing, before being released  
258 back to the wild. The first experiment individuals participated in in the aviaries was the reversal learning  
259 experiment, as described in the introduction. A total of 19 grackles participated in the serial reversal learning  
260 experiment, where they learned to associate a reward with one color before experiencing one reversal to learn

261 that the other color was rewarded (initial rewarded option was counterbalanced and randomly assigned as  
 262 either a dark gray or a light gray tube). The rewarded option was switched when grackles passed the  
 263 criterion of choosing the rewarded option in 17 of the most recent 20 trials. This criterion was set based  
 264 on earlier serial reversal learning studies, and is based on the chi-square test, which indicates that 17 out of  
 265 20 represents a significant association. With this criterion, individuals can be assumed to have learned the  
 266 association between the cue and the reward rather than having randomly chosen one option more than the  
 267 other (Logan et al., 2022). A subset of 8 individuals were randomly assigned to the trained group and went  
 268 through a series of reversals until they reached the criterion of having formed an association (17 out of 20  
 269 choices correct) in 50 trials or less in two consecutive reversals. The individuals in the trained group needed  
 270 between 6-8 reversals to consistently reach this threshold, with the number of reversals not being linked to  
 271 their performance at the beginning or at the end of the experiment. A subset of 11 grackles were part of  
 272 the control group, who experienced only a single reversal, before participating in trials with two identically  
 273 colored tubes (yellow) where both contained a reward. The number of yellow tube trials was set to the  
 274 average number of trials it took a bird in the trained group to pass their serial reversals.

275 For question 6, we additionally used data from an experiment the grackles participated in after they had  
 276 completed the reversal learning experiment. Both the control and trained individuals were provided access  
 277 to two multi-option puzzle boxes, one made of wood and one made of plastic. The two boxes were designed  
 278 with slight differences to explore how general their performance was. The wooden box was made from a  
 279 natural log, thus was more representative of something the grackles might encounter in the wild. In addition,  
 280 while both boxes had four possible ways (options) to access food, the four options on the wooden box were  
 281 distinct compartments, each containing rewards, while the four options on the plastic box all led to the same  
 282 reward. Grackles were tested sequentially on both boxes, in a counterbalanced order, where individuals could  
 283 initially explore all options. After proficiency at an option was achieved (gaining food from this locus three  
 284 times in a row), this option became non-functional by closing access to the option, and then the latency  
 285 of the grackle to switch to attempting a different option was measured. If they again successfully solved  
 286 another option, this second option was also made non-functional, and so on. The outcome measures for each  
 287 individual on each box were the average latency it took to switch to a new option and the total number of  
 288 options they successfully solved.

## 289 The Bayesian reinforcement learning model

290 For both the simulated and the observed grackle data, we used the Bayesian reinforcement learning model to  
 291 estimate for each individual their  $\phi$  and  $\lambda$  values based on the choices they made during the reversal learning  
 292 experiments. The estimated  $\phi$  and  $\lambda$  values were then used as outcome and/or predictor variables in the  
 293 statistical models built to assess questions 3-6. We used the version of the Bayesian model that was developed  
 294 in Blaisdell et al. (2021) and modified in Logan et al. (2023c) (see their Analysis Plan > “Flexibility analysis”  
 295 for model specifications and validation). This model uses data from every trial of reversal learning (rather  
 296 than only using the total number of trials to pass criterion) and represents behavioral flexibility using two  
 297 parameters: the association-updating rate ( $\phi$ ) and the sensitivity to learned associations ( $\lambda$ ). The model  
 298 transforms the series of choices each grackle made based on two equations to estimate the most likely  $\phi$  and  
 299  $\lambda$  that generated the observed behavior.

300 Equation 1 (learning and  $\phi$ ):  $A_{b,o,t+1} = (1 - \phi_b)A_{b,o,t} + \phi_b \pi_{b,o,t}$ .

301 Equation 1 estimates how the associations  $A$ , that individual  $b$  forms between the two different options ( $o$ ,  
 302 option 1 or 2) and their expected rewards, change from one trial to the next (trial  $t+1$ ) as a function of  
 303 their previously formed associations  $A_{b,o,t}$  (how preferable option  $o$  is to grackle  $b$  at trial  $t$ ) and recently  
 304 experienced payoff  $\pi$  (in our case,  $\pi = 1$  when they chose the correct option and received a reward in a  
 305 given trial, and 0 when they chose the unrewarded option). The parameter  $\phi_b$  modifies how much individual  
 306  $b$  updates its associations based on its most recent experience. The higher the value of  $\phi_b$ , the faster the  
 307 individual updates its associations, paying more attention to recent experiences, whereas when  $\phi_b$  is lower,  
 308 a grackle’s associations reflect averages across many trials. Association scores thus reflect the accumulated  
 309 learning history up to trial  $t$ . The association with the option that is not explored in a given trial remains  
 310 unchanged. At the beginning of the experiment (trial  $t$  equals 0), we assumed that individuals had the same  
 311 low association between both options and rewards ( $A_{b,1,0} = A_{b,2,0} = 0.1$ ).

312 Equation 2 (choice and  $\lambda$ ): 
$$P_{b,o,t} = \frac{\exp(\lambda_b A_{b,o,t})}{\sum_{o=1}^2 \exp(\lambda_b A_{b,o,t})}$$

313 Equation 2 is a normalized exponential (softmax) function to convert the learned associations of the two  
 314 options with rewards into the probability,  $P$ , that an individual,  $b$ , chooses one of the two options,  $o$ , in  
 315 the current trial,  $t$ . The parameter  $\lambda_b$  represents the sensitivity of a given grackle,  $b$ , to how different its  
 316 associations to the two options are. As  $\lambda_b$  gets larger, choices become more deterministic and individuals  
 317 consistently choose the option with the higher association even if associations are very similar. As  $\lambda_b$  gets  
 318 smaller, choices become more exploratory, with individuals choosing randomly between the two options  
 319 independently of their learned associations if  $\lambda_b$  is 0.

320 We implemented the Bayesian reinforcement learning model in the statistical language Stan (Stan Develop-  
 321 ment Team, 2023), calling the model and analyzing its output in  $R$  (version 4.3.3) (R Core Team, 2023).  
 322 The model takes the full series of choices individuals make (which of the two options did they choose, which  
 323 option was rewarded, did they make the correct choice) across all their trials to find the  $\phi$  and  $\lambda$  values  
 324 that best fit these choices given the two equations. Which option individuals chose was estimated with a  
 325 categorical distribution with the probability,  $P$ , as estimated from equation 2 for each of the two options  
 326 (categories), before updating the associations using equation 1. The model was fit across all choices, with  
 327 individual  $\phi$  and  $\lambda$  values estimated as varying effects. In the model,  $\phi$  is estimated on the logit-scale to  
 328 reflect that it is a proportion (can only take values between 0 and 1), and  $\lambda$  is estimated on the log-scale to  
 329 reflect that values have to be positive (there is no upper bound). The limitation that, with an estimation on  
 330 the log-scale  $\lambda$  can never be equal to 0, is not an issue because we only included individuals in the analyses  
 331 who did not pick options at random. We set the priors for the logit of  $\phi$  and the log of  $\lambda$  to come from a  
 332 normal distribution with a mean of zero and a standard deviation of one. We set the initial associations  
 333 to both options for all individuals at the beginning of the experiment to 0.1 to indicate that they do not  
 334 have an initial preference for either option but are likely to be somewhat curious about exploring the tubes  
 335 because they underwent habituation and training with a differently colored tube (see below). For estimations  
 336 at the end of each reversal, we set the association with the option that was rewarded before the reversal  
 337 to 0.7 and to the option that was previously not rewarded to 0.1. Note that when applying equation 1 in  
 338 the context of the reversal learning experiment, as is most commonly used, where there are only rewards  
 339 (positive association) or no rewards (zero association) but no punishment (negative association), associations  
 340 can never reach zero because they change proportionally.

341 For each estimation (simulated and observed grackle data), we ran four chains with 2000 samples each (half  
 342 of which were warm up). We used functions in the package “posterior” (Vehtari et al., 2021) to draw 4000  
 343 samples from the posterior (the default). We report the estimates for  $\phi$  and  $\lambda$  for each individual (simulated or  
 344 observed grackle) as the mean from these samples from the posterior. For the subsequent analyses where the  
 345 estimated  $\phi$  and  $\lambda$  values were response or predictor variables, we ran the analyses both with the single mean  
 346 per individual as well as looping over the full 4000 samples from the posterior to reflect the uncertainty in  
 347 the estimates. The analyses with the samples from the posterior provided the same estimates as the analyses  
 348 with the single mean values, though with larger compatibility intervals because of the increased uncertainty.  
 349 In the results, we report the estimates from the analyses with the mean values. The estimates with the  
 350 samples from the posterior can be found in the code in the `rmd` file at the repository [https://github.com/  
 351 corinalogan/grackles/blob/master/Files/Preregistrations/g\\_flexmanip2post.Rmd](https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/g_flexmanip2post.Rmd). In analyses where  $\phi$  and  
 352  $\lambda$  are predictor variables, we standardized the values that went into each analysis (either the means, or  
 353 the respective samples from the posterior) by subtracting the average from each value and dividing by the  
 354 standard deviation. We did this to define the priors for the relationships on a more standard scale and to  
 355 be able to more directly compare the respective influence of  $\phi$  and  $\lambda$  on the outcome variable.

## 356 1) Using simulations to determine whether the Bayesian serial reinforcement 357 learning models have sufficient power to detect changes through the serial re- 358 versal learning experiment

359 We ran the Bayesian reinforcement learning model on the simulated data to understand the minimum number  
 360 of choices per individual that would be necessary to recover the association-updating rate  $\phi$  and the sensitivity



361 to the learned associations  $\lambda$  assigned to each individual.

362 To determine whether the Bayesian reinforcement learning model can accurately recover the simulated  $\phi$   
363 and  $\lambda$  values from limited data, we applied the model first to only the choices from the initial association  
364 learning phase, next to only the choices from the first reversal learning phase, and finally from both phases  
365 combined. To estimate whether the Bayesian reinforcement learning model can recover the simulated  $\phi$  and  
366  $\lambda$  values without bias from either the single or the combined phases, we correlated the estimated values with  
367 the values individuals were initially assigned:

$$\begin{aligned} 368 \phi_{b,1} \text{ or } \lambda_{b,1} &\sim \text{Normal}(\mu_b, \sigma), \\ 369 \mu_b &= \alpha + \beta \times \phi_{b,0} \text{ or } \lambda_{b,0}, \\ 370 \alpha &\sim \text{Normal}(0,0.1), \\ 371 \beta &\sim \text{Normal}(1,1), \\ 372 \sigma &\sim \text{Exponential}(1), \end{aligned}$$

373 where  $\phi_{b,1}$  or  $\lambda_{b,1}$ , the values estimated for each bird, indexed by  $b$ , from the simulated behavior are assumed  
374 to come from a normal distribution with a mean that can vary for each bird,  $\mu_b$ , and overall variance,  $\sigma$ .  
375 The mean for each bird is constructed from an overall intercept,  $\alpha$ , and the change in expectation, the  
376 slope,  $\beta$ , depending on the values assigned to each bird at the beginning of the simulation ( $\phi_{b,0}$  or  $\lambda_{b,0}$ ).  
377 The combination of  $\alpha$  close to 0 and of  $\beta$  close to 1 would indicate that the estimated values matched the  
378 assigned values.

379 This, and all following statistical models, were implemented using functions of the package ‘rethinking’  
380 (McElreath, 2020) in  $R$  to call Stan and estimate the relationships. Following the social convention set in  
381 (McElreath, 2020), we report the mean estimates and the 89% compatibility intervals from the posterior  
382 estimates from these models. For each model, we ran four chains with 10,000 iterations each (half of which  
383 were warm up). We checked that the number of effective samples was sufficiently high and evenly distributed  
384 across all estimated variables such that autocorrelation did not influence the estimates. We also confirmed  
385 that in all cases the Gelman-Rubin convergence diagnostic,  $\hat{R}$ , was 1.01 or smaller, indicating that the  
386 chains had converged on the final estimates (Gelman & Rubin, 1995). In all cases, we also plotted the  
387 model inferences onto the distribution of the raw data to confirm that the estimated predictions matched  
388 the observed patterns.

