REBUTTAL LETTER

Associate Editor Comments to Author:

The reviewers feel that the topic of your paper is interesting and that your study has the potential to make a good contribution to Ecology and Evolution. However they also feel that your paper as currently written lacks sufficient details about your methodology and analysis to enable your study to be fully assessed. Hopefully you can add these details to a revised version of your paper.

RESPONSE: Thanks for this opportunity, we have addressed the reviewers’ concerns below.

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author

In this manuscript, the authors examined the influence of corticosterone and incubation temperature on foraging decisions in the common garden skink. They tested for the presence of different decision making strategies by exposing the lizards to choices of different prey items. This is an interesting study, although I think the methods as currently presented are insufficient to determine what was done in this study.  
  
Main comments

Corticosterone treatments – There is not enough information presented on how treatments were conducted. For example, how often were eggs treated? How old were the eggs when treated?

RESPONSE:   
  
Behavior testing – It is unclear when these tests were conducted. How old were the hatchlings when testing began? When was each test done, how often per individual? How many days between testing before an animal was tested again? Was the order of testing randomized? If so, how? If not, how did you control for the order of testing?

RESPONSE:   
  
Discussion – one question that came to mind while reading the discussion was related to the fasting paradigm that was employed. As the animals were fasted prior to behavior testing, is it possible that they did not discriminate between different foraging options because they were hungry? Perhaps the lizards just ate indiscriminately and would have made different choices if they weren’t so hungry. I think the discussion would benefit from a few sentences discussing this idea.

RESPONSE:   
  
  
Minor comments

Line 133 – “by” should be replaced with “for”  
Line 170 – “follows” should be replaced with “followed”  
Line 177 – The b in based should be capitalized.

Reviewer: 2

Comments to the Author

In this paper, the authors conduct a series of two-choice tests to determine the ability of common garden skinks to discriminate between small numeric quantities (1 vs. 4, 1 vs. 3, 2 vs. 4, 2 vs. 3 and 3 vs. 4) in a food-related context. Prior to being tested, they exposed skinks to the combination of two incubation conditions: temperature (High/Low) and corticosterone levels (control/CORT) in a 2x2 factorial design to determine whether these conditions affect their cognition abilities in the two-choice tests. They found no evidence of quantity discrimination and no differences among the 4 groups of early-life conditions, as skinks were as likely to choose the large and small quantity of food during tests and showed a similar latency and interest in the large amount of food. However, they did find some differences depending on the sex and age of individuals on the latency of choice. They conclude that common garden skinks do not seem to discriminate small quantities of food and these choices are not affected by early-life conditions (temperature and corticosterone).

Despite the lack of positive results, I find the methodology and paradigm of this study to be ambitious and its findings could offer valuable insight to other researchers conducting studies on this topic in these underrepresented taxa. The sample sizes (although unclear, see specific comments) seem adequate, the manuscript is well written, and the conclusions are based on the results described within. There are, however, important details missing from the methods regarding the sample sizes, cort treatment and some trial details, and I missed a more in depth evaluation of the potential flaws in the design and more specific examples of how these results may make sense based on the biology and ecology of the study species.  
I have made (perhaps too many) specific comments to clarify the above, which I hope the authors find useful in revising their manuscript.

Specific comments:

L25-29: The way this is written here is almost like the/one of the goals of this study was to test whether OFS or ANS is used in quantity discrimination in L. guichenoti but in L29-L32 this aspect is not mentioned. If that was indeed the goal, L30 could read (e.g.): “…to discriminate between quantities using either OFS or ANS and the effects of prenatal GCs…” or something like that. I’m left wondering how important is to discuss the difference between OFS and ANS systems (particularly in the abstract)? If I understood correctly, all of the results in the main MS are from OFS tests, and there is no explicit mention in the second half of the abstract to which system was tested or how the results link to those systems.

L54 replace “etc” by “e.g.”: “(e.g. in number, size, volume)”  
-In several sections of the introduction, the argumentation seems supported by a single citation. E.g. Agrillo and Bisazza 2014 in L63-70 and L80-84, Hyde 2011 in L71-76. Are these references reviews? If not, I would encourage authors to include more bibliographic support. If they are reviews, please consider whether citing some of the original papers within the review is more appropriate (e.g. if the review is only a summary of existing research).

L79-80 Wouldn’t this be the case only when quantities are small? If quantities are large, OFS would not be used regardless of the presence/absence of continuous cues, according to L76: “[OFS] is thought to be limited to small quantities, typically up to three or four”

L82-83 Would relying on cumulative surface area (a continuous trait) in this example indicate ANS is being employed? Conversely, would relying on numerical information alone imply OFS? It would be helpful to clarify this more explicitly here.

