



Peer Community In Ecology

From cognition to range dynamics – and from preregistration to peer-reviewed preprint

Emanuel A. Fronhofer based on peer reviews by **Laure Cauchard** and 1
anonymous reviewer

Blaisdell A, Seitz B, Rowney C, Folsom M, MacPherson M, Deffner D, Logan CJ (2021) Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles. Missing preprint_server, ver. Missing article_version, peer-reviewed and recommended by Peer Community in Ecology.

[10.31234/osf.io/z4p6s](https://doi.org/10.31234/osf.io/z4p6s)

Submitted: 27 November 2020, Recommended: 30 March 2021

Cite this recommendation as:

Fronhofer, E. (2021) From cognition to range dynamics – and from preregistration to peer-reviewed preprint. *Peer Community in Ecology*, 100076. [10.24072/pci.ecology.100076](https://doi.org/10.24072/pci.ecology.100076)

Published: 30 March 2021

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

In 2018 Blaisdell and colleagues set out to study how causal cognition may impact large scale macroecological patterns, more specifically range dynamics, in the great-tailed grackle (Fronhofer 2019). This line of research is at the forefront of current thought in macroecology, a field that has started to recognize the importance of animal behaviour more generally (see e.g. Keith and Bull (2017)). Importantly, the authors were pioneering the use of preregistrations in ecology and evolution with the aim of improving the quality of academic research.

Now, nearly 3 years later, it is thanks to their endeavour of making research better that we learn that the authors are “[...] unable to speculate about the potential role of causal cognition in a species that is rapidly expanding its geographic range.” (Blaisdell et al. 2021; page 2). Is this a success or a failure? Every reader will have to find an answer to this question individually and there will certainly be variation in these answers as becomes clear from the referees’ comments. In my opinion, this is a success story of a more stringent and transparent approach to doing research which will help us move forward, both methodologically and conceptually.

References:

Fronhofer (2019) From cognition to range dynamics: advancing our understanding of macroecological patterns. *Peer Community in Ecology*, 100014. doi: <https://doi.org/10.24072/pci.ecology.100014>

Keith, S. A. and Bull, J. W. (2017) Animal culture impacts species' capacity to realise climate-driven range shifts. *Ecography*, 40: 296-304. doi: <https://doi.org/10.1111/ecog.02481>

Blaisdell, A., Seitz, B., Rowney, C., Folsom, M., MacPherson, M., Deffner, D., and Logan, C. J. (2021) Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles. *PsyArXiv*, ver. 5 peer-reviewed and recommended by Peer community in Ecology. doi: <https://doi.org/10.31234/osf.io/z4p6s>

Reviews

Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.31234/osf.io/z4p6s>

Version of the preprint: 3

Authors' reply, 23 March 2021

Dear Dr. Fronhofer,

Thank you very much for evaluating our revised manuscript and for your further comments. No need to apologize for a delay! We were also very busy during this time and totally understand. We revised the manuscript per your comments - please see our responses below and the revised manuscript at <https://psyarxiv.com/z4p6s> (version 4).

Additionally, we realized that we hadn't added all of the background for the validation of the Bayesian model based on the Santa Barbara grackle data, so we did this and also fleshed out the description for the Bayesian model that included the causal data (see Response 11 for details).

Many thanks again for your feedback and for making our research better!

All our best,

Corina and Aaron (on behalf of all co-authors)

Dear Dr. Blaisdell,

thank you for your revisions. I'd like to start by apologizing for the delay in handling your manuscript. Before I proceed to the recommendation of your manuscript, I would like you to go through one last round of revisions to address some remaining minor points listed below. I am looking forward to receiving a revised version of your preprint.

Sincerely yours,

Emanuel A. Fronhofer

****COMMENT 1:**** General points: Please check the formatting of the manuscript as some lines go over the margins, for instance.

> ****RESPONSE 1:**** Thank you for your thoroughness! Exporting to PDF from rmd gets pretty tricky and we endeavor (below) to clean it up.

****COMMENT 2:**** Links: Please use DOIs as much as possible. Links to github or other websites are not a priori stable.

> ****RESPONSE 2:**** We cited references with DOIs as much as we could, but ended up needing to cite two preregistrations at GitHub that are not yet post-study submitted (see citations below). We are in the process of analyzing these results, and we hope to have the post-study versions out soon. Thomas Guillemaud and Denis Bourguet at PCI suggested citing the pre-study peer reviewed preregistrations as we have, even though they do not yet have DOIs associated with them, because there are unique identifiers at GitHub that can point to the exact version that was approved. Regardless, we think that citing these preregistrations is a better solution than just saying “in prep.” because readers will know where to go to find the results (when they are posted).

Logan et al. 2019 Is behavioral flexibility manipulatable and, if so, does it improve flexibility and problem solving in a new context?

Logan et al. 2019 Are the more flexible individuals also better at inhibition?

****COMMENT 3:**** Specific points: Referencing the pre-registration: I don't see that my previous comment has been addressed. On page 1, lines 13-17 you are referencing the preregistration, as far as I understand. If this is the case, the authors that have been added on the manuscript should not be in the authors.

> ****RESPONSE 3:**** Sorry about this oversight! It turns out we didn't exactly understand what you meant before. Your point makes sense now and we changed the citation of the preregistration to omit the authors who were not already on the preregistration at the time of in principle acceptance (we removed Folsom and MacPherson).

****COMMENT 4:**** Line 54: sentence seem incomplete

> ****RESPONSE 4:**** Thank you for catching this! We meant to finish the sentence with “as well as by exerting more control over events [@blaisdell2006causal; @leising2008special; @blaisdell2012rational]”. We now corrected the error.

