



## Great-tailed grackle research reveals need for researchers to consider their own flexibility and test limitations in cognitive test batteries.

**Aliza le Roux** based on reviews by *Pizza Ka Yee Chow* and *Alex DeCasian*  
A recommendation of:

Are the more flexible great-tailed grackles also better at behavioral inhibition?  
Logan CJ, McCune KB, MacPherson M, Johnson-Ulrich Z, Rowney C, Seitz B, Blaisdell AP, Deffner D, Wascher CAF (2021), *PsyArXiv*, vpc39, ver. 7 peer-reviewed and recommended by Peer community in Ecology [10.31234/osf.io/vpc39](https://doi.org/10.31234/osf.io/vpc39)

### Open Access

Submitted: 04 December 2020, Recommended: 04 May 2021

Published: 4 May 2021

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

#### Cite this recommendation as:

Aliza le Roux (2021) Great-tailed grackle research reveals need for researchers to consider their own flexibility and test limitations in cognitive test batteries.. *Peer Community in Ecology*, 100081. [10.24072/pci.ecology.100081](https://doi.org/10.24072/pci.ecology.100081)

### Recommendation

In the article, "Are the more flexible great-tailed grackles also better at behavioral inhibition?", Logan and colleagues (2021) are setting an excellent standard for cognitive research on wild-caught animals. Using a decent sample (N=18) of wild-caught birds, they set out to test the ambiguous link between behavioral flexibility and behavioral inhibition, which is supported by some studies but rejected by others. Where this study is more thorough and therefore also more revealing than most extant research, the authors ran a battery of tests, examining both flexibility (reversal learning and solution switching) and inhibition (go/no go task; detour task; delay of gratification) through multiple different test series. They also -- somewhat accidentally -- performed their experiments and analyses with and without different criteria for correctness (85%, 100%). Their mistakes, assumptions and amendments of plans made during preregistration are clearly stated and this demonstrates the thought-process of the researchers very clearly.

Logan et al. (2021) show that inhibition in great-tailed grackles is a multi-faceted construct, and demonstrate that the traditional go/no go task likely tests a very different aspect of inhibition than the detour task, which was never linked to any of their flexibility measures. Their comprehensive Bayesian analyses held up the results of some of the frequentist statistics, indicating a consistent relationship between flexibility and inhibition, with more flexible individuals also showing better inhibition (in the go/no go task). This same model, combined with inconsistencies in the GLM analyses (depending on the inclusion or exclusion of an outlier), led them to recommend caution in the

creation of arbitrary thresholds for "success" in any cognitive tasks. Their accidental longer-term data collection also hinted at patterns of behaviour that shorter-term data collection did not. Of course, researchers have to decide on success criteria in order to conduct experiments, but in the same way that frequentist statistics are acknowledged to have flaws, the setting of success criteria must be acknowledged as inherently arbitrary. Where possible, researchers could reveal novel, biologically salient patterns by continuing beyond the point where a convenient success criterion has been reached. This research also underscores that tests may not be examining the features we expected them to measure, and are highly sensitive to biological and ecological variation between species as well as individual variation within populations.

To me, this study is an excellent argument for pre-registration of research (registered as Logan et al. 2019 and accepted by Vogel 2019), as the authors did not end up cherry-picking only those results or methods that worked. The fact that some of the tests did not "work", but was still examined, added much value to the study. The current paper is a bit densely written because of the comprehensiveness of the research. Some editorial polishing would likely make for more elegant writing. However, the arguments are clear, the results novel, and the questions thoroughly examined. The results are important not only for cognitive research on birds, but are potentially valuable to any cognitive scientist. I recommend this article as excellent food for thought.

## References

Logan CJ, McCune K, Johnson-Ulrich Z, Bergeron L, Seitz B, Blaisdell AP, Wascher CAF. (2019) Are the more flexible individuals also better at inhibition? [http://corinalogan.com/Preregistrations/g\\_inhibition.html](http://corinalogan.com/Preregistrations/g_inhibition.html) In principle acceptance by PCI Ecology of the version on 6 Mar 2019

Logan CJ, McCune KB, MacPherson M, Johnson-Ulrich Z, Rowney C, Seitz B, Blaisdell AP, Deffner D, Wascher CAF (2021) Are the more flexible great-tailed grackles also better at behavioral inhibition? PsyArXiv, ver. 7 peer-reviewed and recommended by Peer community in Ecology. <https://doi.org/10.31234/osf.io/vpc39>

Vogel E (2019) Adapting to a changing environment: advancing our understanding of the mechanisms that lead to behavioral flexibility. Peer Community in Ecology, 100016. <https://doi.org/10.24072/pci.ecology.100016>

## Reviews

Toggle reviews

---

## *Revision round #1*

2021-03-19

### *Author's Reply*

Dear Dr.'s le Roux, Chow, and DeCasian,

Thank you so much for your extremely useful feedback on our manuscript! We really appreciate the time you have taken through this whole process, which started a couple of years ago. You really have made this research better. We responded to your comments below and revised the manuscript accordingly (PDF version 4 at <https://doi.org/10.31234/osf.io/vpc39> or HTML at [http://corinalogan.com/Preregistrations/g\\_inhibition.html](http://corinalogan.com/Preregistrations/g_inhibition.html)).