## 389 **2) Using mathematical derivations to determine whether variation in $\phi$ or $\lambda$ has** 390 **a stronger influence on the number of trials individuals might need to reach** 391 **criterion in serial reversal learning experiments**

392 We mathematically derived predictions about the choice behavior of individuals using equations 1-3. We  
393 determined the values for  $\phi$  and  $\lambda$  that individuals would need to reach the passing criterion in 50 trials or  
394 fewer in the serial reversal learning experiment. To derive the learning curves for individuals with different  
395  $\phi$  and  $\lambda$ , we incorporated the dynamic aspect of change over time by inserting the probabilities of choosing  
396 either the rewarded or the non-rewarded option from trial  $t$  as the likelihood for the changes in associations  
397 at trial  $t+1$ .

398 Equation 3a (dynamic association for the rewarded option):

$$399 A_{r,t+1} = ((1-\phi) \times A_{r,t} + \phi \times \pi) \times P_t + (1-P_t) \times A_{r,t}.$$

400 Equation 3b (dynamic association for the non-rewarded option):

$$401 A_{n,t+1} = (1-P_t) \times (1-\phi) \times A_{n,t} + P_t + (1-P_t) \times A_{n,t}.$$

402 In equations 3a and 3b, the association with both the rewarded,  $A_r$ , and the non-rewarded,  $A_n$ , options  
403 change from trial  $t$  to trial  $t+1$  depending on the association updating rate  $\phi$  and the probability,  $P$ , that  
404 the association was chosen during trial  $t$ . The probability,  $P$ , is calculated using equation 2. The reward  $\pi$   
405 is set to 1. We used these equations to explore which combinations of  $\phi$  and  $\lambda$  would lead to an individual  
406 choosing the rewarded option above the passing criterion in 50 trials or less after a reversal in the rewarded  
407 option. We assumed serial reversals, and therefore set the initial associations after the reversal to 0.1 for the

408 now rewarded option (previously unrewarded, so low association) and to 0.7 for the now unrewarded option  
 409 (previously rewarded, so high association). We obtained these associations from the end of the reversal  
 410 learning simulation in question 1. For a given combination of  $\phi$  and  $\lambda$ , we first used equation 2 to calculate  
 411 the probability that an individual would choose the rewarded option during this first trial after the reversal  
 412 (where the remaining probability reflects the individual choosing the non-rewarded option). We then used  
 413 equations 3a and 3b to update the associations. We repeated the calculations of the probabilities and the  
 414 updates of the associations 50 times to determine whether individuals with a given combination of  $\phi$  and  
 415  $\lambda$  would reach the passing criterion within either 50 (the serial reversal passing criterion) or 40 trials (the  
 416 average observed among the trained grackles). For  $\phi$  ranging between 0.02 and 0.10, we manually explored  
 417 which  $\lambda$  would be needed such that an individual would choose the rewarded option with more than 50%  
 418 probability at trial 31 (or 21) and with more than 85% probability at trial 50 (or 40), to match the passing  
 419 criterion of 17 correct out of the last 20 trials (17/20=0.85).

### 420 **3) Estimating $\phi$ and $\lambda$ from the observed reversal learning performances of grack-** 421 **les to determine which has more influence on variation in how many trials indi-** 422 **viduals needed to reach the passing criterion**

423 We fit the Bayesian reinforcement learning model to the data of both the control and the trained grackles.  
 424 Based on the simulation results indicating that the minimum sample per individual required for accurate  
 425 estimation are two learning phases across one reversal (see below), we fit the model first to only the choices  
 426 from the initial association learning phase and the first reversal learning phase for both control and trained  
 427 individuals. For the control grackles, these estimated  $\phi$  and  $\lambda$  values also reflected their behavioral flexibility  
 428 at the end of the reversal learning experiment. For the trained grackles, we additionally calculated  $\phi$  and  $\lambda$   
 429 separately for their final two reversals at the end of the serial reversal to infer the potential changes in the  
 430 parameters.

431 We determined how the  $\phi$  and  $\lambda$  values influenced the number of trials individuals needed during a reversal  
 432 by building a regression model to determine which of the two parameters had a more direct influence on  
 433 the number of trials individuals needed to reach the passing criterion. We fit this model to the data from  
 434 the simulated individuals, as well as to the data from the grackles. We assumed that the number of trials  
 435 followed a Poisson distribution because the number of trials to reach criterion is a count that is bounded at  
 436 smaller numbers (individuals need at least 20 trials to reach the criterion) with a log-linear link because we  
 437 expect there are diminishing influences of further increases in  $\phi$  or  $\lambda$ . The model is as follows:

$$\begin{aligned}
 438 \quad & v_b \sim \text{Poisson}(\mu), \\
 439 \quad & \log \mu = \alpha + \beta_1 \times \phi_b + \beta_2 \times \lambda_b, \\
 440 \quad & \alpha \sim \text{Normal}(4.5, 1), \\
 441 \quad & \beta_1 \sim \text{Normal}(0, 1), \\
 442 \quad & \beta_2 \sim \text{Normal}(0, 1),
 \end{aligned}$$

443 where the number of trials each individual needed during their reversal,  $v_b$ , was linked with separate slopes,  
 444  $\beta_1$  and  $\beta_2$ , to both the  $\phi$  and  $\lambda$  of each individual. The mean of the prior distribution for the intercept,  $\alpha$ ,  
 445 was based on the average number of trials (90) grackles in Santa Barbara were observed to need to reach the  
 446 criterion during their one reversal (mean of 4.5 is equal to logarithm of 90, standard deviation set to 1 to  
 447 constrain the estimate to the range observed across individuals). The priors for the relationships  $\beta_1$  and  $\beta_2$   
 448 with  $\phi$  and  $\lambda$  were centered on zero, indicating that, *a priori*, we did not bias these toward a relationship.

### 449 **4) Comparing $\phi$ and $\lambda$ from the beginning and end of the observed serial reversal** 450 **learning experiment to assess which changes more as grackles improve their** 451 **performance**

452 For the subset of grackles that were part of the serial reversal group, we calculated how much their  $\phi$  and  $\lambda$   
 453 changed from their first to their last reversal. The model is as follows:

$$\begin{aligned}
 454 \quad & \phi_{b,r} \text{ or } \lambda_{b,r} \sim \text{Normal}(\mu_b, \sigma), \\
 455 \quad & \mu_b = \alpha_b + \beta_b \times r,
 \end{aligned}$$

$$\begin{bmatrix} \alpha_b \\ \beta_b \end{bmatrix} \sim \text{MVNormal} \left( \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, S \right),$$

$$S = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} Z \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix},$$

$$Z \sim \text{LKJcorr}(2),$$

$$\alpha \sim \text{Normal}(5, 2),$$

$$\beta \sim \text{Normal}(-1, 0.5),$$

$$\delta_b \sim \text{Exponential}(1),$$

$$\sigma \sim \text{Exponential}(1),$$

where each grackle,  $b$ , has two  $\phi$  and  $\lambda$  values, one from the beginning ( $r = 0$ ) and one from the end of the serial reversal experiment ( $r = 1$ ). We assume that there are individual differences that persist through the experiment (intercept  $\alpha_b$ ), and that how much individuals change from the first to the last reversal,  $r$ , estimated by  $\beta_b$ , might also depend on their values at the beginning. Each bird has an intercept and slope with a prior distribution defined by the two dimensional Gaussian distribution (*MVNormal*) with means,  $\sigma_\alpha$  and  $\sigma_\beta$ , and covariance matrix,  $S$ . The covariance matrix,  $S$ , is factored into separate standard deviations,  $\delta_b$ , and a correlation matrix,  $Z$ . The prior for the correlation matrix is set to come from the Lewandowski-Kurowicka-Joe (*LKJcorr*) distribution, and is set to be weakly informative and skeptical of extreme correlations near -1 or 1.

We also fit a model to assess whether individual improvement in the number of trials from their first to their last reversal was linked more to their change in  $\phi$  or to their change in  $\lambda$ . The model is as follows:

$$\Delta v_b \sim \text{Normal}(\mu_b, \sigma),$$

$$\mu_b = \alpha + \beta_1 \times \Delta\phi_b + \beta_2 \times \Delta\lambda_b,$$

$$\alpha_b \sim \text{Normal}(40, 10),$$

$$\beta_1 \sim \text{Normal}(0, 10),$$

$$\beta_2 \sim \text{Normal}(0, 10),$$

$$\sigma \sim \text{Exponential}(1),$$

where  $\Delta v_b$ , the improvement in the number of trials, is the difference in the number of trials between the first and the last reversal, and  $\Delta\phi_b$  and  $\Delta\lambda_b$  are the respective differences in these parameters between the beginning and the end of the serial reversal experiment. The remaining parameters in the model are as defined above.

## 5) Calculating whether individual differences in $\phi$ and $\lambda$ persist throughout the serial reversal learning experiment and whether grackles differ in how much they change throughout the experiment

We checked whether the  $\phi$  and  $\lambda$  values of grackles at the beginning were associated with how much they changed (difference in values between beginning and end):

$$\Delta\phi_b \text{ or } \Delta\lambda_b \sim \text{Normal}(\mu_b, \sigma),$$

$$\mu_b = \alpha + \beta \times \phi_{b,0} \text{ or } \lambda_{b,0},$$

$$\alpha \sim \text{Normal}(0, 1),$$

$$\beta \sim \text{Normal}(0, 1),$$

$$\sigma \sim \text{Exponential}(1),$$

where  $\Delta\phi_b$  and  $\Delta\lambda_b$  are the changes in these values, and  $\phi_{b,0}$  and  $\lambda_{b,0}$  are the bird's values from their first reversal. The remaining parameters are as defined above.

We also checked whether the  $\phi$  or  $\lambda$  values of grackles at the beginning were associated with the values they had at the end:

497  $\phi_{b,1}$  or  $\lambda_{b,1} \sim \text{Normal}(\mu_b, \sigma)$ ,  
 498  $\mu_b = \alpha + \beta \times \phi_{b,0}$  or  $\lambda_{b,0}$ ,  
 499  $\alpha \sim \text{Normal}(0,1)$ ,  
 500  $\beta \sim \text{Normal}(0,1)$ ,  
 501  $\sigma \sim \text{Exponential}(1)$ ,

502 where  $\phi_{b,1}$  and  $\lambda_{b,1}$  are from the last reversal. The remaining parameters are as defined above.

503 In addition, we assessed whether grackles at the end of the serial reversal experiment focused more on one  
 504 of the processes,  $\phi$  or  $\lambda$ , than the other. The model is as follows:

505  $\phi_{b,1} \sim \text{Normal}(\mu_b, \sigma)$ ,  
 506  $\mu_b = \alpha + \beta \times \lambda_{b,1}$ ,  
 507  $\alpha \sim \text{Normal}(0,1)$ ,  
 508  $\beta \sim \text{Normal}(0,1)$ ,  
 509  $\sigma \sim \text{Exponential}(1)$ ,

510 where the values estimated for birds from their last reversal are assessed for an association. All parameters  
 511 as defined above.

512 We used the  $\phi$  and  $\lambda$  values estimated from individuals after they completed the serial reversal learning  
 513 experiment to better understand how individuals behave after a reversal in which option is rewarded. We  
 514 chose two combinations of  $\phi$  and  $\lambda$  from the end of the range of values observed among the individuals who  
 515 completed the serial reversal learning experiment. The first combines a slightly higher  $\phi$  (0.09) with a slightly  
 516 lower  $\lambda$  (3), and the second combines a slightly lower  $\phi$  (0.06) with a slightly higher  $\lambda$  (4). We entered these  
 517 values in equations 2, 3a, and 3b. We plotted the change in the probability that an individual will choose  
 518 the rewarded option across the first 40 trials after a switch. As above, we set the initial associations to the  
 519 now rewarded option to 0.1 and to the now non-rewarded option to 0.7.