****COMMENT 5:**** Line 69: delete “the bird”

> ****RESPONSE 5:**** Great edit, we made the change.

****COMMENT 6:**** Line 164: Results here and this entire paragraph were accompanied by a statistical results table (previously Table 1) which has been removed since the last round of revisions. While you can of course remove the table, you may want to provide the reader with some more results regarding the statistics.

> ****RESPONSE 6:**** This is a good suggestion. We have now reported the main effects of audio cue and visual cue type as well as the interaction.

“Evidence of causal cognition in grackles would be apparent by an interaction in responding to trial type (intervene or observe) and the associated audio cue (tone or noise). Specifically, if grackles learned the common cause structure, they should respond less to the screen when they intervene to cause the tone than when they merely observe the tone. However, there should be no difference in responses to the screen whether the grackles intervene to cause the noise or simply observe the noise; thus resulting in an interaction. A 2 (trial type: intervene vs observe) x 2 (audio cue type: tone vs noise) repeated measures ANOVA revealed no significant main effect of trial type, $F(1,7) = 3.698$, $p = 0.096$, no significant main effect of audio cue, $F(1,7) < 1.0$, and critically, no significant interaction between trial type and audio cue, $F(1,7) < 1.0$. The lack of interaction suggests that there is no evidence of causal reasoning in grackles. That said, there was a low response rate in the Observe condition (Figure 3) and we only have the power to detect very large effects, which makes it difficult to rely on this conclusion.”

****COMMENT 7:**** Line 184: including the code to generate the tables was probably unintentional. Please remove here and throughout the manuscript (for all tables) to increase readability.

> ****RESPONSE 7:**** We actually did intend to include the table code so that people can fully replicate our manuscript. However, we can see how this is distracting in the PDF. Therefore, we set the table code in the rmd file to hide when exported to PDF.

****COMMENT 8:**** Line 517, 522: text goes beyond page

> ****RESPONSE 8:**** Thank you! We figured out how to fix this.

****COMMENT 9:**** Page 26, 28, 29, 31: paths go beyond page

> ****RESPONSE 9:**** These are all file paths, which are treated differently than regular text because they are one long string. We have been unsuccessfully trying for months to find a way to get this text to wrap in PDFs. Therefore, we implemented a work around by inserting text just below the file path that says: “#PDF readers: for the full file path, please see the rmd file”. The html and rmd versions of the manuscript are listed on page 1 of the pdf.

****COMMENT 10:**** Page 35: This is a local path, could you provide as for the other parts of the code a path that is accessible to all readers?

> ****RESPONSE 10:**** Good catch! This code ended up not being used because it was for experiment 2, which was not conducted because the grackles did not pass experiment 1. So we deleted the file path. However, we found another local file path in the interobserver reliability code and replaced it with the path to the data sheet on GitHub so anyone can run this code.

> ****RESPONSE 11 (self added, not in response to reviewer or the recommender):**** We realized that we hadn't previously added all of the background for the validation of the Bayesian model based on the Santa Barbara grackle data (figure 4). Therefore, we added the data sheet from the Santa Barbara grackles to the GitHub repository so that the analysis in the rmd file can be run from any computer (this data set was already published at KNB in 2016), added text to that provides background for how phi and lambda were estimated, and added the code to generate the Santa Barbara grackle figure (fig 4), and added references. We also thought of a couple of ways to more clearly explain the Bayesian model to people who are not statisticians, so we added examples for what phi and lambda mean, and added a description of the model regarding how it works when adding the causal score to it.

METHODS > ANALYSIS PLAN > Flexibility comprehensive:

Regarding the model development using Santa Barbara data: “We adapted the specific implementation of a social learning reinforcement model developed for human laboratory experiments [deffner2020dynamic].”

Regarding phi: “A value of $\phi=0.04$, for example, means that receiving a single reward for one of the two options will shift preferences by 0.02 from initial 0.5-0.5 attractions, a value of $\phi=0.06$ will shift preferences by 0.03 and so on.”

Regarding lambda: “For instance, if an individual has a 0.6-0.4 preference for option A, a value of $\lambda = 3$ means they choose A 65% of the time, a value of $\lambda = 10$ means they choose A 88% of the time and a value of $\lambda = 0.5$ means they choose A only 53% of the time.”

“We validated this computational model by analyzing data previously collected from great-tailed grackles in Santa Barbara, California [logan2016flexibilityproblem]. The following code first prepares the Santa Barbara

data for the reinforcement learning model, runs the model and extracts samples from the posterior distribution. We then use those population estimates of both learning parameters, ϕ_j and λ_j , to simulate experimental data for 8 new birds. Finally, this code plots the empirical and simulated learning curves (see figure 4)."

"This computational analysis yields posterior distributions for ϕ and λ for each individual bird. To use these estimates in a linear model that predicts Grackles' causal score, we need to propagate the full *uncertainty* from the reinforcement learning model, which is achieved by directly passing the variables to the linear model within a single large *stan* model. We include both parameters (ϕ and λ) as predictors and estimate their respective independent effect on causal score as well as an interaction term. To account for potential differences between experimenters, we also included experimenter ID as a random effect (omitted from previous equations to enhance readability, but available in the code below)."

Decision by **Emanuel A. Fronhofer**, posted 10 March 2021

Dear Dr. Blaisdell,

thank you for your revisions. I'd like to start by apologizing for the delay in handling your manuscript. Before I proceed to the recommendation of your manuscript, I would like you to go through one last round of revisions to address some remaining minor points listed below. I am looking forward to receiving a revised version of your preprint.

Sincerely yours,
Emanuel A. Fronhofer

General points:

Please check the formatting of the manuscript as some lines go over the margins, for instance.