Note that the version-tracked version of this manuscript is in rmarkdown at GitHub: [https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/g\\_inhibition.Rmd](https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/g_inhibition.Rmd). In case you want to see the history of track changes for this document, click the link and then click the “History” button on the right near the top. From there, you can scroll through our comments on what was changed for each save event and, if you want to see exactly what was changed, click on the text that describes the change and it will show you the text that was replaced (in red) next to the new text (in green).

All our best,

Corina (on behalf of all co-authors)

## Round

#1

by Aliza le Roux, 2021-03-19 09:41

Manuscript: <https://doi.org/10.31234/osf.io/vpc39> version 3

Flexibility and inhibition in great-tailed grackles

**\*\*COMMENT 1:\*\*** The current manuscript is well-written and clear, with well-defined hypotheses and experimental set-ups. The results and intellectual contribution to the field are highly worthwhile and publishable. Both reviewers recommended changes mostly related to clarity, and I agree with them.

What stops me from recommending the article at this point is a valid question that both reviewers are asking: How would the results change if the outlier grackle (Taquito) were removed from the analyses? The authors note the unusual behaviour in this individual but do not assess how his inclusion or exclusion would affect overall outcomes. In agreement with the reviewers, I therefore ask that analyses be done with that individual removed, to assess the impact of this clear outlier. Likely, it would make sense to still retain the original analyses as well as the outcomes with Taquito removed.

Apart from that point, the rest of the suggested changes are mostly minor and would not necessitate another round of reviews. Thank you for your patience with the review process, which has been slower than usual due to COVID-related challenges and some technical glitches.

> **\*\*RESPONSE 1:\*\*** Thank you for your evaluation and direction! We address these issues, including Taquito, below. Absolutely no worries about any delays - we were also very busy during this time and totally understand.

Reviewed by Pizza Ka Yee Chow, 2021-03-12 19:44

**\*\*COMMENT 2:\*\*** Thank you for inviting me to review this paper. The authors examined the relationship between behavioural flexibility (defined as the ability to inhibit a previously learned behaviour) and inhibitory control using multiple cognitive tasks established in the field. They found that the go/no-go and detour inhibition tasks showed a positive correlation with the number of trial used to reach the learning criterion in the reversal learning task and a negative correlation with the mean latency to switch to an alternative solution after the proficient solution is blocked in the multi-access box (MAB).

I fully support this investigation and think that it is much needed in the field to clarify what actually we are measuring when using different tasks. I applaud the hard work that the authors paid for the study and totally

agree with them that given the mixed findings on the existing literature, we shall use several tasks to examine the relationship between flexibility and inhibitory control. It is a shame that the grackles did not fully habituate to the the delay gratification apparatus, and I hope there would opportunity to conduct this task later.

I am mostly happy with the study, and i only have a couple of minor comments, which I hope they would be useful for the revision:

> **RESPONSE 2:** Thank you very much for your support! And also for your comments, which we address below.

**COMMENT 3:** Introduction. I like the introduction very much - very concise and straight to the point, with a clear study goal that has proper rationale behind. Line 73-87 When describing each task, it may be more straightforward to relate how each task measure inhibitory control/ flexibility consistently. For example, whether successfully not pecking the no-go stimulus indicates higher inhibitory control - it is a matter of phrasing...

> **RESPONSE 3:** Great point. We added these clarifications:

INTRODUCTION: "The go/no go experiment consisted of two different shapes sequentially presented on a touchscreen where one shape must be pecked to receive a food reward (automatically provided by a food hopper under the screen) and the other shape must not be pecked **(indicating more inhibitory control)** or there will be a penalty of a longer intertrial interval **(indicating less inhibitory control)**. In the detour task, individuals are assessed on their ability to inhibit the motor impulse to try to reach a reward through the long side of a transparent cylinder **(indicating less inhibitory control)**, and instead to detour and take the reward from an open end **(indicating more inhibitory control)** [ @kabadayi2018detour; methods as in @maclean2014evolution who call it the "cylinder task"]. In the delay of gratification task, grackles must wait longer **(indicating more inhibitory control)** for higher quality (more preferred) food or for higher quantities [methods as in @hillemann2014waiting]. The reversal learning of a color preference task involved one reversal (half the birds) or serial reversals **(to increase flexibility)**; half the birds) of a light gray and a dark gray colored tube, one of which contained a food reward [the experiments and data are in @logan2019flexmanip]. **Those grackles that are faster to reverse are more flexible.** The multi-access box experimental paradigm is modeled after @auersperg\_flexibility\_2011 and consists of four different access options to obtain food where each option requires a different type of action to solve it [the experiments and data are in @logan2019flexmanip]. Once a grackle passes criterion for demonstrating proficiency in solving an option, that option becomes non-functional in all future trials. The measure of flexibility is the latency to switch to attempting a new option after a proficient option becomes non-functional, **with shorter latencies indicating more flexibility.**"

**COMMENT 4:** Line 88 The introduction so far write it in a way that it is not certain about the relationship between flexibility and inhibitory control, so I would keep it more conservative for the hypothesis as in 'Employing several experimental assays to measure flexibility and inhibition supports a rigorous approach to testing whether the two trait are linked.' Unless you are specifically testing the alternative (hypothesis 1)?