## 520 **6) Linking $\phi$ and $\lambda$ from the observed serial reversal learning performances to** 521 **the performance on the multi-option puzzle boxes**

522 We modified the statistical models in the original article (Logan et al., 2023a) that linked performance on  
 523 the serial reversal learning tasks to performance on the multi-option puzzle boxes, replacing the previously  
 524 used independent variable of the number of trials needed to reach criterion in the last reversal with the  
 525 estimated  $\phi$  and  $\lambda$  values from the last two reversals (trained grackles) or the initial discrimination and the  
 526 first reversal (control grackles). We assumed that there also might be non-linear, U-shaped relationships  
 527 between  $\phi$  and/or  $\lambda$  and the performance on the multi-option puzzle box. For the number of options solved,  
 528 we fit a binomial model with a logit link:

529  $o_b \sim \text{Binomial}(4, p)$ ,  
 530  $\text{logit}(p) \sim \alpha + \beta_1 \times \phi + \beta_2 \times \phi^2 + \beta_3 \times \lambda + \beta_4 \times \lambda^2$ ,  
 531  $\alpha \sim \text{Normal}(1, 1)$ ,  
 532  $\beta_1 \sim \text{Normal}(0, 1)$ ,  
 533  $\beta_2 \sim \text{Normal}(0, 1)$ ,  
 534  $\beta_3 \sim \text{Normal}(0, 1)$ ,  
 535  $\beta_4 \sim \text{Normal}(0, 1)$ ,

536 where  $o_b$  is the number of options solved on the multi-option puzzle box, 4 is the total number of options  
 537 on the multi-option puzzle box,  $p$  is the probability of solving any one option across the whole experiment,  
 538  $\alpha$  is the intercept,  $\beta_1$  is the expected linear amount of change in  $p$  for every one unit change in  $\phi$  in the  
 539 reversal learning experiments,  $\beta_2$  is the expected non-linear amount of change in  $p$  for every one unit change  
 540 in  $\phi^2$ ,  $\beta_3$  the expected linear amount of change for changes in  $\lambda$ , and  $\beta_4$  is the expected non-linear amount  
 541 of change for changes in  $\lambda^2$ .

542 For the average latency to attempt a new option on the multi-option puzzle box as it relates to  $\phi$  and  $\lambda$ , we  
 543 fit a Gamma-Poisson model with a log-link:

544  $n_b \sim \text{Gamma-Poisson}(m_b, s)$ ,  
545  $\log(m_b) \sim \alpha + \beta_1 \times \phi + \beta_2 \times \phi^2 + \beta_3 \times \lambda + \beta_4 \times \lambda^2$ ,  
546  $\alpha \sim \text{Normal}(1, 1)$ ,  
547  $\beta_1 \sim \text{Normal}(0, 1)$ ,  
548  $\beta_2 \sim \text{Normal}(0, 1)$ ,  
549  $\beta_3 \sim \text{Normal}(0, 1)$ ,  
550  $\beta_4 \sim \text{Normal}(0, 1)$ ,  
551  $s \sim \text{Exponential}(1)$ ,

552 where  $n_b$  is the average latency, counted as the number of seconds, to attempt a new option on the multi-  
553 option puzzle box,  $m_b$  reflects the tendency of each grackle to wait (if they have a higher tendency to wait,  
554 they have a longer latency),  $s$  controls the variance (larger values mean the overall distribution is more like  
555 a pure Poisson process in which all grackles have the same tendency to wait),  $\alpha$  is the intercept,  $\beta_1$  is the  
556 expected linear amount of change in latency for every one unit change in  $\phi$ ,  $\beta_2$  is the expected non-linear  
557 amount of change in latency for every one unit change in  $\phi^2$ ,  $\beta_3$  the expected linear amount of change for  
558 changes in  $\lambda$ , and  $\beta_4$  is the expected non-linear amount of change for changes in  $\lambda^2$ .

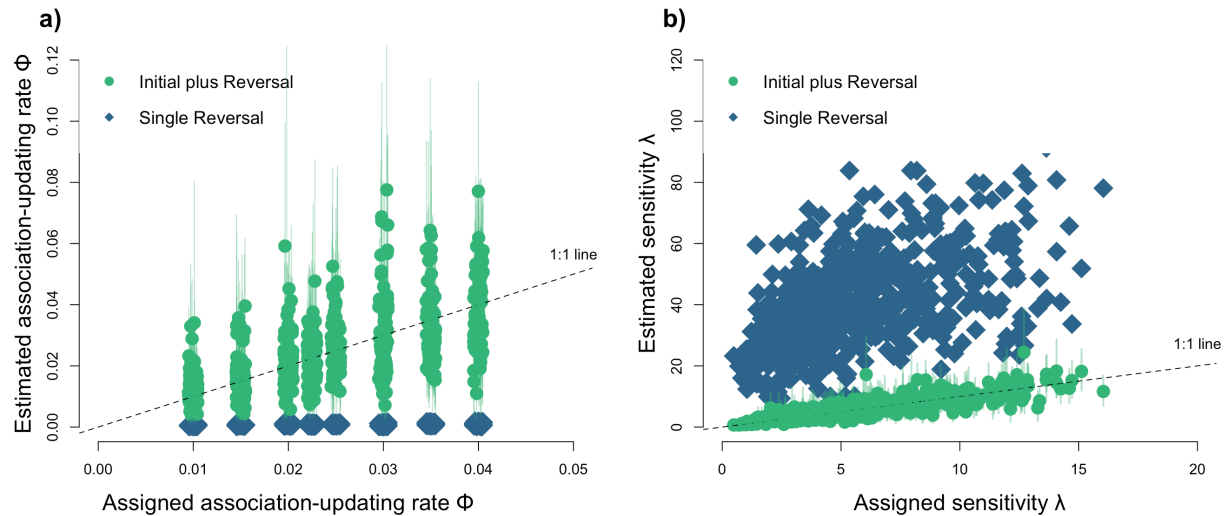
## 559 Results

### 560 1) Power of the Bayesian reinforcement learning model to detect short-term 561 changes in the association-updating rate, $\phi$ , and the sensitivity to learned asso- 562 ciations, $\lambda$

563 Applying the Bayesian reinforcement learning model to simulated data from only a single phase (initial  
564 association or first reversal) revealed that, while the model recovered the differences among individuals, the  
565 estimated  $\phi$  and  $\lambda$  values did not match those the individuals had been assigned (Figure 2). The estimated  
566  $\phi$  and  $\lambda$  values were consistently shifted away from the values assigned to the simulated individuals. The  
567 estimated  $\phi$  values were consistently smaller than those assigned to the simulated individuals (here and  
568 hereafter, we report the posterior mean slope of the association, the  $\beta$  factor in the statistical models, with  
569 the 89% compatibility interval; +0.15, +0.06 to +0.23, n=626 simulated individuals), while the estimated  
570  $\lambda$  values were consistently estimated to be larger than the assigned  $\lambda$  values (+6.04, +5.86 to +6.22, n=626  
571 simulated individuals)(Figure 2). The model assumed that, during the initial association learning, individuals  
572 only needed to experience each option once to learn which of the two options to choose. This would lead to  
573 a difference in the associations between the two options. The model assumed that the simulated individuals  
574 would not require a large  $\phi$  because a small difference in the associations would already be informative.  
575 Individuals would then be expected to consistently choose the option that was just rewarded, and they would  
576 because of their large  $\lambda$ . In addition, these shifts mean that  $\phi$  and  $\lambda$  are no longer estimated independently.  
577 The model estimated that, if an individual had a particularly low  $\phi$  value, it would require a particularly  
578 high  $\lambda$  value. This dependency (which was due to inaccurate estimation) between  $\phi$  and  $\lambda$  led to a strong  
579 positive correlation in the estimated values of  $\phi$  and  $\lambda$  (+505, +435 to +570, n=626 simulated individuals).  
580 This correlation is erroneous because individuals were assigned their  $\lambda$  values independent of their  $\phi$  values,  
581 with the different combinations across the populations meaning that high and low values of  $\lambda$  were assigned  
582 to individuals with both high and with low  $\phi$  values.

583 In contrast, when we combined data from across the initial discrimination learning and the first reversal,  
584 the model recovered the  $\phi$  and  $\lambda$  values that the simulated individuals had been assigned ( $\phi$ : intercept  
585 0.00, -0.01 to +0.01, slope +0.96, +0.70 to +1.21, n=626 simulated individuals;  $\lambda$ : intercept +0.01, -0.15 to  
586 +0.16, slope +0.98, +0.92 to +1.05, n=626 simulated individuals) (Figure 2). While different combinations  
587 of  $\phi$  and  $\lambda$  could potentially explain the series of choices during a single phase (initial discrimination and  
588 single reversal), these different combinations lead to different assumptions about how an individual would  
589 behave right after a reversal when the reward is switched. In combination, the choices before and after a  
590 reversal make it possible to infer the assigned values (initial learning plus first reversal, or two subsequent  
591 reversals). Given that the choices individuals make during any given trial are probabilistic, the estimation  
592 can show slight deviations from the assigned values. However, this was also reflected in the uncertainties of

593 the estimated values, and the compatibility intervals of the estimated values included the value assigned to  
 594 the simulated individuals (Figure 2).

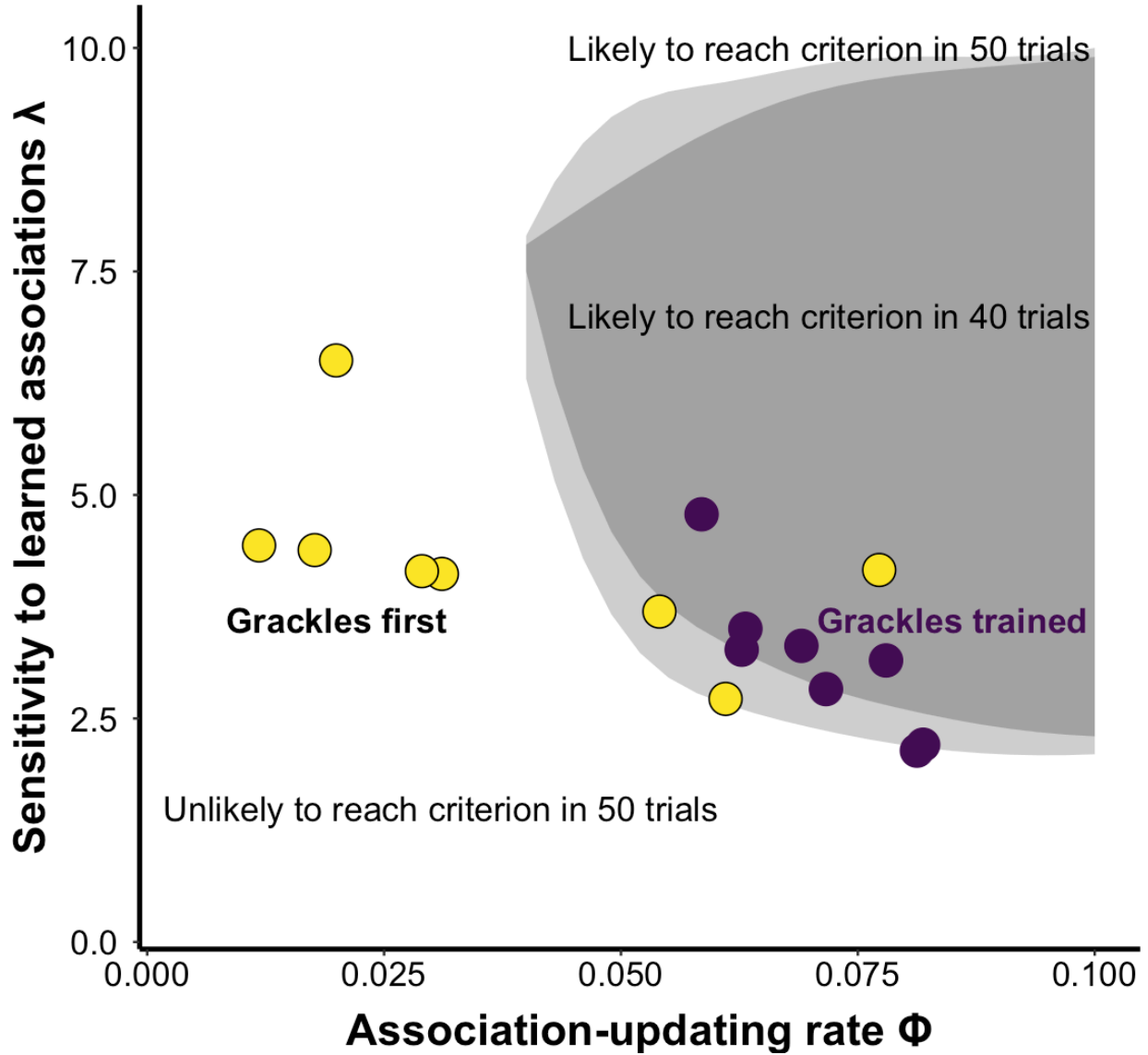


595

596 **Figure 2.** Both the  $\phi$  (a) and the  $\lambda$  (b) values are only estimated correctly by the Bayesian reinforcement  
 597 model when the choices from the simulated reversal learning are combined with the previous initial association  
 598 learning (green circles). When  $\phi$  was estimated based on the choices made only during the first reversal, the  
 599 estimates were consistently lower than the assigned values, particularly for large  $\phi$  values (a, blue diamonds).  
 600 The model assumed that the simulated individuals chose the rewarded option consistently not because they  
 601 updated their associations, but because they consistently chose the rewarded option as soon as they had  
 602 learned which option was rewarded. Accordingly, the model wrongly assigned individuals very high  $\lambda$  values  
 603 (b, blue diamonds). Lines around the points indicate the 89% compatibility intervals of the estimated values  
 604 and are only shown for the estimation from the combined choices from the initial and reversal learning - the  
 605 approach we ended up using for the remaining analyses.