Links: Please use DOIs as much as possible. Links to github or other websites are not a priori stable.

Specific points:

Referencing the pre-registration: I don't see that my previous comment has been addressed. On page 1, lines 13-17 you are referencing the preregistration, as far as I understand. If this is the case, the authors that have been added on the manuscript should not be in the authors.

Line 54: sentence seem incomplete

Line 69: delete "the bird"

Line 164: Results here and this entire paragraph were accompanied by a statistical results table (previously Table 1) which has been removed since the last round of revisions. While you can of course remove the table, you may want to provide the reader with some more results regarding the statistics.

Line 184: including the code to generate the tables was probably unintentional. Please remove here and throughout the manuscript (for all tables) to increase readability.

Line 517, 522: text goes beyond page

Page 26, 28, 29, 31: paths go beyond page

Page 35: This is a local path, could you provide as for the other parts of the code a path that is accessible to all readers?

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.31234/osf.io/z4p6s>

Authors' reply, 12 February 2021

Round #1

by Emanuel A. Fronhofer, 2021-01-22 16:21

Manuscript: <https://doi.org/10.31234/osf.io/z4p6s>

Dear Dr. Blaisdell,

thank you for submitting your preprint “Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles” to PCI Ecology, it is great to see a preregistration developing into a preprint.

COMMENT 1: Your preprint has been reviewed by two referees and you will see that both have a number of points that need to be addressed. I see myself in agreement with referee 1, in the sense that we appreciate the processes this work has gone through and the honesty of the preprint. Nevertheless, the results, especially given the very small sample size, are not conclusive. Referee 2 notes that even negative results merit to be published. I agree, but I ask myself, what should the reader take home from an inconclusive experiment with a very small sample size? Whatever your concrete answer to this question will be, this must be the strength of the preprint. In this context, I would like to point to referee 1’s point related to lines 222-223, which is very relevant. Beyond this, both referees point to some improvement possibilities that I would like to encourage you to follow.

RESPONSE 1: Thank you very much for your assessment! We address these points in our responses to the specific comments below.

COMMENT 2: Besides these comments, I have some additional questions: On page 1 you reference the preregistration. If I am not mistaken, Folsom and MacPherson were not on the recommended preregistration (see: https://raw.githubusercontent.com/corinalogan/grackles/master/Files/Preregistrations/g_causalPassedPreStudyPeerReview31Jan2019.pdf). Please correct the reference. Along these lines, Deffner has been added as an author and Johnson-Ulrich has been left out in comparison to the preregistration. I am not aware of PCI having rules regarding authorship on preregistration vs. preprint. Nevertheless, I would like you to make sure that all authors who merit authorship have been included.

RESPONSE 2: There are differences between the authors on the preregistration and those on the post-study manuscript, which is quite common with registered reports. This is because the authors on the preregistration were planning on collecting/analyzing data and contributing to the writing and editing process (both of which are a requirement on the grackle project and are discussed in advance of any offers of co-authorship. We follow the ICMJE guidelines and this lab policy is listed at our website: <http://corinalogan.com/ethics.html>). The authors who were on the preregistration, but not on the post-study manuscript (Johnson-Ulrich, Bergeron, and McCune) were removed because they either left the project before data collection on this experiment started and/or they chose not to contribute to the writing/editing process. The authors who were not on the preregistration, but are on the post-study manuscript (Folsom, MacPherson, and Deffner) were added because they joined the project after the preregistration passed pre-study peer review and they contributed to data collection or analysis and writing/editing. As such, the preregistration citation is different from the post-study manuscript citation.

COMMENT 3: Lines 38-43, conclusions in the abstract. Similar to lines 222-223, this should be reformulated to be more cautious or changed.

RESPONSE 3: We revised the text in the abstract to acknowledge the limitation in the claims from our study (see Response 7 for details).

COMMENT 4: Page 3: The first figure you reference is Fig. 4. Please adapt the figure numbering to start with 1.

RESPONSE 4: Good catch! We fixed it.

COMMENT 5: Table 1: correct formatting of “ η^2 ”.

RESPONSE 5: Thanks for catching this - we fixed it.

COMMENT 6: Page 10, 1): If I am not mistaken, the minimum sample size aimed for in the preregistration was $2 \times 8 = 16$ samples. Please adjust the text accordingly. In addition, referee 1 notes that the power was already not very high with $N=16$. This reduction in power should be discussed. I suggest revising your preprint

in light of the referees' comments, accompanied by a detailed response to their criticism. I am looking forward to receiving a revised version of your preprint. Sincerely yours, Emanuel A. Fronhofer

RESPONSE 6: In the Methods > Sample Size Rationale from the preregistration (which is the Methods section of the current manuscript) we had set our minimum sample size at "The minimum sample size will be 8 birds per experiment (n=16 total)". We only conducted experiment 1, which had a minimum sample size of 8 birds, which we met. We did not conduct experiment 2 (which was planned to have a sample size of 8) because the plan was only to conduct experiment 2 if the grackles showed evidence of causal cognition in experiment 1 (noted in Methods > Experiment 1). Thanks to your comment, we now realize that we used the wrong sample size in our power analysis (n=16 instead of n=8). We ran a new power analysis with the planned sample size of 8 per experiment and we now include this in the Methods > Analysis plan. The chance of us being able to detect an effect increased from 0.77 to 1.11 and we added a sentence about this in the results section.

