This preregistration describes a study aiming to measure how individual variation in behavioral flexibility relates to other behavioral traits such as exploration and boldness. It is argued that elucidating these behavioral patterns will help improve our understanding of how species' are able to adapt to new or changing environments. The authors plan to measure behavior in captive and wild great-tailed grackles which seem like a good system for examining questions relating behavior to species expansion. The overall goal is to manipulate individual flexibility by utilizing different training protocols such as a serial reversal learning task in a subset of individuals and measuring if this manipulation impacts other behavioral traits. Group differences in these traits would suggest some connection between flexibility and that trait, while a lack of group differences would suggest independence with flexibility. Even if the manipulation does not work, the authors state that they would have the ability to examine individual-level behavioral patterns between flexibility, exploration, and boldness in this system.

Overall, I find the topic to be of particular scientific merit as there is clearly growing interest in the animal cognition field for both measuring individual level variation in behavioral flexibility and tying that variation to other behavioral traits (e.g., coping styles, cognitive syndromes, etc). To then tie that variation to species expansion would be particularly exciting. While I am a little skeptical in terms of the manipulation working because it is unclear how well it will generalize across contexts, I think the approach is well thought out, and I agree with the authors that even if it doesn't work they will still have a worthwhile dataset in order to examine underlying behavioral patterns. These patterns along with measures of repeatability both involving captive and wild individuals would be a worthwhile dataset for publication. Below I have included some comments and questions.

Predictions: I understand that the section "P1-P5 alternative" is describing the alternative to the section "Predictions 1-5". I believe it is saying that even if the manipulation doesn't work you should still be able to examine underlying patterns of correlation between these traits among individuals. However, I am a little confused by sections following (e.g., P1 alternative 1). Are these predictions being made under the assumption that the manipulation failed?? I think this could be a little clearer.

P1 alternative 1: Unclear what is meant by "… could indicate another trait is present, such as boldness." Do you mean that this other trait could be explaining both individual variation in exploration and flexibility?

P1 alternative 2: Would it matter which of the dependent outcomes it was correlated with (i.e., the one that accounts for exploration in reversal learning or the one that does not)? Also, how can flexibility be described as totally independent in this case if one of the two measures of flexibility are associated with exploration?

Figure 1: Is time 1 before or after the manipulation?

H2: What if the manipulation itself manipulates these other traits independently. For example, the repeated trials of the manipulation habituate the animal to handling and other

experimental stressors and therefore results in them showing more exploratory behavior because they are no longer shutting down behaviorally from these stressors.

P6: Alternative. Repeatability and changing behavior are not mutually exclusive as it is how behavior changes relative to other individuals. All individuals can change their behavior across time and still have high repeatability (e.g., those with the highest scores at time 1 still have highest scores at time 2 even though the exact scores may differ considerably). Also, even with lower repeatability you would still say that the traits are at least partially a property of the individual.

Novel Environment: What is the rationale for having the familiar environment measure always first? Are you comparing main effects in terms of movement between familiar and novel environments, or just relative differences between individuals?

- This protocol seems different from your reference in that the bird is really examining a large object in a familiar space vs. entering a whole new space as was tested in Mettke-Hoffman et al. 2009. Seems arguable if this is novel space or novel object. This might be important in terms of interpretation of results and distinguishing between predictions 1 and 2.

Are you tracking unsuccessful wild assay attempts? I think it will be important to track overall participation in order to argue against possible critiques of self-selection bias due to personality differences.

Analysis: It is unclear why flexibility measures are being used as dependent variables when the research is being framed as examining how manipulation of flexibility may or may not affect other behavioral traits. This also seems particularly troublesome when condition is