## 606 2) Role of $\phi$ and $\lambda$ on performance in the serial reversal learning task based on 607 analytical predictions

608 To determine how  $\phi$  and  $\lambda$  influence behavior during the serial reversals, we performed a mathematical  
 609 derivation using equations 2, 3a, and 3b. We identified the range of values for  $\phi$  and for  $\lambda$  that we would expect  
 610 in individuals who quickly change their behavior after a reversal in the serial reversal learning experiment.  
 611 We found that  $\phi$  needs to be 0.04 or larger for individuals to be able to reach the passing criterion in 40 or  
 612 50 trials after a reversal (Figure 3). With smaller  $\phi$  values, individuals are expected to take longer before  
 613 switching to the newly rewarded option because they would not update their associations fast enough. We  
 614 also found that, as  $\phi$  values increased beyond 0.04, individuals could have a larger range of  $\lambda$  values and still  
 615 reach the passing criterion in 40 or 50 trials. However, the  $\lambda$  values are expected to be less than 10 and as  
 616 low as 2.4.



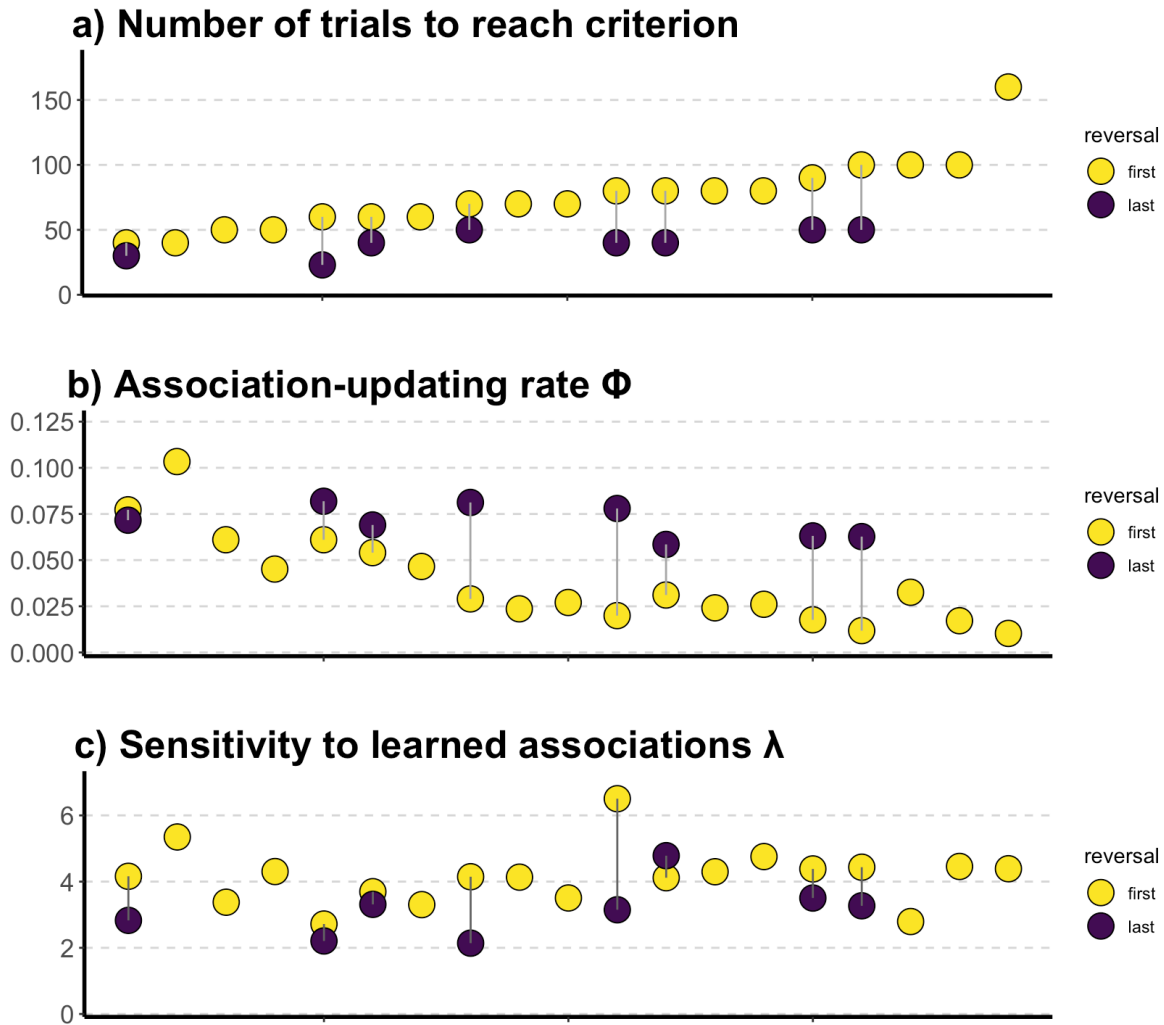
617

618 **Figure 3.** Individuals are more likely to reach the criterion of choosing the correct option 17 out of 20 times  
 619 during the serial reversal trials if they update their associations quickly (high  $\phi$ ). Using the equations, we  
 620 found the space of values individuals are predicted to need to reach the passing criterion in 40 trials or less  
 621 (dark gray shading) or 50 trials or less (light gray shading). Individuals are predicted to need a large  $\phi$  to  
 622 completely reverse their associations with the two options presented in the serial reversal learning experiment.  
 623 The predicted  $\lambda$  values are expected to be relatively small given that there is no upper limit. The figure also  
 624 shows the median  $\phi$  and  $\lambda$  values estimated for the trained grackles during their first reversal (yellow), when  
 625 they needed on average 70 trials to reach criterion, and during their last reversal (purple) when they needed  
 626 on average 40 trials to reach criterion. During the training, grackles increased their  $\phi$  to become efficient at  
 627 gaining the reward and reaching the criterion. They also showed a slight decline in their  $\lambda$ , allowing them  
 628 to explore the alternative option after a reversal.

629 **3) Observed role of  $\phi$  and  $\lambda$  on performance of grackles in the reversal learning**  
 630 **task**

631 We estimated  $\phi$  and  $\lambda$  after the first reversal for all grackles, and additionally after the final reversal for the  
 632 individuals who experienced the serial reversal learning experiment. The findings from the simulated data

633 indicated that  $\lambda$  and  $\phi$  can only be estimated accurately when calculated across at least one reversal. In  
 634 the simulation, we could combine the performance of individuals during the initial learning with the first  
 635 reversal to estimate the parameters because the behavior during those two phases in the simulations was  
 636 determined in the same way by the  $\phi$  and  $\lambda$  values that individuals were assigned. We determined that  
 637 we can also combine the first two phases for the observed grackle data because we found that the number  
 638 of trials grackles needed to reach criterion during the initial learning and the first reversal learning were  
 639 correlated (+1.61, +1.53 to +1.69, n=19 grackles), where grackles needed about 28 trials more to reach  
 640 criterion during the first reversal than they needed during the initial association learning. Therefore, we  
 641 estimated  $\phi$  and  $\lambda$  for the grackles based on their performance in the initial discrimination plus first reversal,  
 642 and for the trained grackles additionally based on their performance in their final two reversals. The inferred  
 643  $\phi$  values for the grackles in Arizona ranged between 0.01 and 0.10, and the  $\lambda$  values between 2.1 and 6.5  
 644 (Figure 4).

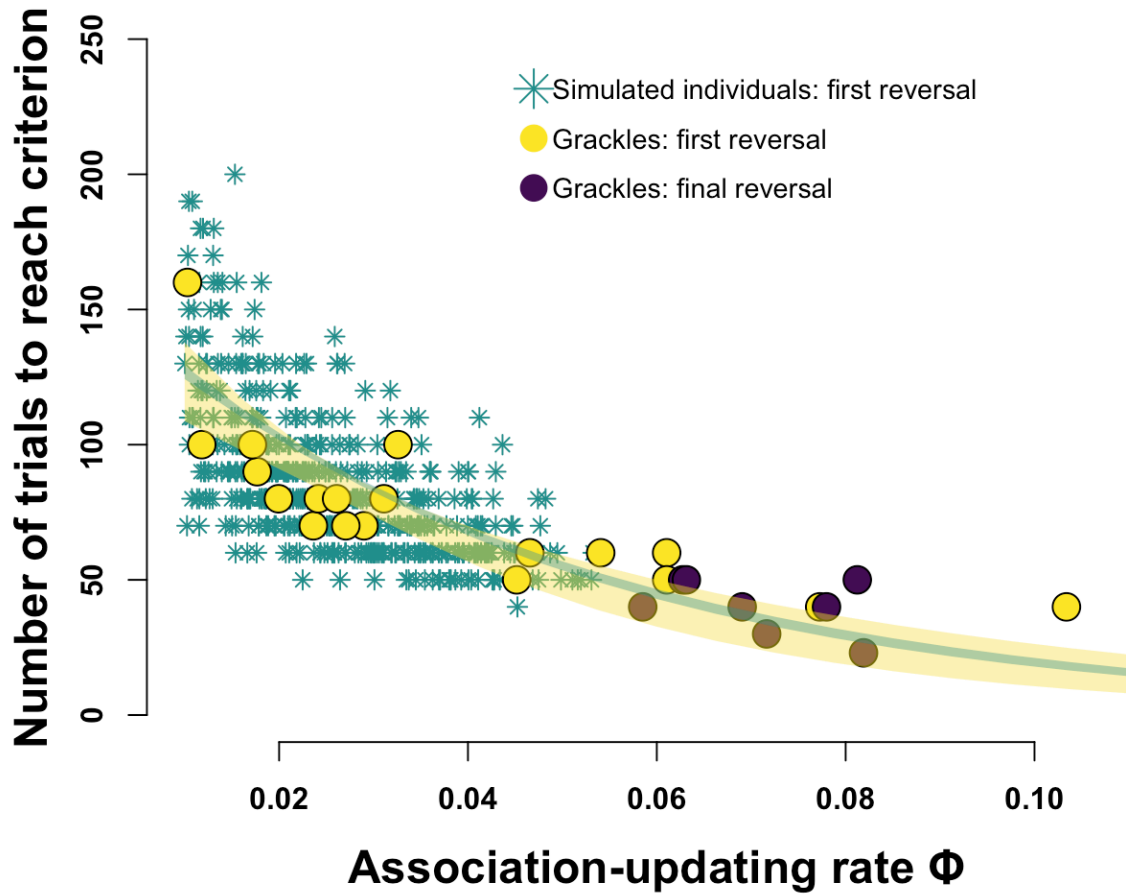


645  
 646 **Figure 4.** Comparisons of the parameters estimated from the behavior of 19 grackles in the serial reversal  
 647 task. The figure shows a) the number of trials to pass criterion for the first reversal (yellow; all grackles) and  
 648 the last reversal (purple; only trained grackles); b) the  $\phi$  values reflecting the rate of updating associations  
 649 with the two options inferred from the initial discrimination and first reversal (yellow; all grackles) and  
 650 from the last two reversals (purple; trained grackles); and c) the  $\lambda$  values reflecting the sensitivity to the learned  
 651 associations inferred from the initial discrimination and first reversal (yellow; all grackles) and from the last  
 652 two reversals (purple; trained grackles). Individual grackles have the same position along the x-axis in all



653 three panels. Grackles that needed fewer trials to reverse their preference generally had higher  $\phi$  values,  
 654 whereas  $\lambda$  appeared unrelated to the number of trials grackles needed during the first reversal. For the  
 655 trained grackles, their  $\phi$  values changed more consistently than their  $\lambda$  values: their  $\phi$  values were generally  
 656 higher than those observed in the control individuals, while their  $\lambda$  values remained within the range observed  
 657 for the control group.