RESULTS: "However, there was a low response rate in the Observe condition (Figure 2) and we only have the power to detect very large effects, which makes it difficult to rely on this conclusion"

Reviewed by anonymous reviewer, 2020-12-23 13:25

COMMENT 7: Authors studied the causal cognition in the great-tailed grackles and its link to behavioral flexibility with indirect implications for range-expanding species. The question is interesting. Authors failed to find evidence for causal evidence and even more so a link with flexibility. While the absence of causal cognition is worth publishing, there are several issues that prevent a firm conclusion on the biological underpinnings of a lack of causal cognition. The issues are: 1) not sure birds were "very attentive to visual events presented on the touchscreen" (lines 187-194), 2) "the touchscreen might be inappropriate for testing causal processes associated with obtaining food" (lines 195-202), 3) the very low sample size (i.e. 8 birds) and the limited initial power analysis of 71% chance of detecting a large effect, 4) "the changes in protocol over the course of the experiment" (line 258), 5) the "However, note that the very low response rate in the Observe condition (Figure 2) makes it difficult to rely on this conclusion." (line 124-125). While I appreciate the great honesty all along the manuscript, I regret to say that all these points make the current results hardly conclusive. And frankly speaking, while again I do value non-significant (wrongly called negative) results, publishing a study one cannot rely on the results because of multiple methodological issues is debatable. I respectfully disagree with (lines 222-223) "This failure to find evidence of causal cognition in the grackle provides a cautionary tale for comparative psychologists interested in testing wild-caught animals using traditional laboratory apparatuses and techniques". This sentence mixed up two ideas. Scientists should indeed make sure they use appropriate methods to test their questions on the focal species. I fully agree. However, I find misleading to link it with "the failure to find evidence of causal cognition" which implies causal cognition was properly tested. Authors may have failed to test for causal cognition and not failed to find evidence for it, or at least it is currently difficult to tease apart these two. These two statements (i.e. failing to test for and failing to find evidence) are clearly different.

RESPONSE 7: Because this article passed pre-study peer review at PCI Ecology based on our methods and planned sample size, which we met, the manuscript cannot be rejected for methodological or sample size reasons. Nevertheless, we acknowledge that lines 222-223 could be misleading. We added text:

ABSTRACT: "This could indicate that our test was inadequate to assess causal cognition. Because of this, we are unable to speculate about the potential role of causal cognition in a species that is rapidly expanding its geographic range. We suggest further exploration of this hypothesis using larger sample sizes and multiple test paradigms"

And the following text has been placed in the...

DISCUSSION after lines 222-223: "Such failures can occur not due to the true lack of the behavioral process in test subjects, but rather, to shortcomings in the approach itself. The caveats raised above preclude us from determining the actual absence of causal cognition in the great-tailed grackles we tested."

Although there were changes to the protocol as we conducted the tests, these were purely related to increasing the motivation of the grackles to participate. The changes did not affect the planned experimental

design of the tests.

We have now quantified each bird's attentiveness during the tests using the number of sessions it took to complete their training sessions and their Observe and Intervene tests (data from the original data sheet, summarized in the Discussion in the new Table 5). Ideally, each test is conducted in one session, however due to the lack of motivation of many of the grackles to participate, multiple sessions were conducted to complete all trials in each training or test program. We added this to the...

DISCUSSION: "Half of the grackles completed the intervene test in one session, indicating that they were attentive to the screen, while the other half needed two to fifteen sessions, indicating they were less interested in attending to the screen. Whereas, six grackles completed the observe test in one session, while the other three required two to three sessions (Table 5). A similar level of attentiveness occurred in the training sessions: about half of the training programs were completed in one session, while the other half required two to eight sessions (Table 5)"

COMMENT 8: I also find hazardous that authors discuss and even mention range expansion ("could indicate that it (causal cognition) is not implicated as a key factor involved in a rapid geographic range expansion", lines 231-232). None of the analysis relates to range expanding processes (e.g. birds sampled in populations of varying ages) and authors are not sure they properly measured causal cognition and can rely on their conclusions. I am sorry my comments may sound harsh. I advise to ensure the protocol is a valid and functional protocol for the studied species and increase the sample size. Other comments:

RESPONSE 8: Good point that we need to tone down our speculation about the potential role of causal cognition in a range expansion. We changed the text as follows:

DISCUSSION: Given that there is ambiguity around whether the grackles do not use causal cognition or whether the test did not work, we will refrain from speculating about whether it is involved in a rapid geographic range expansion with regards to the lack of a correlation with behavioral flexibility.

COMMENT 9: 1) Methods are sometimes mixed with the introduction and the results (e.g. lines 57-59, 134-145)

RESPONSE 9: We removed the methods sentences from the introduction and results sections as suggested.

COMMENT 10: 2) In a few places (e.g. lines 61-63 and 67-80), the introduction is hardly understandable for non-specialists.

RESPONSE 10: We added an example to make causal maps easier to understand, and we added a new figure to illustrate the schematic of the Blaisdell et al (2006) study described (see Response 23 for the new text from the manuscript).

COMMENT 11: 3) Cannot find anywhere the supplementary information for the statistical analysis for the repeated ANOVA with JASP (not clear why R is not used for this analysis). The exact model used is not clear, how it controls for repeated measures, where do all the residuals come from. An experimenter random intercept is added in some analyses. How many experimenters are running the experiment? Is an experimenter effect well-distributed among or nested within a bird ID or a condition? With 8 individuals, this point is very important.

RESPONSE 11: Thank you! We apologize for the confusion. Three different people conducted analyses on this article and two use R and one does not use R. This is the reason for the ANOVA being conducted with a different program. The ANOVA analysis was performed in JASP, free and open source software that acts as GUI for R. We have now included the .jasp file for this analysis in the data package at KNB (*Gracklecausal*JASP.jasp), which can be downloaded and opened by anyone who has also downloaded JASP. This also allows one to review and recreate our analysis. We also apologize for including Tab 1, which may have caused some confusion as "residuals" are not typically reported for a repeated measures ANOVA. We have revised how those data are reported. With that said, the model remains the same, a 2 (Audio Cue: Tone vs Noise) x 2 (Cue Type: Observe vs Intervene) repeated measures ANOVA. There is no need to correct for sphericity, because with only 2 levels of repeated measures the assumption of sphericity cannot be violated (Hinton et al. 2004).