L112-115 It is unclear whether these stress responses are happening to embryos (directly) or to mothers (indirectly) and how (the mechanism) are GCs being transmitted to offspring. Adding a specific example may be illustrative.

L108-121 As currently presented, singling-out temperature and GCs among the array of early-life conditions feels somewhat arbitrary (e.g. L107 “For instance, temperature…”). Why are these two specific factors comparatively more important? I’m sure there are other developmental factors that also affect cognition. I guess my point is I find the justification for studying these two in particular slightly lacking. Are reptiles experiencing incubation temperatures outside optimal ranges frequently? Are they at risk of experiencing them more now with climate change? Are there any other ways in which reptiles can experience elevated prenatal GCs or is this simply a side-effect of temperature stress? How prevalent are these effects across species?

L124-125 replace hyphens with an M dash

L136-140 These predictions do not quite follow from the information presented in the introduction. E.g. Why is it expected that lizards incubated at higher temperatures will outperform those incubated at lower temperatures? For example, when discussing the impacts of temperature (L108 onwards) the nature of the effects of temperature (high/low temperature is good/bad) on cognition are not mentioned. They are also not mentioned here as a way to justify this prediction. I assume some of these references (L137) are studies where the results have been in this expected direction? If so, please say so and expand on the basis of these predictions.

L143 What is life expectancy for garden skinks? My understanding is that this is a short-lived species (2-3 years max?) and thus the lizards tested here are the 2nd-3rd (or more?) generation born in captivity in this colony. I did not see any discussion on the potential effects of using lab-bred versus wild-captured animals in cognition tests, which may be warranted, especially in light of the lack of significant results. Is inbreeding a concern within the colony?

L152 Were eggs incubated in isolation or were complete clutches placed in the same cup? If clutches were kept complete, clutch size may have affected the microenvironment experienced by the eggs. Was there any egg/hatchling mortality? This is important information to add in my opinion.

L155 Egg length and mass was measured. I may have missed it but was anything done with these data? If nothing, this information is perhaps superfluous.

L173 please justify why these temperatures were chosen. Move L180-182 here.

L175 What was the frequency in the application of CORT and the duration of the treatment?

L176 100% Ethanol was used as the vehicle for CORT. I’m familiar with several studies where CORT was mixed with sesame oil for transdermal application in lizards, but this is the first time I hear of ethanol. Is topically applied ethanol not harmful to tissues? Could the authors provide some references for this approach?

L179 Please give the value of the dose used. The dose of CORT was chosen based on a study on another species (same genus). Is there any information about the natural range of CORT levels on the focus species? How do we know a ~3.7x difference is biologically meaningful?

L186 I’m curious why clutches under 4 eggs were not excluded from the analyses as an alternative approach? Data on the number of large (≥4 eggs) and small (<4 eggs) clutches included in the analysis would be helpful to assuage concerns on potential methodological biases.

L188 Does the sample size of the experiment consist of hatchlings from the above described clutches tested after at least ≥2 weeks of life as this seems to imply? Neither the number, the sex nor the age of the lizards tested in this study is provided in the methods. This is essential information that is missing. The only sample sizes given that I can see are the “n=20” in the legends of the figures. The only reference we have is the 180 individuals that were used to establish the colony in 2019, but this tells us nothing about the current experiment. It is especially relevant to know how was the distribution of early-developmental treatments (Temp+CORT levels) across the different categories (sex, age). Were these balanced?

L206 one->once. Please provide somewhere the total number of tests/trials performed.

L208 I have some concerns about the use of frozen crickets rather than live prey for the cognitive tests since many of the cues that lizards use to forage (and possibly “count”) are likely strongly affected by this. Is there any data on how appealing (for lack of a better word), these crickets are to skinks in comparison to live crickets? They are eating them so there is that, but this may have factored in in the results (i.e. the lack of evidence for quantity discrimination). In my experience conducting behavioral experiments with lizards, the responses elicited can be significantly impacted when using artificial versus real stimuli and lab versus free-ranging lizards. I understand there were reasons behind this choice, but I can think of alternatives, such as live crickets that have been cooled (rather than frozen), or have hind legs removed (not sure on the ethics on this one), or the use of smaller containers for the crickets, to minimize movement. It may have been more difficult to control for the effect of some continuous variables in this case but since the quantities being compared are still small, and the methods to control for length and total area did not completely remove the differences anyway (L215-216), an educated guess of the discrimination system being used could still have been made.

L209-210 This adds to the confusion regarding the age of the lizards tested and the timeline of the experiment. Is this month overlapping with the 2 weeks of acclimation to the arenas or did this happen before/after?