658 For the 19 grackles that finished the initial learning and the first reversal, only their  $\phi$  (-20.69, -26.17 to  
 659 -15.13; n=19 grackles), but not their  $\lambda$  (-0.22, -5.66 to +5.26, n=19 grackles), predicted the number of trials  
 660 they needed to reach criterion during their first reversal (Figure 4, increase left to right in panel a), decrease  
 661 in panel b), no pattern in panel c)). A grackle with a  $\phi$  of 0.01 higher than another individual needed about  
 662 10 fewer trials to reach the criterion. The slope between  $\phi$  and the number of trials for the grackles was  
 663 essentially the same as the slope from the simulations (-20.69 vs -20.48, Figure 5). The number of trials  
 664 grackles needed to reach the criterion given their  $\phi$  values fell right into the range for the relationship between  
 665  $\phi$  and the number of trials for simulated individuals (Figure 5). Even though the 8 trained grackles also  
 666 appeared to need slightly fewer trials to reach criterion in their final two reversals if they had a higher  $\phi$ ,  
 667 the limited variation in the number of trials and in  $\phi$  and  $\lambda$  values among individuals means that there is  
 668 no clear association between the number of trials and either parameter in the last reversals ( $\phi$ : -7.38, -15.97  
 669 to +1.28;  $\lambda$ : -4.00, -12.53 to +4.61, n=8 grackles).



670

671 **Figure 5.** Relationship between  $\phi$  and the number of trials needed to reach criterion observed among grackles  
 672 during their first reversal (yellow points; all grackles) and last reversal (purple points; trained grackles), as

673 well as for the first reversal for the simulated individuals (green stars). The observed grackle data falls within  
674 the range of the number of trials individuals with a given  $\phi$  value are expected to need. Grackles show the  
675 same negative correlation between their  $\phi$  and the number of trials needed to reach criterion as the simulated  
676 individuals (the shaded lines display the 89% compatibility interval of the estimated relationships between  
677  $\phi$  and the number of trials for both the simulated individuals, green line, and for the grackles during their  
678 first reversal, yellow line). We did not simulate individuals with  $\phi$  values larger than 0.05 because we did  
679 not observe larger values among grackles in the Santa Barbara population, which we used to parameterize  
680 the simulations.

#### 681 4) Changes in $\phi$ and $\lambda$ through the serial reversal learning task

682 Grackles who experienced the serial reversal learning reduced the number of trials they needed to reach the  
683 criterion from an average of 75 to an average of 40 by the end of their experiment (-30.02, -36.05 to -24.16,  
684 n=8 grackles). For the trained grackles, the estimated  $\phi$  values more than doubled from 0.03 in their initial  
685 discrimination and first reversal (which is identical to the average observed among the control grackles who  
686 did not experience the serial reversals) to 0.07 in their last two reversals (+0.03, +0.02 to +0.05, n=8). The  
687  $\lambda$  values of the trained grackles went slightly down from 4.2 (again, similar to control grackles) to 3.2 (-1.07,  
688 -1.63 to -0.56, n=8 grackles) (Figure 4). The number of trials to reverse that we observed in the last reversal,  
689 as well as the  $\phi$  and  $\lambda$  values estimated from the last reversal, all fall within the range of variation we observed  
690 among the control grackles in their first and only reversal (Figure 5). This means that the training did not  
691 push grackles to new levels, but changed them within the boundaries of their natural abilities observed in  
692 the population.

693 As predicted, the increase in  $\phi$  during the training fits with the outcome from the mathematical predictions:  
694 larger  $\phi$  values were associated with fewer trials to reverse. The improvement the grackles showed in the  
695 number of trials they needed to reach the criterion from the first to the last reversal matched the increase  
696 in their  $\phi$  values (+7.59, +1.54 to +14.22, n=8 grackles). The improvement did not match the change in  
697 their  $\lambda$  values (+2.17, -4.66 to +9.46, n=8 grackles) because, as predicted, the trained grackles showed a  
698 decreased  $\lambda$  in their last reversal. This decrease in  $\lambda$  meant that grackles quickly found the rewarded option  
699 after a reversal in which option was rewarded. Across all grackles, in their first reversal, grackles chose the  
700 newly rewarded option in 25% of the first 20 trials, while the trained grackles in their final reversal chose  
701 correctly in 35% of the first 20 trials. Despite their low  $\lambda$  values, trained grackles still chose the rewarded  
702 option consistently because the increase in  $\phi$  compensated for this reduced sensitivity (Figure 3; also see  
703 below).

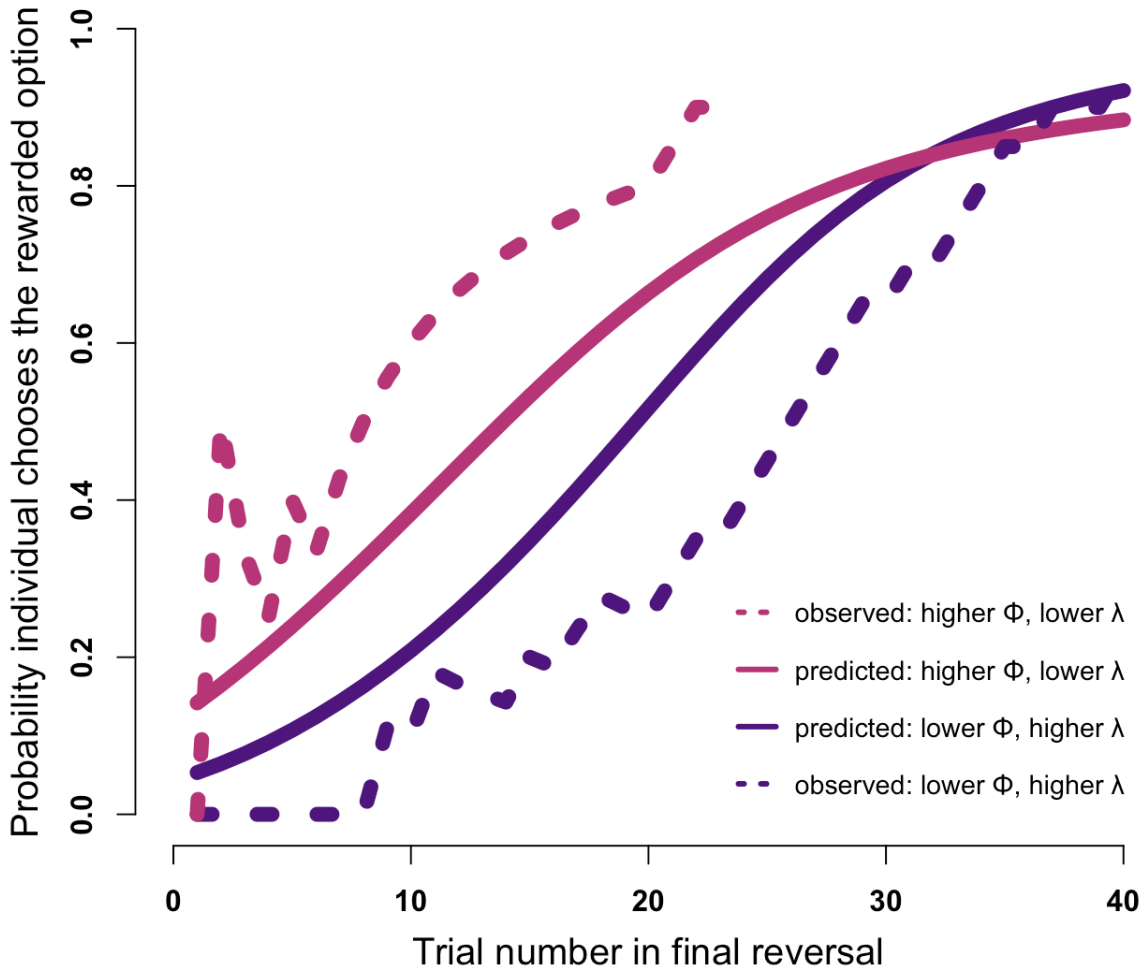
#### 704 5) Individual consistency in the serial reversal learning task

705 We found a negative correlation between the  $\phi$  estimated from an individual's performance in the first  
706 reversal and how much their  $\phi$  changed through the serial reversals (-0.84, -1.14 to -0.52, n=8 grackles). The  
707 larger increases in  $\phi$  for individuals who had smaller  $\phi$  values at the beginning made it so that individuals  
708 ended up with similar  $\phi$  values at the end of the serial reversals. We did not find consistent individual  
709 variation among grackles in  $\phi$ : their beginning and end  $\phi$  values were not correlated (-0.21, -1.55 to +1.35,  
710 n=8 grackles). Similarly, individuals who started with a high  $\lambda$  changed more than individuals who already  
711 had a lower  $\lambda$  during the first reversal (-0.44, -0.76 to -0.10, n=8 grackles). Individuals changed to different  
712 degrees, such that those with higher  $\lambda$  values in the beginning did not necessarily have higher  $\lambda$  values than  
713 other individuals at the end of the serial reversal learning: their values at the beginning and end were not  
714 associated (+0.17, -0.67 to +0.97, n=8 grackles).

715 Individuals appeared to adjust their behavior differently to improve their performance through the serial  
716 reversals. There was a negative correlation between an individual's  $\phi$  and  $\lambda$  after their last reversal (-0.39,  
717 -0.72 to -0.06, n=8 grackles). While, as predicted, essentially all grackles who experienced the serial reversal  
718 learning experiments increased their  $\phi$  and decreased their  $\lambda$  (Figure 5), individuals ended up with different  
719 combinations of the two parameters and all combinations allowed them to switch to the newly rewarded  
720 option in 50 trials or less. Individuals ended up along the lower (on the y-axis) side of the space of values  
721 that are needed to reach criterion in the serial reversal learning experiment (the lower edge of the light gray

722 shading in Figure 3).

723 We illustrate how these differences in  $\phi$  and  $\lambda$  lead to slightly different ways of reaching the passing criterion  
724 during the final reversal. We used the values from the two individuals at the ends of the spectrum, the one  
725 with the highest  $\phi$  and lowest  $\lambda$ , and the one with the lowest  $\phi$  and highest  $\lambda$ , to explore how individuals  
726 switched from the previous option to the option that is now being rewarded. Based on equations 1-3,  
727 individuals with a slightly higher  $\phi$  and slightly lower  $\lambda$  are expected to learn the new reward associations  
728 after a reversal more quickly. However, they continue to explore the alternative option even after they  
729 learned the new association and therefore do not exclusively choose the rewarded option (red line in Figure  
730 6). Individuals with a slightly lower  $\phi$  and a slightly higher  $\lambda$  are expected to take slightly longer to learn  
731 that the reward has switched, but once they reversed their association, they rarely choose the unrewarded  
732 option (purple line in Figure 6). Together, this suggests that all individuals improved by the same extent  
733 through the training such that the differences in their performances persisted, but they utilized slightly  
734 different behaviors to quickly reach criterion after a reversal.