> **RESPONSE 4:** You are correct - the point is that we use multiple measures of flexibility and multiple measures of inhibition to determine whether flexibility and inhibition are linked. We changed the text as you suggested.

**COMMENT 5:** Line 105. Well, it is good that the authors set a very stringent criterion for the grackles.

But we also know that grackles, like most if not all other animals rarely (consistently) do a perfect score...and the authors also noted in line 196-197 that such perfect criterion is not entirely ecological relevant.

> **RESPONSE 5:** Thank you for bringing this topic up - it was a point of confusion for us that would be good to clarify further because we realize now that we forgot to update the Deviations from the Preregistration section with this explanation. What happened was that we had two versions of the passing criterion: 85% correct was listed in our preregistration, and 100% correct by trial 150 or 85% correct thereafter was listed in the testing protocol associated with our preregistration. Because experimenters use the testing protocol when they are running the tests, the 150 trial criterion was the criterion we used to determine whether and when a bird passed the experiment. Logan is mystified as to how she included and kept the 150 trial criterion and cannot find any references that use this criterion (though had a memory that there was one reference at one point, which is where this criterion came from). Because we were confused, we failed to realize that the 85% criterion was not a post-hoc passing criterion, but actually the one we preregistered. We realized that the Deviations from the Preregistration section did not accurately reflect this, therefore we revised it to say:

“1) Jan 2020 go/no go performance: the preregistration listed the passing criterion as the number of trials to reach 85% correct, while the protocol associated with the preregistration that the experimenters used when testing the birds listed the passing criterion as 100% correct within 150 trials or 85% correct between 150-200 trials. Therefore, we tested birds according to the latter criterion and conducted all analyses for both criteria. It was previously not specified over what number of trials 85% accuracy was calculated, therefore we decided to calculate it at the level of the most recent sliding 10 trial block (i.e., the most recent 10 trials, regardless of whether it is an even 20, 30, 40 trials).”

Thanks to your comment, we also realized that on line 105 we stated it was reversal performance, however that was a typo and we changed this to “1) Jan 2020 go/no go performance:”

**COMMENT 6:** Results. Line 149. This sub-heading only partial cover what this section is actually about. May be something like ‘relationship between reversal learning and go/no-go task’

> **RESPONSE 6:** Good point. We updated this to: “Relationship between go/no go (inhibition) and reversal learning (flexibility)”

**COMMENT 7:** Line 151. Hang on, so the analyses of the reversal learning had only included those only had serial reversal learning, right? please clarify. Also some descriptive results about the number of trials that grackles took to reach the criterion in the last reversal phase would be useful.

> **RESPONSE 7:** Good points! We revised the text as follows:

RESULTS > Prediction 1 > Model 2a > Relationship between go/no go and reversal learning:

There was a positive correlation between the number of trials to pass criterion in the go/no go experiment and the number of trials to reverse a preference **(average=59 trials, standard deviation=41, range=23-160 trials, n=9 grackles)** in the colored tube reversal experiment (in their **last reversal**, **thus for the control grackles, this was their first and only reversal, while for the manipulated grackles, this was their last reversal in the serial reversal manipulation**) when using one of the two go/no go passing criteria: the number of trials to reach 85% correct (measured in the most recent 20 trial block; **(measured in the most recent 20 trial block; average=149 trials, standard deviation=71, range=60-290 trials, n=9 grackles; Table 2a, Figure 1)**). The other passing criterion of achieving 100% correct performance by trial 150, and if this is not met then they pass when they reach 85% correct after trial 150 (measured in the most recent 20 trial

block; **average=178 trials, standard deviation=15, range=160-200 trials**) did not correlate with reversal performance.

**COMMENT 8:** Line 197-200 I agree with the authors that the reversal learning task and the go/no-go task share a lot of similarities. This may explain why the same type of inhibitory control is correlated between the two tasks. But this bit of information is better to be placed in the discussion.

> **RESPONSE 8:** We moved this paragraph to the discussion.

**COMMENT 9:** Line 201-204. These are descriptive results of the go/no-go task, should they presented earlier in the section before, say in line 151?

> **RESPONSE 9:** Good point! We made the change (see Response 7 for the revision).

**COMMENT 10:** Line 206-207. So what if the analysis exclude Taquito in both tasks, does the trend still hold?

> **RESPONSE 10:** Please see our Response 23.2 below.