While we expect no significant experimenter effects (we have detailed protocols and all experimenters undergo extensive training), we wanted to make sure that any differences that might have accidentally occurred

could be accounted for by the model such that we could see the actual performance of the birds more clearly. We are only able to include experimenter as a random effect if the response variable has more than one row per bird, as it did with this model. There were 4 different experimenters for the reversal learning task. While trials were not evenly distributed among experimenters for each bird, 6 birds were tested by at least 2 experimenters and 4 birds by 3 different experimenters, such that the model could clearly differentiate bird ID effects from experimenter ID effects, which were relatively minor.

References:

Hinton, P. R., Brownlow, C., & McMurray, I. (2004). *SPSS Explained*. Routledge.

COMMENT 12: 4) Lines 680-683: "We have chosen to keep the models as simple as possible because the sample sizes for each experiment are small. These experiments were designed to determine whether grackles attend to causal cues or not. If results show that they do, then we will conduct further tests to investigate the extent of these abilities." This is circular thinking. The sample size may prevent authors from finding a causal cognition and so the decision to extend this experiment to other places and to a larger sample size is created by this small sample size.

RESPONSE 12: Because we used a completely within-subject design, unlike the study by Blaisdell et al, which used a mixed design, we matched their study in terms of group size for each predicted effect, which allowed us to use fewer subjects in total than in the Blaisdell et al study. The n was, therefore, powered enough to enable replication of the Blaisdell results.

COMMENT 13: 5) Figure 1 legend: circle should be changed for stars.

RESPONSE 13: Thank you for catching this! We made the change.

COMMENT 14: 6) Figure 3 legend: a legend for grey bars (confidence intervals?) and the black curve (means) should be added.

RESPONSE 14: Thanks! We changed the caption to clarify.

Reviewed by Laure Cauchard, 2021-01-14 16:10

COMMENT 15: This study explores the role of causal cognition in behavioural flexibility in 8 wild great-tailed grackles temporarily kept in captivity (up to 6 months). The study has 2 goals: 1) are the grackles displaying causal inference using a touch screen task, and 2) is this ability correlated with behavioural flexibility, measured in another study on the same individuals. Unfortunately, results show no evidence of causal inference, and the performance of the touch screen task does not correlate with behavioural flexibility. Negative results are as informative as positive results and deserve to be published, so that future studies can rely on it to go further. The conclusion of the study is well adapted to the results and the authors acknowledge their limits. My main comments would be: - The introduction should be reorganized so that the goals of the study arrive at the end of the introduction and the arguments supporting the study should be developed. Ideas maybe: what are the processes underlying beh flex, are they well known? Is causal cognition well spread in the animal kingdom, if not why? Maybe also why this species: they are expanding rapidly yes, but is this the only reason?

RESPONSE 15: In the introduction, we moved the grackle goals near the end. Great question about why this species is rapidly expanding its range - no one knows, which is why we are really interested in this question and why our larger research program is attempting to find some answers. So no empirical answers on that front yet (hopefully in the next couple of years!). The processes underlying behavioral flexibility are also unknown. Therefore, we took your advice and discussed causal cognition across the animal kingdom in a new paragraph.

INTRODUCTION: "While a smattering of studies have investigated understanding of physical causality, such as in tool construction and use, in birds and mammals [see reviews by @emery2009tool; @lambert2019birds; @volter2017causal] there are no existing studies of causal perception aside from those in rats from Blaisdell's lab (see below) and chimpanzees [@premack1994levels]."

COMMENT 16: - If causal cognition is not working, maybe the authors can find another score to relate performance to the task to another process, such as trial and error learning? Attention at least? Just to show that there is at least one simple learning process in action at some point and that the touch screen task is working in measuring something? A negative result is a result but at least we have to show that the task is

measuring something.

RESPONSE 16: We conducted three experiments using this touchscreen: the causal experiment here, a reversal learning task (Logan et al. 2019 http://corinalogan.com/Preregistrations/g_flexmanip.html, results not yet analyzed), and a go no-go inhibition task (Logan et al. 2020 http://corinalogan.com/Preregistrations/g_inhibition.html, results included at the link). We have evidence that the grackles (indeed, the individuals in the causal cognition test) were able to successfully interact with the touchscreen in the inhibition task where they took an average amount of trials to pass criterion compared with other species. We mention this in the Discussion as evidence that they are able to learn on the touchscreen. They were extremely slow at reversal learning on the touchscreen apparatus compared with physical colored tubes so it appears that shape or color discrimination works a bit differently for them depending on the testing apparatus. Your comment about attention spurred us to summarize how many sessions it took them to complete the training, intervene, and observe tests (see Response 7 for the full description and revision that was made). Trials were manually initiated by the experimenter only when the grackle was attending to the screen, therefore session number is a useful proxy for attention.

COMMENT 17: ABSTRACT: L31-34: "by allowing... by making... by exerting..." a little bit hard to follow, rephrase? I would add the sample size in the abstract because it is important to know it to understand the conclusion of the study. And I think references can be removed from the abstract, it will make space for a sentence or two about the touchscreen task maybe?