735

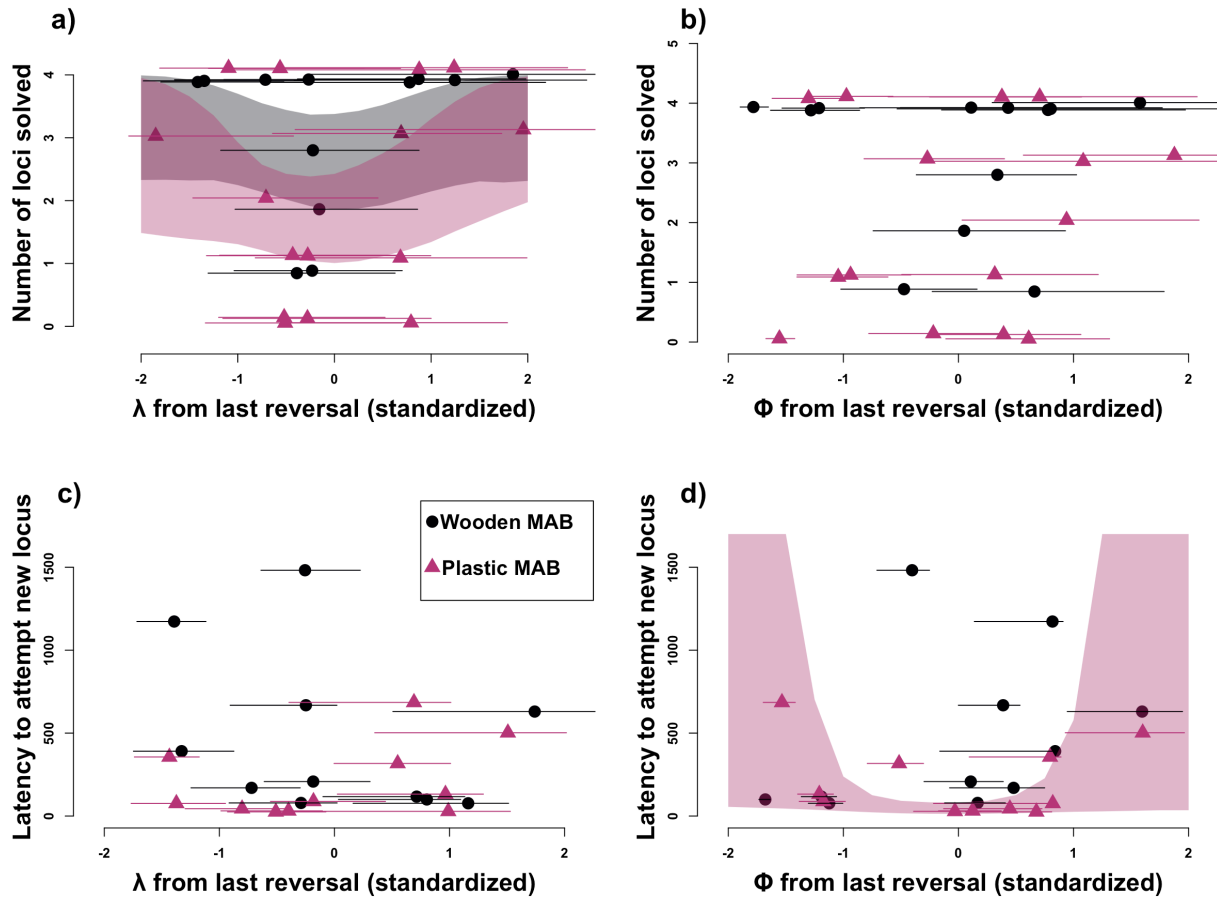
736 **Figure 6.** Predicted and observed performance curves of individuals with different  $\phi$  and  $\lambda$  values in their  
737 last reversal in the serial reversal learning experiment. The dotted lines present the behavior of the grackles

738 Burrito (red on the top,  $\phi = 0.08$ ,  $\lambda = 2.1$ ) and Habanero (purple on the bottom,  $\phi = 0.06$ ,  $\lambda = 4.8$ ) during  
739 their last reversal. The dotted lines show the probability with which they chose the rewarded option during  
740 their last 20 trials. We used their  $\phi$  and  $\lambda$  values in the analytical equations 2, 3a, and 3b to derive the  
741 predicted curves (solid lines) of the probability that an individual will choose the option that is currently  
742 rewarded for each trial number. Individuals with a higher  $\phi$  and lower  $\lambda$  (red lines on the top) are expected  
743 and observed to quickly learn the new association, but continue to explore the unrewarded option even after  
744 they learned the association, leading to a curve with a more gradual increase through the trials. Individuals  
745 with a lower  $\phi$  and higher  $\lambda$  (purple lines on the bottom) are expected and observed to take longer to switch  
746 their association, but, once they do, they rarely choose the non-rewarded option, leading to a more S-shaped  
747 curve where the initial increase in probability is lower and more rapid later.

## 748 6) Association between $\phi$ and $\lambda$ with performance on the multi-option puzzle 749 boxes

750 We found that the number of options solved for both the wooden and the plastic multi-option puzzle boxes as  
751 well as the latency to solve a new option on both boxes correlated with the underlying flexibility parameters  
752  $\phi$  and  $\lambda$ . In particular, the  $\lambda$  values individuals had after their last reversal had a U-shaped relationship with  
753 the number of options solved on both the plastic ( $\lambda +0.17$ ,  $-0.27$  to  $+0.61$ ;  $\lambda^2 +0.59$ ,  $+0.18$  to  $+1.02$ ;  $n=15$   
754 grackles) and the wooden multi-option puzzle boxes ( $\lambda +0.03$ ,  $-0.50$  to  $+0.59$ ;  $\lambda^2 +0.63$ ,  $+0.12$  to  $+1.19$ ;  
755  $n=12$  grackles). There was no association between the number of options solved on either box and  $\phi$  (plastic  
756 box:  $\phi +0.03$ ,  $-0.38$  to  $+0.43$ ;  $\phi^2 -0.16$ ,  $-0.59$  to  $+0.28$ ,  $n=15$  grackles; wooden box:  $\phi -0.08$ ,  $-0.62$  to  $+0.47$ ,  
757  $\phi^2 +0.43$ ,  $-0.08$  to  $+0.97$ ,  $n=12$  grackles). Grackles who had either particularly low or particularly high  
758 sensitivities to their previously learned associations were more likely to solve all four options than grackles  
759 with intermediate values of  $\lambda$  (Figure 7).

760 For the latency to attempt a new option on the plastic box, there was also a U-shaped association, but  
761 only with  $\phi$  ( $\phi -0.66$ ,  $-1.30$  to  $+0.06$ ;  $\phi^2 +0.58$ ,  $-0.06$  to  $+1.30$ ;  $\lambda +0.14$ ,  $-0.45$  to  $+0.70$ ;  $\lambda^2 +1.09$ ,  $+0.28$   
762 to  $+1.87$ ;  $n=11$  grackles). Grackles with either particularly high or particularly low rates of updating their  
763 associations took longer to attempt a new option than grackles with intermediate values of  $\phi$  (Figure 8).  
764 There was no association between the latency to attempt a new option on the wooden box with either  $\phi$   
765 ( $-0.62$ ,  $-1.46$  to  $+0.14$ ;  $\phi^2 +0.39$ ,  $-0.47$  to  $+1.26$ ; 11 grackles) or  $\lambda$  ( $+0.13$ ,  $-0.66$  to  $+0.86$ ;  $\lambda^2 +0.32$ ,  $-0.62$  to  
766  $+1.35$ ;  $n=11$  grackles).



767

768 **Figure 7.** Relationships between  $\phi$  and  $\lambda$  from the last reversal and performance on the wooden (black  
769 dots) and plastic (magenta triangles) multi-option puzzle boxes. Grackles with intermediate  $\lambda$  values in their  
770 last reversal (a) were less likely to solve all four options on both multi-option puzzle boxes than grackles  
771 with either high or low  $\lambda$  values. Grackles with intermediate  $\phi$  values had a shorter latency to attempt a  
772 new option on the plastic box (d). There were no clear relationships between  $\phi$  and the number of options  
773 solved on either box (b),  $\lambda$  and the latency to attempt an option on either box (c), or  $\phi$  and the latency to  
774 attempt a new option on the wooden box (d). An individual's  $\phi$  and  $\lambda$  values changed slightly between the  
775 top and bottom rows because values were standardized for each plot and not all individuals were tested on  
776 both boxes, therefore values changed relative to the mean of the points included in each plot. The shaded  
777 areas (black for the data for the wooden box, magenta for the data from the plastic box) show the 89%  
778 compatibility intervals for the detected relationships between  $\phi / \lambda$  and the respective outcome variable.  
779 Lines around each point indicate the 89% compatibility intervals for the estimated  $\phi$  and  $\lambda$  values.

## 780 Discussion

781 Our analyses show that grackles change their behavioral flexibility to match the reliability and stability  
782 of the environment they experience. The application of the Bayesian reinforcement learning model to the  
783 grackle serial reversal learning data revealed that the association-updating rate,  $\phi$ , explained more of the  
784 interindividual variation in how many trials individuals needed to reach criterion during a reversal than the  
785 sensitivity to learned associations,  $\lambda$ . We found that, as predicted given the reliability of cues and frequent  
786 switches in the serial reversal learning experiment,  $\phi$  more than doubled between the first and last reversals,  
787 whereas  $\lambda$  slightly declined. Even though all grackles changed their behavior in the expected direction by  
788 the end of the serial reversal learning experiment, we found that these trained individuals used slightly

789 different approaches from across the range of possible behaviors. Finally, these changes in how the trained  
790 individuals explored alternative options and switched preferences in light of recent information subsequently  
791 also influenced their behavior in a different experimental test of behavioral flexibility and innovativeness.  
792 Grackles with intermediate sensitivities to learned associations solved fewer options on both multi-option  
793 puzzle boxes than grackles with either low or high sensitivities. Accordingly, the trained grackles not only  
794 changed their behavior within the specific serial reversal learning task, they also more generally changed  
795 their behavior across contexts in response to their training. Our findings show that grackles modulate their  
796 behavioral flexibility in response to the high reliability of cues and frequent changes in associations they  
797 experienced in the serial reversal learning experiment.

798 Applying the Bayesian reinforcement learning model to serial reversal data shows that participating in the  
799 serial reversal learning experiment made grackles change how much they value new information over old  
800 to update their associations, and how much they continue to explore alternative options or whether they  
801 are sensitive to the reward they are receiving at their current choice. Grackles coming into the experiment  
802 already had different rates of updating their associations and different sensitivities to learned associations,  
803 suggesting they had different experiences of how predictable cues are and how frequently their environment  
804 changes. In the urban environment they live in, changes are presumably frequent, so they would be expected  
805 to change their associations frequently (Lee & Thornton, 2021; Breen & Deffner, 2023). In line with this,  
806 the association-updating rate,  $\phi$ , appeared to explain more of the variation in how many trials individuals  
807 needed to reach the criterion of consistently choosing the rewarded option during a single phase as early  
808 as in their first reversal. Other recent applications of the Bayesian reinforcement learning model to serial  
809 reversal learning experiments also found that the association-updating rate explains more of the variation in  
810 the number of trials to pass criterion (squirrel monkeys Bari et al., 2022; mice Metha et al., 2020; Woo et al.,  
811 2023). In response to learning that the cues are highly reliable and the reversals are relatively frequent, the  
812 grackles increased their association-updating rate,  $\phi$ , which on average doubled across individuals, changing  
813 more for individuals who started off with lower  $\phi$  values. Grackles also changed their sensitivity to the  
814 learned associations,  $\lambda$ , during the serial reversals in line with the prediction that they benefit from being  
815 open to exploring the alternative option when the associations between cues and rewards switch frequently.  
816 Individuals changed their  $\phi$  and  $\lambda$  more if their initial values were further from those necessary to reach  
817 the passing criterion quickly. Individuals who passed their first reversal in 50 trials or less, changed  $\phi$  and  
818  $\lambda$  only slightly by the end of the serial reversal learning experiment. Among the trained grackles, who all  
819 required very few trials to consistently reach the criterion by the end of the experiment, we observed different  
820 approaches (see also Chen et al., 2021). Some individuals seemed more focused on the frequent changes, such  
821 that they kept exploring the alternative options and changed their associations as soon as they encountered  
822 new information. These individuals reached the passing criterion quickly because they switched to the newly  
823 rewarded option soon after a reversal. However, their continued exploration of the alternative option meant  
824 that they still needed several trials to reach the criterion. Other individuals seemed to place more emphasis  
825 on the reliability of the cues, focusing on the rewarded option after they learned that the cues had reversed.  
826 These individuals reached the passing criterion quickly because they consistently chose the rewarded option.  
827 However, these grackles needed a few more trials after a reversal began to switch to the new option. At the  
828 beginning of the experiment, the grackles showed a diversity of  $\phi$  and  $\lambda$  values and, because they had no  
829 prior experience, they did not show specific approaches to quickly reach the criterion. With the variables we  
830 measured at the beginning of the serial reversal learning experiment, we could not predict which approach  
831 grackles ended up with after the serial reversals.