**COMMENT 11:** Line 260 Some descriptive results about the detour task?

> **RESPONSE 11:** We originally included this as the last sentence in this paragraph, but now we moved them to a more prominent location in the first sentence:

RESULTS > Prediction 1 > Relationship between go/no go (inhibition) and reversal learning (flexibility):

“There was no correlation between the proportion correct on the detour experiment **(average=0.71, standard deviation=0.25, range=0.20-1.00, n=18 grackles)**”

**COMMENT 12:** Line 223&261 again, the subheading only partially reflect what the section is about

> **RESPONSE 12:** We changed these to:

Relationship between go/no go (inhibition) and reversal learning (flexibility)

Relationship between detour (inhibition) and reversal learning (flexibility)

**COMMENT 13:** Line 306 some information about the MAB performance?

> **RESPONSE 13:** We added these descriptive stats to the paragraph that describes that the two flexibility experiments did not correlate:

RESULTS > Prediction 1 > Relationship between go/no go (inhibition) and multi-access box (flexibility):

“On the plastic multi-access box, the average of the average latency per bird to attempt a new solution was 167 seconds (standard deviation=188, range=25-502, n=7 grackles). On the log multi-access box, the average of the average latency per bird to attempt a new solution was 513 seconds (standard deviation=544, range=77-1482, n=8 grackles).”

**\*\*COMMENT 14:\*\*** Discussion. Line 349-350. The findings here support the go/no-go task and the detour task are not measuring the same inhibitory control. Would there be also other factors at play here? For example, the go/no-go requires entails much more costs than the detour task – a peck on the no-go resulted in a longer inter-trial interval and no reward whereas the detour will end up having a reward anyway...

> **\*\*RESPONSE 14:\*\*** Thank you, this is an interesting point. This is indeed a possibility. We have added some discussion on this.

DISCUSSION: “A major concern associated with the comparison of performance on inhibition tasks is that measures are not always consistent when different experimental paradigms are used [addressi2013delay; @brucks2017measures; @van2018detour], which is further confirmed by our findings. **\*\*One potential explanation could be the cost-benefit trade-off associated with the task where certain contexts affect performance in inhibitory control tasks [bray2014context; brucks2017measures]. In the detour task, the cost of failing a trial is very low - the individual will receive the food reward regardless of their first touch. Whereas the cost is higher in the go/no go task, where individuals did not receive food and had to wait longer before the next trial started when they pecked the no go key.\*\*** This indicates that it is crucial to compare inhibition paradigms with each other on the same individuals to understand whether and how they relate to each other and in which contexts.”

**\*\*COMMENT 15:\*\*** Line 373-374 well, as the authors have mentioned previously about the latency to peck the screen data...it is a shame that it could not be analysed...would the authors have suggestions regarding how to modify the paradigm so that future studies that use touch screen for this task could also record the latency on no-go stimulus?

> **\*\*RESPONSE 15:\*\*** Unfortunately, we are not able to get latencies for all no go trials on the touchscreen because, if they make the correct choice, they do not touch the screen during the 10 s the stimulus is present. To get some kind of latency for delay trials where the individual makes the correct choice, we would have to conduct a different experiment. For example, in the delay of gratification experiment, we can calculate the latency to give up on waiting to eat the food for each trial.

**\*\*COMMENT 16:\*\*** For all figures - Indicate sample size

> **\*\*RESPONSE 16:\*\*** We implemented this change.

**\*\*COMMENT 17:\*\*** Table 2, 3. Why used GLM if the authors would like to show individual variation in the two trait? isn't using GLMM and setting individual identity as random variable more appropriate for the analyses?

> **\*\*RESPONSE 17:\*\*** We used a GLM because we there was only one row of data per bird (i.e., number of trials to pass criterion, average number of seconds to switch loci), therefore we wouldn't be able to have individual as a random effect.

Reviewed by Alex DeCasian, 2021-03-11 15:17  
Review of the preprint “Are the more flexible great-tailed grackles also better at behavioral inhibition?”

**\*\*COMMENT 18:\*\*** This is an interesting manuscript which can improve our understanding of a topic critical to behavioral ecology – the role that inhibition may (or may not) play in the evolution of behavioral flexibility. The authors investigate this by examining relationships between grackle performance on two inhibition tasks and two flexibility tasks. I recommend that this manuscript be published following major revisions. My main concerns include: 1) the effects of an outlier individual who was accidentally measured for too many trials; 2) the organization of the Results section; and 3) the lack of explanation/discussion of one of the significant findings. My recommendations/questions are outlined below by section:

Throughout: When discussing the relationship between inhibition and flexibility, I recommend mentioning them in this order since it follows the order of proposed causation (e.g., Lines 33-35, 54-55, 91, 147).

> **\*\*RESPONSE 18:\*\*** Thank you for your feedback! We changed the order so that in the subheadings, inhibition comes before flexibility, which helps improve the flow. We also implemented this order in the introduction (except for the first paragraph), results (except for the Bayesian model section which used reversal learning parameters), and discussion (except for one sentence each in the first and last paragraph, which follows the order in the article title). However, we were careful not to hypothesize about which trait (flexibility or inhibition) causes which because this is a correlational study and cannot test causality.

**\*\*COMMENT 19:\*\*** Abstract: Lines 37-42: These two sentences provide the same information – one could be cut to shorten the abstract.

> **\*\*RESPONSE 19:\*\*** Good point. We deleted the first of these two sentences.

