The proposed project tests for sex differences in reversal learning in Grackles sampled from three different sites. The authors have clearly justified the project, and provided code and simulated data. I think the RR should be accepted. I just have some minor comments which are aimed at making the RR and any resulting reports easier for readers to access and understand.

**1A. The scientific validity of the research question(s).**

The research question is perfectly logical. I would like to see the authors set some of their reasoning out a little more explicitly, though.

For example, when the authors write (lines 36-37): “Given the influence of sex-biased dispersal and learning ability on range expansion, it is perhaps surprising, then, that their potential joint impact on this aspect of movement ecology remains unexamined…” What do they mean by ‘joint impact’? I’m under the impression that the hypothesis of the report is that male Grackles will have faster reversal learning because of sex-biased dispersal. I’m not sure what the implications are of this on range expansion. (I think what I’m asking is “what DAG do you draw from this sentence?”)

**1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.**

Again, I think this is perfectly valid. A couple of comments on spelling out the reasoning follow.

Lines 44-52. If I understand correctly, the argument here is that sex difference in learning speed can be estimated in a way that is ‘uncontaminated’ by these various other factors. Given that this argument is based on the absence of a correlation, can the authors comment at all on either: a) the strength of evidence for null correlations? b) whether these correlations have been estimated in the same populations that they are now studying?

Lines 81-82: is the argument here that recent selection pressures for fast learning should have been stronger for males than for females? If so, could this be stated explicitly? If not, could the argument be clarified?

**1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).**

This RR describes an analysis of data that are already being collected by a third party. I therefore have no real concerns about the methodology. I would like to see explicit reference to the primary data collection being approved by a relevant ethics board.

Power analysis based on NHST is not applicable here. I would like to see in the final report some consideration of how representative captured grackles are likely to be of grackles in general.

**1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.**

With regard to method: the authors cite a detailed description of the experimental protocols, which is also published in an open-access journal and is thorough.

With regard to the analysis: I have run the analysis code and got the same outputs as given in the RR (minus a couple of formatting issues). This is great to see, and I commend the clarity of the authors’ code and commenting!

Because I’m not great at using the rethinking package (or indeed R more generally), it took me a wee while of flailing and googling to get the code running on my machine. I’d like to suggest that the authors include links to the Rtools installation page (if indeed this was actually necessary and not just something I ended up doing because I had no clue what I was doing…) and Richard McElreath’s rethinking installations instructions page (links below) for the inexperienced rethinking user. The (clunky, inefficient) code I ended up running to get everything working is below.

**1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).**

I think so. I would like to see a little more detail about how the size of the sex difference in the simulated data was chosen (was it based on a report from another species, for example?) Also it would be useful to see a slightly clearer treatment of how non-choices are to be treated in the analysis, and mention of any plans the authors might have to investigate choice latency as a variable (given that it seems to be being recorded).

R Code:

#NB this ran on R 4.0.3 (“Bunny-Wunnies Freak Out”) and RStudio 1.3.1093.

#installed latest version of rtools from: https://cran.rstudio.com/bin/windows/Rtools/rtools40.html

#make Rtools work

write('PATH="${RTOOLS40\_HOME}\\usr\\bin;${PATH}"', file = "~/.Renviron", append = TRUE)

#verify that Rtools works. correct output is: make "C:\\rtools40\\usr\\bin\\make.exe"

Sys.which("make")

library(cowplot) #For combining plots

library(tidyverse) #For data wrangling

library(rstan) #Needed for rethinking - includes ggplot2 for graphing

library(Rcpp) #Needed for rethinking apparently

#install rethinking as per RM instructions at: #https://www.rdocumentation.org/packages/rethinking/versions/2.13

install.packages(c("coda","mvtnorm","devtools","loo","dagitty"))

devtools::install\_github("rmcelreath/rethinking")

library(rethinking) #for making the model