832 The changes in behavioral flexibility that the grackles showed during the serial reversal learning experiment  
833 influenced their subsequent behavior in other tasks. The analyses linking  $\phi$  and  $\lambda$  to the performance on the  
834 multi-option puzzle boxes show that the different approaches grackles utilized to improve their performance  
835 during the serial reversal learning experiment subsequently appeared to influence how they solved the multi-  
836 option puzzle box. Grackles with intermediate  $\phi$  values showed shorter latencies to attempt a new option.  
837 This could reflect that grackles with high  $\phi$  values take longer because they formed very strong associations  
838 with the previously rewarded option, while grackles with small  $\phi$  values take longer because they either do  
839 not update their associations even though the first option is no longer rewarded or they do not explore as  
840 much due to their small  $\lambda$ . We also found that grackles with intermediate values of  $\lambda$  solved fewer puzzle  
841 box options. This could indicate that grackles with a small  $\lambda$  are more likely to explore new options, while

842 grackles with a large  $\lambda$  and low  $\phi$  are less likely to return to an option that is no longer rewarded. We are  
843 limited in our interpretation by the small sample sizes for the multi-option puzzle boxes. We have some  
844 indication that experiencing the serial reversal learning experiment continued to shape the behavior of the  
845 grackles after releasing them back to the wild. Individuals who changed their  $\phi$  and  $\lambda$  more during the serial  
846 reversal learning experiment appeared to switch more frequently between food types and foraging techniques  
847 (Logan et al., 2024). It took a grackle on average one month to pass the serial reversal learning experiment  
848 (Logan et al., 2023a), and the observations of the foraging behavior in the wild continued for up to 8 months  
849 after individuals were released (Logan et al., 2024). This indicates that the effects of enhancing flexibility are  
850 durable and generalize to other contexts. In grackles, behavioral flexibility does not change within days or  
851 only during certain critical periods. Our results suggest that individuals change their behavioral flexibility  
852 to match their environment if they experience the same conditions repeatedly across weeks.

853 Most individuals that have been tested in serial reversal learning experiments thus far show improvements  
854 throughout the reversals, suggesting that most species can modulate their behavioral flexibility in response  
855 to the predictability and stability of their environments (e.g. Warren & Warren, 1962; Komischke et al., 2002;  
856 Bond et al., 2007; Strang & Sherry, 2014; Chow et al., 2015; Cauchoix et al., 2017; Degrande et al., 2022;  
857 Erdsack et al., 2022). Previous studies used summary statistics to describe how the behavior of individuals  
858 changes during the serial reversal learning experiment (e.g. Federspiel et al., 2017) or show changes in learning  
859 curves (e.g. Gallistel et al., 2004). As shown in Figure 6, we can recreate these learning curves from the  
860 inferred association-updating rates and sensitivities to learned associations. The advantage of the Bayesian  
861 reinforcement learning model with its two parameters of the association-updating rate and the sensitivity to  
862 learned associations is that it has a clear theoretical foundation of what aspects of the experimental setting  
863 should lead to changes in the behavior (Gershman, 2018; Metha et al., 2020; Danwitz et al., 2022; Woo et  
864 al., 2023). Based on our application here, the model appears to be sufficient to accurately represent the  
865 behavior of grackles in the serial reversal experiment. This suggests that the stability and reliability of the  
866 environment has a large influence on how individuals learn about rewards. The importance of experiencing  
867 stable and predictable environments potentially explains the difference between lab-raised and wild-caught  
868 animals in how they change their behavior during the serial reversal learning experiment. Many lab-raised  
869 animals were observed to switch to a “win-stay versus lose-shift” strategy, where only their most recent  
870 experience guided their behavior and they no longer explored alternative options (Mackintosh et al., 1968;  
871 Rayburn-Reeves et al., 2013). These animals generally experience very stable conditions during their lives,  
872 and often participate in large numbers of trials in an experiment. Accordingly, cues are reliable and changes  
873 are rare, so individuals would be expected to show the high association-updating rates and high sensitivities  
874 to learned associations that would lead to the “win stay versus lose shift” strategy. In contrast, wild-  
875 caught animals, including grackles, only slowly move away when an option is no longer rewarded and they  
876 continue to explore alternative options (Chow et al., 2015; Cauchoix et al., 2017). These individuals probably  
877 experience environments in which associations are not perfectly reliable and changes occur more gradually.  
878 These individuals are expected to show smaller sensitivities to their associations and therefore continue to  
879 explore their environment. This focus on the key pieces of information that individuals likely pay attention  
880 to when adjusting their behavior also provides ways to link their performances and inferred cognitive abilities  
881 to their natural behavior. We found that, for the grackles, the behavioral flexibility they exhibited at the  
882 end of the serial reversal learning experiment linked to their foraging behavior in the wild (Logan et al.,  
883 2024). The existing literature on foraging behavior, investigating trade-offs between the exploration versus  
884 exploitation of different options, has a similar focus on gaining information (exploration) versus decision  
885 making (exploitation) (Kramer & Weary, 1991; Berger-Tal et al., 2014; Addicott et al., 2017). Linking this  
886 framework to the concepts of reinforcement learning and decision making could provide further insights into  
887 the cognitive processes that are involved and the information that individuals might pay attention to. The  
888 approach we established here to study behavioral flexibility, linking the theoretical framework of the Bayesian  
889 reinforcement learning model to the specific experimental task of the serial reversal learning experiment and  
890 the natural behavior of individuals, offers opportunities to better understand cognition in the wild (Rosati  
891 et al., 2022).

## 892 **Author contributions**

893 **Lukas:** Hypothesis development, simulation development, data analyses, data interpretation, write up,  
894 revising/editing.

895 **McCune:** Added MAB log experiment, protocol development, data collection, revising/editing.

896 **Blaisdell:** Prediction revision, revising/editing.

897 **Johnson-Ulrich:** Data collection, revising/editing.

898 **MacPherson:** Data collection, revising/editing.

899 **Seitz:** Prediction revision, revising/editing.

900 **Sevchik:** Data collection, revising/editing.

901 **Logan:** Hypothesis development, protocol development, data collection, data analysis, data interpretation,  
902 revising/editing.

## 903 **Funding**

904 This research is funded by the Department of Human Behavior, Ecology and Culture at the Max Planck  
905 Institute for Evolutionary Anthropology.

## 906 **Ethics**

907 The research on the great-tailed grackles followed established ethical guidelines for the involvement and treat-  
908 ment of animals in experiments and received institutional approval prior to conducting the study (US Fish  
909 and Wildlife Service scientific collecting permit number MB76700A-0,1,2; US Geological Survey Bird Band-  
910 ing Laboratory federal bird banding permit number 23872; Arizona Game and Fish Department scientific  
911 collecting license number SP594338 [2017], SP606267 [2018], and SP639866 [2019]; California Department  
912 of Fish and Wildlife scientific collecting permit number S-192100001-19210-001; Institutional Animal Care  
913 and Use Committee at Arizona State University protocol number 17-1594R; Institutional Animal Care and  
914 Use Committee at the University of California Santa Barbara protocol number 958; University of Cambridge  
915 ethical review process non-regulated use of animals in scientific procedures: zoo4/17 [2017]).

## 916 **Conflict of interest disclosure**

917 We, the authors, declare that we have no financial conflicts of interest with the content of this article. CJ  
918 Logan is a Recommender and, until 2022, was on the Managing Board at PCI Ecology. D Lukas is a  
919 Recommender at PCI Ecology.

## 920 **Acknowledgements**

921 We thank our PCI Ecology recommender, Aurelie Coulon, and reviewers, Maxime Dahirel and Andrea  
922 Griffin, for their feedback on the preregistration; Coulon for serving as the recommender on the post-study  
923 manuscript, and reviewers, Maxime Dahirel and an anonymous person, for their constructive feedback on  
924 this manuscript that helped with the framing of the study; Julia Cissewski for tirelessly solving problems  
925 involving financial transactions and contracts; Sophie Kaube for logistical support; and Richard McElreath  
926 for project support.



## References

- 927 Addicott MA, Pearson JM, Sweitzer MM, Barack DL, Platt ML (2017) A primer on foraging and  
928 the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology*, **42**, 1931–1939.  
929 <https://doi.org/10.1038/npp.2017.108>
- 931 Agrawal S, Goyal N (2012) Analysis of thompson sampling for the multi-armed bandit problem. In: *Confer-*  
932 *ence on learning theory*, pp. 39–1. JMLR Workshop; Conference Proceedings.
- 933 Bari BA, Moerke MJ, Jedema HP, Effinger DP, Cohen JY, Bradberry CW (2022) Reinforcement learning  
934 modeling reveals a reward-history-dependent strategy underlying reversal learning in squirrel monkeys.  
935 *Behavioral neuroscience*, **136**, 46. <https://doi.org/10.1037/bne0000492>
- 936 Bartolo R, Averbeck BB (2020) Prefrontal cortex predicts state switches during reversal learning. *Neuron*,  
937 **106**, 1044–1054. <https://doi.org/10.1016/j.neuron.2020.03.024>
- 938 Berger-Tal O, Nathan J, Meron E, Saltz D (2014) The exploration-exploitation dilemma: A multidisciplinary  
939 framework. *PloS one*, **9**, e95693. <https://doi.org/10.1371/journal.pone.0095693>
- 940 Bitterman ME (1975) The comparative analysis of learning: Are the laws of learning the same in all animals?  
941 *Science*, **188**, 699–709. <https://doi.org/10.1126/science.188.4189.699>
- 942 Blaisdell A, Seitz B, Rowney C, Folsom M, MacPherson M, Deffner D, Logan CJ (2021) Do the more flexible  
943 individuals rely more on causal cognition? Observation versus intervention in causal inference in great-  
944 tailed grackles (version 5 of this preprint has been peer reviewed and recommended by peer community  
945 in ecology [<https://doi.org/10.24072/pci.ecology.100076>]). <https://doi.org/10.31234/osf.io/z4p6s>
- 946 Bond AB, Kamil AC, Balda RP (2007) Serial reversal learning and the evolution of behavioral flexibility in  
947 three species of north american corvids (gymnorhinus cyanocephalus, nucifraga columbiana, aphelocoma  
948 californica). *Journal of Comparative Psychology*, **121**, 372. <https://doi.org/10.1037/0735-7036.121.4.372>
- 949 Boyce MS, Haridas CV, Lee CT, Group NSDW, *et al.* (2006) Demography in an increasingly variable world.  
950 *Trends in Ecology & Evolution*, **21**, 141–148.
- 951 Breen AJ, Deffner D (2023) Leading an urban invasion: Risk-sensitive learning is a winning strategy. *eLife*,  
952 **12**, RP89315. <https://doi.org/10.1101/2023.03.19.533319>
- 953 Camerer C, Hua Ho T (1999) Experience-weighted attraction learning in normal form games. *Econometrica*,  
954 **67**, 827–874. <https://doi.org/10.1111/1468-0262.00054>
- 955 Cauchoix M, Hermer E, Chaine A, Morand-Ferron J (2017) Cognition in the field: Comparison  
956 of reversal learning performance in captive and wild passerines. *Scientific reports*, **7**, 12945.  
957 <https://doi.org/10.1038/s41598-017-13179-5>
- 958 Chen CS, Knep E, Han A, Ebitz RB, Grissom NM (2021) Sex differences in learning from exploration. *Elife*,  
959 **10**, e69748. <https://doi.org/10.7554/elife.69748>
- 960 Chow PK, Leaver LA, Wang M, Lea SE (2015) Serial reversal learning in gray squirrels: Learning efficiency  
961 as a function of learning and change of tactics. *Journal of Experimental Psychology: Animal Learning*  
962 *and Cognition*, **41**, 343. <https://doi.org/10.1037/xan0000072>
- 963 Coulon A (2023) An experiment to improve our understanding of the link between behavioral flexibility and  
964 innovativeness. *Peer Community in Ecology*, **1**, 100407. <https://doi.org/10.24072/pci.ecology.100407>
- 965 Danwitz L, Mathar D, Smith E, Tuzsus D, Peters J (2022) Parameter and model recovery of reinforce-  
966 ment learning models for restless bandit problems. *Computational Brain & Behavior*, **5**, 547–563.  
967 <https://doi.org/10.1007/s42113-022-00139-0>
- 968 Daw ND, O’doherly JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions  
969 in humans. *Nature*, **441**, 876–879. <https://doi.org/10.1038/nature04766>
- 970 Degrande R, Cornilleau F, Lansade L, Jardat P, Colson V, Calandreau L (2022) Domestic hens succeed at  
971 serial reversal learning and perceptual concept generalisation using a new automated touchscreen device.  
972 *animal*, **16**, 100607. <https://doi.org/10.1016/j.animal.2022.100607>
- 973 Donaldson-Matasci MC, Bergstrom CT, Lachmann M (2013) When unreliable cues are good enough. *The*  
974 *American Naturalist*, **182**, 313–327.
- 975 Dufort RH, Guttman N, Kimble GA (1954) One-trial discrimination reversal in the white rat. *Journal of*  
976 *Comparative and Physiological Psychology*, **47**, 248. <https://doi.org/10.1037/h0057856>
- 977 Dunlap AS, Stephens DW (2009) Components of change in the evolution of learning and un-  
978 learned preference. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 3201–3208.  
979 <https://doi.org/10.1098/rspb.2009.0602>
- 980 Erdsack N, Dehnhardt G, Hanke FD (2022) Serial visual reversal learning in harbor seals (*phoca vitulina*).