**\*\*COMMENT 20:\*\*** Lines 45-46: The authors should mention that the inhibition tasks did not correlate with each other before discussing their correlations (or lack thereof) with the flexibility measures.

> **\*\*RESPONSE 20:\*\*** Great suggestion! We implemented it such that the ABSTRACT now reads: “Performance on the go/no go and detour inhibition tests did not correlate with each other, indicating that they did not measure the same trait. Individuals who were faster to update their behavior in the reversal experiment were also faster to reach criterion in the go/no go task, but took more time to attempt a new option in the multi-access box experiment.”

**\*\*COMMENT 21:\*\*** Results: The way this section is currently organized is confusing. I recommend the author put all Prediction 1 sections under the same heading, with subheadings indicating the comparison being made (e.g., detour v reversal learning, # go/no go trials v multi-access box, etc.).

> **\*\*RESPONSE 21:\*\*** Yes, this makes sense. We made the change.

**\*\*COMMENT 22:\*\*** Variables can be more clearly defined throughout.

> **\*\*RESPONSE 22:\*\*** We clarified definitions of variables as described in Responses 3, 6, and 7 above.

**\*\*COMMENT 23:\*\*** Prediction 1, go/no go and Reversal Learning --

23.1: Line 150 – Please provide p-values and/or slope estimates for “positive correlation” (even though they are provided in Table 2).



23.2: I am very concerned that the reported relationships would not hold if the individual who was accidentally tested too many times is removed -- doesn't seem like they would from Figure 1? Although this is acknowledged in Lines 206-207, the authors should provide additional analyses excluding this individual (or capping their # of trials).

23.3: It should be made clearer how a negative relationship between the number of go/no go trials and reversal learning rate (Lines 166-168) confirms the finding of a positive relationship between the number of go/no go trials and number of trials to reverse a preference (Lines 150-151).

23.4: Lines 194-200 – this paragraph is not a result and should be moved to the “deviations from preregistration” section as justification for the new cutoff.

23.5: Lines 201-294 – These summary statistics should come earlier in this section.

> **\*\*RESPONSE 23:\*\***

23.1: If we did this here, we would need to do it throughout and we think the text is already too busy with the addition of the reversal and multi-access box summary statistics to these sentences (see Responses 7, 12, and 13). Therefore, we decided not to make the change, but we made sure to reference the appropriate table when we mention the correlations in the text.

23.2: We explicitly stated in the preregistration that we would not exclude any data (Analysis Plan), however we do have grounds to exclude Taquito because he was not supposed to be tested past 200 trials, which is the cap we set. Holding to this cap makes the individuals more comparable because other grackles were not given the chance to go beyond 200 trials - their experiment simply ended and they were not included in the analyses. Therefore, as requested by Dr. le Roux, we replicated the go/no go analyses such that there is one version that includes Taquito, and one version that excludes him. See the new and changed text below, as well as the new tables 2b and 3b and the updated figure 1 (to show first reversals in addition to last reversals).

23.3: Thank you. Birds that take fewer trials to reverse their preference are likely to be characterised by higher learning rates in the reinforcement learning model. Therefore, a negative relationship between learning rate and go/no go trials and a positive relationship between reversal trials and go/no go trials both reflect the same tendency of more flexible individuals performing better on the inhibition go/no go task. We added a sentence in the Results to clarify this relationship: “This confirms the positive relationship between numbers of trials to reverse a preference and trials to reach criterion in the go/no go task, because fewer trials to reverse preferences tend to be reflected in higher learning rates in the computational model.”

23.4: Your point touches on something we needed to clarify further. Please see our Responses 5 and 8 where we describe how we updated the Deviations from the Preregistration and that we moved this text to the Discussion.

23.5: Good point! Please see our Response 7 for how we addressed this.

RESULTS > Model 2a > Relationship between go/no go (inhibition) and reversal learning (flexibility):

“Regardless of the criterion used, we capped the number of trials for the go/no go experiment at 200, with the exception of 2 individuals who were tested past trial 200 **\*\*due to experimenter error (Mofongo continued to trial 249 and did not pass the 85% criterion; and Taquito continued to trial 290 and passed the 85% criterion).** We repeated the above analyses for the 85% criterion using a data set without Taquito because this would make the individuals more comparable as not all grackles were given the chance to pass criterion after trial 200.

**Results for the analyses without Taquito showed that, instead of a positive correlation, there was a negative correlation between the number of trials to pass criterion in the go/no go experiment and the**

number of trials to reverse a preference in the colored tube reversal experiment (in their **\*\*last reversal\*\***; average=47, standard deviation=17, range=23-80, n=8 grackles) using the 85% criterion (average=131 trials, standard deviation=51, range=60-190 trials, n=8 grackles; Table 2b, Figure 1).**\*\***

RESULTS > Model 2a > Relationship between go/no go (inhibition) and reversal learning (flexibility) > Unregistered Analyses:

“We additionally analyzed the relationship between go/no go performance and the number of trials to reverse a color preference (average=76, standard deviation=37, range=40-160, n=8 grackles) in the **\*\*first reversal\*\*** to make our results comparable across more species. This is because most studies do not conduct serial reversals, but only one reversal. **\*\*The results that included Taquito (Table 2a) were the same as the results that excluded Taquito (Table 2b): there was a positive correlation between go/no go and reversal learning performance when using the 85% go/no go criterion, and no relationship when using the 100% by 150 trial criterion. In comparison with the results for the last reversal, these results are the same as those that included Taquito (positive relationship; Table 2a), and the opposite from those that excluded Taquito (negative relationship; Table 2b).\*\***”

“Reassuringly, excluding Taquito did not change the overall patterns. There was still a negative relationship between reversal learning rate and the number of go/no go trials to pass the 85% correct criterion ( $\beta\phi\beta\varphi = -0.26$ , HPDI = -0.47 to -0.01), a positive relationship between random choice rate and go/no go trials ( $\beta\lambda\beta\lambda = -0.34$ , HPDI = -0.53 to -0.06) and a positive interaction between both learning parameters ( $\beta\phi\chi\lambda\beta\varphi\chi\lambda = 0.27$ , HPDI = -0.13 - 0.53). The results for the other go/no go criterion also did not change for the data set that included Taquito.

Overall, these results indicate that those individuals that have more inhibition are also faster at changing their preferences when circumstances change. While the relationship between trials to reverse preference and trials to reach the go/no go criterion was strongly influenced by Taquito, who was very slow in both experiments, the more comprehensive model of flexibility that takes all trials into account and does not rely on an arbitrary passing criterion provided support for the relationship irrespective of whether Taquito was included or not. Still, we would need a larger sample size to determine to what degree the relationship is perturbed by individual variation.”

RESULTS > Relationship between go/no go (inhibition) and multi-access box (flexibility):

“The average latency to attempt a new option on both MAB experiments (plastic and log) negatively correlated with go/no go performance when using the 85% go/no go criterion **\*\* (plastic sample: average=136, standard deviation=54, range=60-190, n=7 grackles, does not include Taquito; log sample: average=146, standard deviation=76, range=60-290, n=8 grackles, includes Taquito). There was no correlation when using the 150 trial threshold (average=176, standard deviation=14, range=160-201, n=7 grackles; Table 3a, Figure 3). Results from the log MAB that exclude Taquito show no relationship between the average latency to attempt a new option (average=572, standard deviation=559, range=77-1482, n=7 grackles) and go/no go performance using the 85% criterion (average=125, standard deviation=53, range=60-190, n=7 grackles).\*\*** On the plastic multi-access box, the average of the average latency per bird to attempt a new solution was 167 seconds (standard deviation=188, range=25-502, n=7 grackles). On the log multi-access box, the average of the average latency per bird to attempt a new solution was 513 seconds (standard deviation=544, range=77-1482, n=8 grackles).”

ABSTRACT: “Individuals who were faster to update their behavior in the reversal experiment **took more time to attempt a new option in the multi-access box experiments, and they were either faster or slower to reach criterion in the go/no go task depending on whether the one bird, Taquito, who was accidentally tested beyond the 200 trial cap was included in the GLM analysis. While the relationship between the number of trials to reverse a preference and the number of trials to reach the go/no go criterion was**

strongly influenced by Taquito, who was very slow in both experiments, the more comprehensive Bayesian model of flexibility that takes all trials into account and does not rely on an arbitrary passing criterion provided support for the positive relationship irrespective of whether Taquito was included.”

DISCUSSION: “We found mixed support for the hypothesis that inhibition and flexibility are associated with each other. Inhibition measured using the go/no go task was associated with flexibility (reversal task and multi-access box tasks), but inhibition measured using the detour task was not associated with either flexibility measure. While the relationship between the number of trials to reverse a preference and the number of trials to reach go/no go criterion depended on the inclusion or exclusion of one individual, flexibility measured through our more mechanistic computational model showed a consistent association with go/no go performance, such that the more flexible learners were also better at inhibition. This shows the need to move beyond rather arbitrary thresholds towards more theoretically grounded measures of cognitive traits, based on, for example, cognitive modeling of behavior. Regardless, the change of direction of the relationship given the addition or removal of one individual from the data set indicates that individuals should be tested beyond an arbitrary threshold in the go/no go test to better understand individual variation at the high end of the spectrum. The negative correlation between performance on go/no go and the multi-access boxes could indicate that solution switching on the multi-access box is hindered by self control. Performance on the multi-access box improves when one explores the other options faster. Perhaps inhibition hinders such exploration, resulting in slower switching times.”

**\*\*COMMENT 24:\*\*** Prediction 1, go/no go and Multi-access box. Lines 224-227 – The lack of correlation between flexibility measures is a separate result and should be reported in its own section. Table 3 – what are the ‘-0.00’ coefficients?

> **\*\*RESPONSE 24:\*\*** Good idea. We moved the text about the lack of correlation between flexibility measures to the second paragraph under RESULTS. The -0.00 coefficients in Table 3 are very slightly negative numbers, which were obscured due to rounding. We now note this in the table legend.