RESPONSE 17: We removed the references from the abstract, added the sample size, and modified the sentence as suggested, and we added a sentence about the touchscreen task:

ABSTRACT: "could play a significant role in rapid range expansions via the ability to learn faster: causal cognition could lead to making better predictions about outcomes through exerting more control over events" and "Causal cognition was measured using a touchscreen where individuals learned about the relationships between a star, a tone, a clicking noise, and food. They were then tested on their expectations about which of these causes the food to become available."

COMMENT 18: INTRODUCTION L46-48: I think this sentence can be explained with an example or two: how being able to change behaviour would help in a new environment?

RESPONSE 18: Good point. We added the following:

INTRODUCTION: For example, flexibility would be useful for changing food preferences in accordance with locally available resources that potentially fluctuate over time.

COMMENT 19: L49: why "however"?

RESPONSE 19: "However" was intended to note that causal cognition is a different trait from flexibility, but we can see how this wasn't clear. We revised the sentence to say:

INTRODUCTION: "It is alternatively or additionally possible that causal cognition, the ability to understand the causality in relationships between events beyond their statistical covariations"

COMMENT 20: L51: add a reference at the end of the definition.

RESPONSE 20: We added an example of what causal cognition is after its first mention where we reference Blaisdell et al. 2006, Leising et al. 2008, and Blaisdell and Waldmann 2012. We also added an additional example and listed the associated citation.

INTRODUCTION: "For example, if a monkey observes an association between a tree branch shaking and a piece of fruit dislodging and falling to the ground where it can be consumed, a causal understanding of this association, that is that shaking the branch caused the fruit to fall, would provide the monkey the opportunity to itself intervene and shake the branch so as to procure the fruit."

INTRODUCTION: "For example, a bird that observes wind moving a branch, and then sees a fruit attached to the branch fall to the ground, the bird might interpret these observations as a causal chain in which wind *causes* the branch to shake, which in turn *causes* the fruit to shake loose and fall to the ground [@tomasello1997primate]"

COMMENT 21: L52: "by exerting more control over event" not sure about what that means?

RESPONSE 21:We added an example to clarify how having causal understanding would enable control over a situation for an individual's benefit (see Response 20 for details).

COMMENT 22: L55-61: the goals of the study are arriving too fast in the introduction, only 10 lines in the introduction and the authors already present the study.

RESPONSE 22: Please see our Response 15 where we addressed this point.

COMMENT 23: L61-66: this is definition of causal inference, this should be placed when the authors first speak about causal cognition.

RESPONSE 23: Causal models are only one aspect of causal cognition, and the one we are specifically investigating in this study. We now clarify this in the text.

INTRODUCTION: "For example, a bird that observes wind moving a branch, and then sees a fruit attached to the branch fall to the ground, the bird might interpret these observations as a causal chain in which wind *causes* the branch to shake, which in turn *causes* the fruit to shake loose and fall to the ground [(@tomasello1997primateTomasello & Call, 1997)]. Causal maps, such as the wind->shake branch->fruit falling map just described, thus provide the causal structure in relationships that go beyond merely observing statistical covariation between events, and allow causal inferences to be derived, such as through diagnostic reasoning and reasoning about one's own interventions on events within the causal model (Blaisdell & Waldmann, 2012) (Waldmann, 1996). Returning to the causal chain example, a bird with such causal knowledge could intervene to shake the branch itself (a causal intervention) with the expectation that they could themselves make the fruit fall to the ground where they would be able to retrieve and eat it. Without such causal knowledge, the bird would not try to shake the fruit loose."

COMMENT 24: L84: Figure 1. L83-89: this is the same results as the paragraph before, explained differently. The authors could instead explain here how they will proceed with their grackles (1 sentence or 2, this is introduction, not methods). And explain the Figure 1 with their own hypotheses and expected results?

RESPONSE 24: Great idea! We removed the text about the rat experiment and we added text about the experiment in the context of the grackles. Here is our addition...

INTRODUCTION: "To do so, we implemented a conceptually similar design, but adapted to the touchscreen. Grackles would first be trained to peck a small white square (the response key) presented on the lower part of the screen just above the food hopper. Pecks to this response key resulted in delivery of food from the hopper. Next, grackles would receive three types of trials intermixed within each training session. One type of trial consisted of the presentation of a white star in the center of the screen followed by the presentation of a tone from a speaker next to the screen. The second type of trial consisted of the presentation of the white star followed by the delivery of food in the food hopper. The third type of trial consisted of presentations of a noise from the speaker followed by delivery of food. As a result of these three types of trial, grackles should develop the causal models shown in the right panel of Figure 2. At test, each grackle will receive each of four types of test trial in separate test sessions. During the Observation test session, the grackle will receive two types of test trial interspersed within the session. One type of test trial consists of presentation of the tone by itself, while the other type of test trial consists of presentation of the noise by itself. If grackles had formed the causal models depicted in Figure 2, then upon hearing the tone the grackle should diagnostically infer that the star must have caused the tone, and because the star is also a cause of food, the grackles should expect food. Likewise, when they hear the noise at test, because the noise is a direct cause of food, they should expect food. Thus, in both cases, grackles should look for food in the hopper, or peck the food key which had been previously associated with food. During the Intervention test session, the grackle will be presented with two novel visual stimuli on the screen as shown in the center panel of Figure 2. Pecks to one of the stimuli, such as the clover, results in the presentation of the tone. Pecks to the other stimulus, such as the triangle, results in the presentation of the noise. When the noise is produced through the intervention of pecking at the triangle, grackles should expect food because noise is a cause of food. When the tone is produced through the intervention of pecking at the clover, however, the grackles should NOT expect food. This is because, if the grackle formed the common cause model of the star being a common cause of tone and food, when the

grackle intervenes to produce the tone, it should attribute the occurrence of that tone to their own causal intervention, and not to the prior cause of the star. Thus, they should not expect that the star had caused THAT tone (the one the bird caused through its intervention) and thus should not expect food either. Thus, we predict more food inspection behavior when the grackles intervenes to cause the noise than when they intervene to cause the tone. This result would replicate the finding in rats by Blaisdell et al.”