981 *Animal Cognition*, **25**, 1183–1193. <https://doi.org/10.1007/s10071-022-01653-1>

982 Federspiel IG, Garland A, Guez D, Bugnyar T, Healy SD, Güntürkün O, Griffin AS (2017) Adjusting foraging  
983 strategies: A comparison of rural and urban common mynas (*acridotheres tristis*). *Animal cognition*, **20**,  
984 65–74. <https://doi.org/10.1007/s10071-016-1045-7>

985 Frömer R, Nassar M (2023) Belief updates, learning and adaptive decision making. <https://doi.org/10.31234/osf.io/qndba>

986 Gallistel CR, Fairhurst S, Balsam P (2004) The learning curve: Implications of a quantitative analysis. *Pro-*  
987 *ceedings of the National Academy of Sciences*, **101**, 13124–13131. <https://doi.org/10.1073/pnas.0404965101>

988 Gelman A, Rubin DB (1995) Avoiding model selection in bayesian social research. *Sociological methodology*,  
989 **25**, 165–173. <https://doi.org/10.2307/271064>

990 Gershman SJ (2018) Deconstructing the human algorithms for exploration. *Cognition*, **173**, 34–42.  
991 <https://doi.org/10.1016/j.cognition.2017.12.014>

992 Izquierdo A, Brigman JL, Radke AK, Rudebeck PH, Holmes A (2017) The neural basis of reversal learning:  
993 An updated perspective. *Neuroscience*, **345**, 12–26. <https://doi.org/10.1016/j.neuroscience.2016.03.021>

994 Komischke B, Giurfa M, Lachnit H, Malun D (2002) Successive olfactory reversal learning in honeybees.  
995 *Learning & memory*, **9**, 122–129. <https://doi.org/10.1101/lm.44602>

996 Kramer DL, Weary DM (1991) Exploration versus exploitation: A field study of time allocation to environ-  
997 mental tracking by foraging chipmunks. *Animal Behaviour*, **41**, 443–449. [https://doi.org/10.1016/s0003-3472\(05\)80846-2](https://doi.org/10.1016/s0003-3472(05)80846-2)

998

999 Lea SE, Chow PK, Leaver LA, McLaren IP (2020) Behavioral flexibility: A review, a model, and some  
1000 exploratory tests. *Learning & Behavior*, **48**, 173–187. <https://doi.org/10.3758/s13420-020-00421-w>

1001 Lee VE, Thornton A (2021) Animal cognition in an urbanised world. *Frontiers in Ecology and Evolution*, **9**,  
1002 120.

1003 Leimar O, Quiñones AE, Bshary R (2024) Flexible learning in complex worlds. *Behavioral Ecology*, **35**,  
1004 arad109. <https://doi.org/10.1093/beheco/arad109>

1005 Liu Y, Day LB, Summers K, Burmeister SS (2016) Learning to learn: Advanced behavioural flexibility in a  
1006 poison frog. *Animal Behaviour*, **111**, 167–172. <https://doi.org/10.1016/j.anbehav.2015.10.018>

1007 Lloyd K, Leslie DS (2013) Context-dependent decision-making: A simple bayesian model. *Journal of The*  
1008 *Royal Society Interface*, **10**, 20130069.

1009 Logan C, Lukas D, Blaisdell A, Johnson-Ulrich Z, MacPherson M, Seitz B, Sevchik A, McCune K (2023a)  
1010 Behavioral flexibility is manipulable and it improves flexibility and innovativeness in a new context. *Peer*  
1011 *Community Journal*, **3**. <https://doi.org/10.24072/pcjournal.284>

1012 Logan C, Lukas D, Blaisdell A, Johnson-Ulrich Z, MacPherson M, Seitz B, Sevchik A, McCune K (2023b)  
1013 Data: Behavioral flexibility is manipulable and it improves flexibility and problem solving in a new  
1014 context. *Knowledge Network for Biocomplexity*, **Data package**. <https://doi.org/10.5063/F1BR8QNC>

1015 Logan C, Lukas D, Geng X, LeGrande-Rolls C, Marfori Z, MacPherson M, Rowney C, Smith C, McCune  
1016 K (2024) Behavioral flexibility is related to foraging, but not social or habitat use behaviors in a species  
1017 that is rapidly expanding its range. *EcoEvoRxiv*. <https://doi.org/10.32942/X2T036>

1018 Logan CJ, McCune KB, LeGrande-Rolls C, Marfori Z, Hubbard J, Lukas D (2023c) Implementing  
1019 a rapid geographic range expansion - the role of behavior changes. *Peer Community Journal*.  
1020 <https://doi.org/10.24072/pcjournal.320>

1021 Logan CJ, Shaw R, Lukas D, McCune KB (2022) How to succeed in human modified environments. *In princi-*  
1022 *ple acceptance by PCI Ecology of the version on 8 Sep 2022*. <https://doi.org/https://doi.org/10.17605/OSF.IO/346AF>

1023 Mackintosh N, McGonigle B, Holgate V (1968) Factors underlying improvement in serial reversal learning.  
1024 *Canadian Journal of Psychology/Revue canadienne de psychologie*, **22**, 85. <https://doi.org/10.1037/h0082753>

1025 McCune K, Blaisdell A, Johnson-Ulrich Z, Sevchik A, Lukas D, MacPherson M, Seitz B, Logan C (2023)  
1026 Repeatability of performance within and across contexts measuring behavioral flexibility. *PeerJ*.  
1027 <https://doi.org/10.7717/peerj.15773>

1028 McElreath R (2020) *Rethinking: Statistical rethinking book package*.

1029 Metha JA, Brian ML, Oberrauch S, Barnes SA, Featherby TJ, Bossaerts P, Murawski C, Hoyer D, Jacobson  
1030 LH (2020) Separating probability and reversal learning in a novel probabilistic reversal learning task for  
1031 mice. *Frontiers in behavioral neuroscience*, **13**, 270.

1032 Mikhalevich I, Powell R, Logan C (2017) Is behavioural flexibility evidence of cognitive complexity? How evo-  
1033 lution can inform comparative cognition. *Interface Focus*, **7**, 20160121. <https://doi.org/10.1098/rsfs.2016.0121>

1034 Minh Le N, Yildirim M, Wang Y, Sugihara H, Jazayeri M, Sur M (2023) Mixtures of strategies

1035 underlie rodent behavior during reversal learning. *PLOS Computational Biology*, **19**, e1011430.  
1036 <https://doi.org/10.1371/journal.pcbi.1011430>

1037 Neftci EO, Averbek BB (2019) Reinforcement learning in artificial and biological systems. *Nature Machine*  
1038 *Intelligence*, **1**, 133–143.

1039 R Core Team (2023) *R: A language and environment for statistical computing*. R Foundation for Statistical  
1040 Computing, Vienna, Austria.

1041 Rayburn-Reeves RM, Stagner JP, Kirk CR, Zentall TR (2013) Reversal learning in rats (*rattus norvegicus*)  
1042 and pigeons (*columba livia*): Qualitative differences in behavioral flexibility. *Journal of Comparative*  
1043 *Psychology*, **127**, 202. <https://doi.org/10.1037/a0026311>

1044 Rescorla RA, Wagner AR (1972) A theory of pavlovian conditioning: Variations in the effectiveness of  
1045 reinforcement and nonreinforcement. In: *Classical conditioning II: Current theory and research* (eds  
1046 Black AH, Prosy WF), pp. 64–99. Appleton-Century-Crofts, New York.

1047 Rosati AG, Machanda ZP, Slocombe KE (2022) Cognition in the wild: Understanding animal thought in its  
1048 natural context. *Current Opinion in Behavioral Sciences*, **47**.

1049 Shettleworth SJ (2010) *Cognition, evolution, and behavior*. Oxford university press.

1050 Sih A (2013) Understanding variation in behavioural responses to human-induced rapid environmental  
1051 change: A conceptual overview. *Animal Behaviour*, **85**, 1077–1088.

1052 Sol D, Timmermans S, Lefebvre L (2002) Behavioural flexibility and invasion success in birds. *Animal*  
1053 *behaviour*, **63**, 495–502. <https://doi.org/10.1006/anbe.2001.1953>

1054 Spence KW (1936) The nature of discrimination learning in animals. *Psychological review*, **43**, 427.  
1055 <https://doi.org/10.1037/h0056975>

1056 Stan Development Team (2023) *Stan modeling language users guide and reference manual, version 2.32.0*,  
1057 <https://mc-stan.org/>.

1058 Starrfelt J, Kokko H (2012) Bet-hedging—a triple trade-off between means, variances and correlations. *Bi-*  
1059 *ological Reviews*, **87**, 742–755.

1060 Strang CG, Sherry DF (2014) Serial reversal learning in bumblebees (*bombus impatiens*). *Animal Cognition*,  
1061 **17**, 723–734. <https://doi.org/10.1007/s10071-013-0704-1>

1062 Tello-Ramos MC, Branch CL, Kozlovsky DY, Pitera AM, Pravosudov VV (2019) Spatial memory and cog-  
1063 nitive flexibility trade-offs: To be or not to be flexible, that is the question. *Animal Behaviour*, **147**,  
1064 129–136. <https://doi.org/10.1016/j.anbehav.2018.02.019>

1065 Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C (2021) Rank-normalization, folding, and  
1066 localization: An improved rhat for assessing convergence of MCMC (with discussion). *Bayesian Analysis*.  
1067 <https://doi.org/10.1214/20-BA1221>

1068 Warren J (1965a) Primate learning in comparative perspective. *Behavior of nonhuman primates*, **1**, 249–281.  
1069 <https://doi.org/10.1016/B978-1-4832-2820-4.50014-7>

1070 Warren JM (1965b) The comparative psychology of learning. *Annual review of psychology*, **16**, 95–118.  
1071 <https://doi.org/10.1146/annurev.ps.16.020165.000523>

1072 Warren J, Warren HB (1962) Reversal learning by horse and raccoon. *The Journal of Genetic Psychology*,  
1073 **100**, 215–220. <https://doi.org/10.1080/00221325.1962.10533590>

1074 Woo JH, Aguirre CG, Bari BA, Tsutsui K-I, Grabenhorst F, Cohen JY, Schultz W, Izquierdo A,  
1075 Soltani A (2023) Mechanisms of adjustments to different types of uncertainty in the reward  
1076 environment across mice and monkeys. *Cognitive, Affective, & Behavioral Neuroscience*, 1–20.  
1077 <https://doi.org/10.1101/2022.10.01.510477>

1078 Wright TF, Eberhard JR, Hobson EA, Avery ML, Russello MA (2010) Behavioral flexibility and  
1079 species invasions: The adaptive flexibility hypothesis. *Ethology Ecology & Evolution*, **22**, 393–404.  
1080 <https://doi.org/10.1080/03949370.2010.505580>