RESULTS: “There was no correlation between the two flexibility experiments: the number of trials to reverse a preference in the last reversal and the average number of seconds (latency) to attempt a new option on the multi-access box after a different locus has become non-functional because they passed criterion on it (Pearson's  $r=0.52$ , 95% confidence interval= $-0.12-0.85$ ,  $t=1.83$ ,  $df=9$ ,  $p=0.10$ ). Additionally, the average latency to attempt a new option did not correlate between the multi-access plastic and multi-access log experiments [@logan2019flexmanip]. Therefore, we conducted separate analyses for each flexibility experiment (reversal and multi-access) as well as separate analyses for the multi-access box and multi-access log.”

Table 3: “Note that an estimate of -0.00 simply means that rounding to two decimal places obscured additional digits that show this is a slightly negative number.”

**\*\*COMMENT 25:\*\*** Discussion. This section is missing a discussion of the possible drivers of negative relationship between go/no go task and multi-access box performance. Why might this be the case, especially since there is no relationship between the flexibility measures (Lines 224-225)?

> **\*\*RESPONSE 25:\*\*** Thank you for pointing this out! We added some potential explanations for the lack of a relationship between the flexibility experiments, and for the negative relationship between multi-access box and go/no go.

RESULTS: “There was no correlation between the two flexibility experiments: the number of trials to reverse a preference in the last reversal and the average number of seconds (latency) to attempt a new option on the

multi-access box after a different locus has become non-functional because they passed criterion on it (Pearson's  $r=0.52$ , 95% confidence interval= $-0.12-0.85$ ,  $t=1.83$ ,  $df=9$ ,  $p=0.10$ ). **This lack of a correlation could have arisen for a variety of reasons: 1) perhaps comparing different types of data, number of trials to pass a criterion versus the number of seconds to switch to attempting a new option, distorts this relationship. Future experiments could obtain switch latencies from reversal learning to make the measures more directly comparable. 2) Perhaps one or both flexibility measures are not repeatable within individuals, in which case, it would be unlikely that a stable correlation would be found. 3) The multi-access box experimental design allows for unknown amounts of learning within a trial, whereas the reversal learning design allows only one learning opportunity per trial. Perhaps this difference in experimental design introduces noise into the multi-access box experiment, thus making the comparison of their results ambiguous."**

DISCUSSION: "The negative correlation between performance on go/no go and the multi-access boxes could indicate that solution switching on the multi-access box is hindered by self control. Performance on the multi-access box improves when one explores the other options faster. Perhaps inhibition hinders such exploration, resulting in slower switching times."

**\*\*COMMENT 26:\*\*** Methods. Why was there no prediction about the relationship between flexibility measures?

> **\*\*RESPONSE 26:\*\***: There is a separate preregistration for the flexibility measure predictions, including how they relate to each other (Logan et al. 2019). This data has now been collected and we are currently writing it up, so hopefully it will also be out soon.

Logan CJ, Breen AJ, MacPherson M, Roney C, Bergeron L, Seitz B, Blaisdell AP, Folsom M, Johnson-Ulrich Z, Sevchik A, McCune KB. 2019. Is behavioral flexibility manipulatable and, if so, does it improve flexibility and problem solving in a new context? ([http://corinalogan.com/Preregistrations/g\\_flexmanip.html](http://corinalogan.com/Preregistrations/g_flexmanip.html)) In principle acceptance by PCI Ecology of the version on 26 Mar 2019 [https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/g\\_flexmanip.Rmd](https://github.com/corinalogan/grackles/blob/master/Files/Preregistrations/g_flexmanip.Rmd)

> **\*\*RESPONSE 27 (self added, not in response to a reviewer or to the recommender):\*\***

Since submitting the inhibition manuscript, we provided full model details and validation for the Flexibility Comprehensive model in a separate article (Blaisdell et al. 2021). Therefore, we updated this inhibition article to refer to Blaisdell et al for further details and we added a more complete summary of what the model does in this article as follows:

METHODS > Analysis Plan > Flexibility comprehensive:

"we also used a more mechanistic multilevel Bayesian reinforcement learning model that takes into account all choices in the reversal learning experiment (see Blaisdell et al. 2021 for details and model validation). From trial to trial, the model updates the latent values of different options and uses those \*attractions\* to explain observed choices. For each bird  $j$ , we estimate a set of two different parameters. The \*learning or updating rate\*  $\phi$  describes the weight of recent experience, the higher the value of  $\phi$ , the faster the bird updates their attraction. This corresponds to the first and third connotation of behavioral flexibility as defined by [Bond et al. 2007], the ability to rapidly and adaptively change behavior in light of new experiences. The \*random choice rate\*  $\lambda$  controls how sensitive choices are to differences in attraction scores. As  $\lambda$  gets larger, choices become more deterministic, as it gets smaller, choices become more exploratory (random choice if  $\lambda=0$ ). This closely corresponds to the second connotation of internally

generated behavioral variation, exploration or creativity [@bond\_serial\_2007]. To account for potential differences between experimenters, we also included experimenter ID as a random effect.”