COMMENT 25: L92-94: I would say “our results will indicate whether grackles exhibit ...” or “our results indicate that grackles exhibited ...”. Like now this sentence is weird, neither exposing the results nor asking a question.

RESPONSE 25: Thanks for the catch! We fixed the grammar as you suggested.

COMMENT 26: Figure legends: the circle is still in the legend but I think it is coming from the previous design right? It should be the star instead.

RESPONSE 26: Thank you! We made the change.

COMMENT 27: RESULTS L124-125: the combined effect of a low sample size + a very low response rate..., especially with an ANOVA.

RESPONSE 27: We are not clear what the question/comment is, but hopefully we covered this concern above: we revised the way we report and discuss the results of the repeated measures ANOVA, and we note the limitations of this analysis with these constraints (see Response 11 for more details about the ANOVA, and Response 7 for more details about the sample size).

COMMENT 28: Table 1: the legend should be more developed so that we can understand the table without the main text.

RESPONSE 28: We now removed the table because it seemed to cause confusion rather than being helpful.

COMMENT 29: L134-145: relate to the first paragraph, do grackles show evidence of causal cognition.

RESPONSE 29: We moved this piece to the Methods so now this section begins with the result instead of discussing the equation (also in accordance with Response 9).

COMMENT 30: L139: what ‘abs’ means? I have to say that I am not familiar with the last part of the results, the mechanistic approach. I am not able to review this section.

RESPONSE 30: abs means the absolute value, which we now explain in the new location for the equation, which is in the Methods.

COMMENT 31: DISCUSSION L187-194: the authors must have a data representing attention that can be used to test this hypothesis.

RESPONSE 31: Please see Response 7, which discusses how we added a summary of how many sessions it took the grackles to complete the training and tests (as an indicator of attention).

COMMENT 32: L195-202: In another species of grackles, the Carib grackle, simple task requiring pecking are working, and color cue-food associative learning have been used (see the work from S. Overington and J Morand Ferron). However, a screen has not been used yet.

RESPONSE 32: Thank you for pointing us in this direction! We are aware of Morand-Ferron’s use of an operant chamber with wild great tits, but we can’t find anything like this from her or from Overington using operant chambers in Carib grackles. Did we miss this? Sorry if we misinterpreted your comment.

Decision by [Emanuel A. Fronhofer](#), posted 22 January 2021

Dear Dr. Blaisdell,

thank you for submitting your preprint “Do the more flexible individuals rely more on causal cognition? Observation versus intervention in causal inference in great-tailed grackles” to *PCI Ecology*, it is great to see a preregistration developing into a preprint.

Your preprint has been reviewed by two referees and you will see that both have a number of points that need to be addressed. I see myself in agreement with referee 1, in the sense that we appreciate the processes this work has gone through and the honesty of the preprint. Nevertheless, the results, especially given the very small sample size, are not conclusive. Referee 2 notes that even negative results merit to be published. I

agree, but I ask myself, what should the reader take home from an inconclusive experiment with a very small sample size? Whatever your concrete answer to this question will be, this must be the strength of the preprint. In this context, I would like to point to referee 1's point related to lines 222-223, which is very relevant. Beyond this, both referees point to some improvement possibilities that I would like to encourage you to follow.

Besides these comments, I have some additional questions:

On page 1 you reference the preregistration. If I am not mistaken, Folsom and MacPherson were not on the recommended preregistration (see: https://raw.githubusercontent.com/corinalogan/grackles/master/Files/Preregistrations/g_causalPassedPreStudyPeerReview31Jan2019.pdf). Please correct the reference.

Along these lines, Deffner has been added as an author and Johnson-Ulrich has been left out in comparison to the preregistration. I am not aware of PCI having rules regarding authorship on preregistration vs. preprint. Nevertheless, I would like you to make sure that all authors who merit authorship have been included.

Lines 38-43, conclusions in the abstract. Similar to lines 222-223, this should be reformulated to be more cautious or changed.

Page 3: The first figure you reference is Fig. 4. Please adapt the figure numbering to start with 1.

Table 1: correct formatting of " η^2 ".

Page 10, 1): If I am not mistaken, the minimum sample size aimed for in the preregistration was $2 \times 8 = 16$ samples. Please adjust the text accordingly. In addition, referee 1 notes that the power was already not very high with $N=16$. This reduction in power should be discussed.

I suggest revising your preprint in light of the referees' comments, accompanied by a detailed response to their criticism. I am looking forward to receiving a revised version of your preprint.

Sincerely yours, Emanuel A. Fronhofer

Reviewed by anonymous reviewer 1, 23 December 2020

Authors studied the causal cognition in the great-tailed grackles and its link to behavioral flexibility with indirect implications for range-expanding species. The question is interesting. Authors failed to find evidence for causal evidence and even more so a link with flexibility. While the absence of causal cognition is worth publishing, there are several issues that prevent a firm conclusion on the biological underpinnings of a lack of causal cognition.

The issues are: 1) not sure birds were "very attentive to visual events presented on the touchscreen" (lines 187-194), 2) "the touchscreen might be inappropriate for testing causal processes associated with obtaining food" (lines 195-202), 3) the very low sample size (i.e. 8 birds) and the limited initial power analysis of 71% chance of detecting a large effect, 4) "the changes in protocol over the course of the experiment" (line 258), 5) the "However, note that the very low response rate in the Observe condition (Figure 2) makes it difficult to rely on this conclusion." (line 124-125).