“This analysis yields posterior distributions for phi and lambda for each individual bird. To use these estimates in a GLM that predicts their inhibition score, we propagate the full \*uncertainty\* from the reinforcement learning model by directly passing the variables to the linear model within a single large \*stan\* model. We include both parameters (phi and lambda) as predictors and estimate their respective independent effect on the number of trials to pass criterion in go/no go as well as an interaction term. To model the number of trials to pass criterion, we used a Poisson likelihood and a standard log link function as appropriate for count data with an unknown maximum.”

### ***Decision round #1***

The current manuscript is well-written and clear, with well-defined hypotheses and experimental set-ups. The results and intellectual contribution to the field are highly worthwhile and publishable. Both reviewers recommended changes mostly related to clarity, and I agree with them.

What stops me from recommending the article at this point is a valid question that both reviewers are asking: How would the results change if the outlier grackle (Taquito) were removed from the analyses? The authors note the unusual behaviour in this individual but do not assess how his inclusion or exclusion would affect overall outcomes. In agreement with the reviewers, I therefore ask that analyses be done with that individual removed, to assess the impact of this clear outlier. Likely, it would make sense to still retain the original analyses as well as the outcomes with Taquito removed.

Apart from that point, the rest of the suggested changes are mostly minor and would not necessitate another round of reviews. Thank you for your patience with the review process, which has been slower than usual due to COVID-related challenges and some technical glitches.

Preprint DOI: <https://doi.org/10.31234/osf.io/vpc39>

### ***Reviewed by Pizza Ka Yee Chow, 2021-03-12 19:44***

Thank you for inviting me to review this paper. The authors examined the relationship between behavioural flexibility (defined as the ability to inhibit a previously learned behaviour) and inhibitory control using multiple cognitive tasks established in the field. They found that the go/no-go and detour inhibition tasks showed a positive correlation with the number of trial used to reach the learning criterion in the reversal learning task and a negative correlation with the mean latency to switch to an alternative solution after the proficient solution is blocked in the multi-access box (MAB).

I fully support this investigation and think that it is much needed in the field to clarify what actually we are measuring when using different tasks. I applaud the hard work that the authors paid for the study and totally agree with them that given the mixed findings on the existing literature, we shall use several tasks to examine the relationship between flexibility and inhibitory control. It is a shame that the grackles did not fully habituate to the the delay gratification apparatus, and I hope there would opportunity to conduct this task later.

I am mostly happy with the study, and i only have a couple of minor comments, which I hope they would be useful for the revision:

#### **Introduction**

I like the introduction very much - very concise and straight to the point, with a clear study goal that has proper rationale behind.

Line 73-87 When describing each task, it may be more straightforward to relate how each task measure inhibitory control/ flexibility consistently. For example, whether successfully not pecking the no-go stimulus indicates higher inhibitory control - it is a matter of phrasing...

Line 88 The introduction so far write it in a way that it is not certain about the relationship between flexibility and inhibitory control, so I would keep it more conservative for the hypothesis as in 'Employing several experimental assays to measure flexibility and inhibition supports a rigorous approach to testing whether the two trait are linked.' Unless you are specifically testing the alternative (hypothesis 1)?

Line 105. Well, it is good that the authors set a very stringent criterion for the grackles. But we also know that grackles, like most if not all other animals rarely (consistently) do a perfect score....and the authors also noted in line 196-197 that such perfect criterion is not entirely ecological relevant.

#### Results

Line 149. This sub-heading only partial cover what this section is actually about. May be something like 'relationship between reversal learning and go/no-go task'

Line 151. Hang on, so the analyses of the reversal learning had only included those only had serial reversal learning, right? please clarify. Also some descriptive results about the number of trials that grackles took to reach the criterion in the last reversal phase would be useful.

Line 197-200 I agree with the authors that the reversal learning task and the go/no-go task share a lot of similarities. This may explain why the same type of inhibitory control is correlated between the two tasks. But this bit of information is better to be placed in the discussion.

Line 201-204. These are descriptive results of the go/no-go task, should they presented earlier in the section before, say in line 151?

Line 206-207. So what if the analysis exclude Taquito in both tasks, does the trend still hold?

Line 260 Some descriptive results about the detour task?

Line 223&261 again, the subheading only partially reflect what the section is about

Line 306 some infomration about the MAB performance?

#### Discussion

Line 349-350. The findings here support the go/no-go task and the detour task are not measuring the same inhibitory control. Would there be also other factors at play here? For example, the go/no-go requires entails much more costs than the detour task – a peck on the no-go resulted in a longer inter-trial interval and no reward whereas the detour will end up having a reward anyway...

Line 373-374 well, as the authors have mentioned previously about the latency to peck the screen data....it is a shame that it could not be analysed...would the authors have suggestions regarding how to modify the paradigm so that future studies that use touch screen for this task could also record the latency on no-go stimulus?

For all figures - Indicate sample size

Table 2, 3. Why used GLM if the authors would like to show individual variation in the two trait? isn't using GLMM and setting individual identity as random variable more appropriate for the analyses?

***Reviewed by Alex DeCasian, 2021-03-11 15:17***

[Download the review \(PDF file\)](#)