While I appreciate the great honesty all along the manuscript, I regret to say that all these points make the current results hardly conclusive. And frankly speaking, while again I do value non-significant (wrongly called negative) results, publishing a study one cannot rely on the results because of multiple methodological issues is debatable.

I respectfully disagree with (lines 222-223) "This failure to find evidence of causal cognition in the grackle provides a cautionary tale for comparative psychologists interested in testing wild-caught animals using traditional laboratory apparatuses and techniques". This sentence mixed up two ideas. Scientists should indeed make sure they use appropriate methods to test their questions on the focal species. I fully agree. However, I find misleading to link it with "the failure to find evidence of causal cognition" which implies causal cognition was properly tested. Authors may have failed to test for causal cognition and not failed to find evidence for it, or at least it is currently difficult to tease apart these two. These two statements (i.e. failing to test for and failing to find evidence) are clearly different.

I also find hazardous that authors discuss and even mention range expansion (“could indicate that it (causal cognition) is not implicated as a key factor involved in a rapid geographic range expansion”, lines 231-232). None of the analysis relates to range expanding processes (e.g. birds sampled in populations of varying ages) and authors are not sure they properly measured causal cognition and can rely on their conclusions.

I am sorry my comments may sound harsh. I advise to ensure the protocol is a valid and functional protocol for the studied species and increase the sample size.

Other comments:

- 1) Methods are sometimes mixed with the introduction and the results (e.g. lines 57-59, 134-145)
- 2) In a few places (e.g. lines 61-63 and 67-80), the introduction is hardly understandable for non-specialists.
- 3) Cannot find anywhere the supplementary information for the statistical analysis for the repeated ANOVA with JASP (not clear why R is not used for this analysis). The exact model used is not clear, how it controls for repeated measures, where do all the residuals come from. An experimenter random intercept is added in some analyses. How many experimenters are running the experiment? Is an experimenter effect well-distributed among or nested within a bird ID or a condition? With 8 individuals, this point is very important.
- 4) Lines 680-683: “We have chosen to keep the models as simple as possible because the sample sizes for each experiment are small. These experiments were designed to determine whether grackles attend to causal cues or not. If results show that they do, then we will conduct further tests to investigate the extent of these abilities.” This is circular thinking. The sample size may prevent authors from finding a causal cognition and so the decision to extend this experiment to other places and to a larger sample size is created by this small sample size.
- 5) Figure 1 legend: circle should be changed for stars.
- 6) Figure 3 legend: a legend for grey bars (confidence intervals?) and the black curve (means) should be added.

Reviewed by Laure Cauchard, 14 January 2021

This study explores the role of causal cognition in behavioural flexibility in 8 wild great-tailed grackles temporarily kept in captivity (up to 6 months). The study has 2 goals: 1) are the grackles displaying causal inference using a touch screen task, and 2) is this ability correlated with behavioural flexibility, measured in another study on the same individuals. Unfortunately, results show no evidence of causal inference, and the performance of the touch screen task does not correlate with behavioural flexibility. Negative results are as informative as positive results and deserve to be published, so that future studies can rely on it to go further. The conclusion of the study is well adapted to the results and the authors acknowledge their limits. My main comments would be: - The introduction should be reorganized so that the goals of the study arrive at the end of the introduction and the arguments supporting the study should be developed. Ideas maybe: what are the processes underlying beh flex, are they well known? Is causal cognition well spread in the animal kingdom, if not why? Maybe also why this species: they are expanding rapidly yes, but is this the only reason? - If causal cognition is not working, maybe the authors can find another score to relate performance to the task to another process, such as trial and error learning? Attention at least? Just to show that there is at least one simple learning process in action at some point and that the touch screen task is working in measuring something? A negative result is a result but at least we have to show that the task is measuring something.

ABSTRACT: L31-34: “by allowing... by making... by exerting...” a little bit hard to follow, rephrase? I would add the sample size in the abstract because it is important to know it to understand the conclusion of the study . And I think references can be removed from the abstract, it will make space for a sentence or two about the touchscreen task maybe?

INTRODUCTION L46-48: I think this sentence can be explained with an example or two: how being able to change behaviour would help in a new environment? L49: why “however”? L51: add a reference at the end of the definition. L52: “by exerting more control over event” not sure about what that means? L55-61: the goals of the study are arriving too fast in the introduction, only 10 lines in the introduction and the authors

already present the study. L61-66: this is definition of causal inference, this should be placed when the authors first speak about causal cognition. L84: Figure 1. L83-89: this is the same results as the paragraph before, explained differently. The authors could instead explain here how they will proceed with their grackles (1 sentence or 2, this is introduction, not methods). And explain the Figure 1 with their own hypotheses and expected results? L92-94: I would say "our results will indicate whether grackles exhibit" or "our results indicate that grackles exhibited ...". Like now this sentence is weird, neither exposing the results nor asking a question. Figure legends: the circle is still in the legend but I think it is coming from the previous design right? It should be the star instead.

RESULTS L124-125: the combined effect of a low sample size + a very low response rate..., especially with an ANOVA. Table 1: the legend should be more developed so that we can understand the table without the main text. L134-145: relate to the first paragraph, do grackles show evidence of causal cognition. L139: what 'abs' means? I have to say that I am no familiar with the last part of the results, the mechanistic approach. I am not able to review this section.

DISCUSSION L187-194: the authors must have a data representing attention that can be used to test this hypothesis. L195-202: In another species of grackles, the Carib grackle, simple task requiring pecking are working, and color cue-food associative learning have been used (see the work from S. Overington and J Morand Ferron). However, a screen has not been used yet.