L213-214 If I understand this correctly, the actual comparison in length was done after, not before the numerical testing (using the videos recorded during the experiment). Thus, the change in the orientation of the crickets was done somewhat ‘blind’ as to what effect this would have (whether it would effectively remove the differences in length)? It does not seem ideal and I’m not sure it should be described in the methods as a part of the experimental design, or simply discussed in the results/discussion. Maybe keeping only L211-213 here makes more sense. In addition, if “total area” was not measured it should be removed from L213.

L215-220 The results of these experiments and are better featured in the Results section (e.g. 215-218 and L220-221 should be moved to results).

L220-221 Was this test done using different lizards than those tested for quantity discrimination? It could be argued that orientation and length are intrinsically correlated and both of these tests (L214, L220) are comparing length and orientation simultaneously. In order to split these factors, the prey orientation test needs to specify whether a horizontally- and a vertically-positioned cricket had the same length or not (as measured in the control of size test).

L225-226 “the higher number of crickets was randomly placed on the right or the left platform” should be moved to somewhere around L206-210 since it refers to the experimental design, not the habituation phase. During habituation, only 1 cricket was fed at a time per L197 and this was confusing to read here. I think L209-210 should be moved here or this (L222-224) moved to L210 when speaking on the habituation.

L226-227 It is always clearer to provide specific numbers than general statements. There’s a lot of confusion stemming from the lack of clear statements on the sample sizes used in this study. Maybe Fig.1 split the text in separate paragraphs but this phrase would better placed in L229.

L229 What was the duration of a trial? Provide the details on when a trial began (when the platform is left in the enclosure per L230) and when it ended.

L236 I have some trouble understanding the variable Interest. After the first interaction with one option, the time interacting with that option was recorded for a maximum of 5 minutes. Does this mean you stopped the timer every time a lizard walked away from that option after having first interacted with it and resumed it if it came back, up to a maximum of 5 min? Information on trial duration is needed to gauge what these 5 minutes represent. Here, it seems that latency will determine how much time is left for lizards to interact with one or both options before the trial ends, which makes trials that showed longer latencies more likely to be those where lizards only interacted with one of the options, perhaps overestimating interest in the first choice. Interest could perhaps be better measured by accounting for this limitation with respect to the available time left. Or trial duration could have been adjusted to a fixed time starting to count after the first interaction occurred, making it independent from latency. Clarification on trial duration is needed.

L241 Indicate how many cases were these.

L244 Same here, how many lizards were excluded?

L249 Is this one cricket “per day”? Add the frequency for clarity. In my experience housing different species of lizards, one cricket per day, or 2-3 crickets every other day can be a perfectly fine feeding rate that keeps lizards satiated. Are these small crickets or what percentage of body weight does one cricket represent?

L253-254. If this happened in n=0 cases, then no trials were excluded.

L268 Specify how you coded the predictor age (e.g. measured in days?).

L268-270 “We did not find…” If this did not lead to any changes in the model specification, remove it from here and mention it only in the results.

L270 Could you add a reason why age was centered to zero and the predicted values for males and females averaged?

L272 add “identity” after clutch as well. Or does this refer to first or second clutch?

L297 Do any of these estimates have units?

L301 What does “performance” mean in this context? If it refers to latency, choice and interest, I’d suggest using that directly for improved clarity.

L325-328 How is an insectivorous diet preventing quantity discrimination exactly? This is not explained. This sentence and the following argumentation is also mixing the type of diet (insects vs. vegetables) with quantity (few vs. many) with status (alive vs. dead prey). These are all good points that could be made separately.

L328 It’s unclear to me what the Podarcis example is contrasting since they also failed to discriminate numerical quantities.

L336-340 Based on its ecology and foraging behavior, is there any indication that any of this could be happening in the common garden skink in particular? Again, I would prefer if specific references were used here instead of Agrillo and Bisazza 2014, provided none of this is a novel conclusion stemming from the review paper itself but from the original research featured in the paper.

L343-346 This was however not presented as a specific aim of this study in the introduction, specifically with regards to the OFS part. If this was indeed the aim, some adjustment to L122-140 is needed to make the role of OFS more explicit.

L346 Check “lizards decision” for grammar

L347 See comment L213-214 above. The method for controlling for this variable was only partially successful. I’m not sure it should be described as part of the experimental design since it’s not quite accomplishing what it wanted, suggesting that it wasn’t implement correctly.

L360 Is this conclusion based on “latency” alone?

L373-374 According to the methods (L254) there were zero cases where lizards failed to interact or did not eat the cricket so no trials were excluded.

L376-379 I find this result interesting and I think it’d be useful if the authors could provide a little more context based on the literature in order to give an idea of how extended cognition differences between sexes and ages are in reptiles and how these may connect to differences in early developmental environment.

L399 “decision-making” and “foraging” are rather broad concepts and thus I find this conclusion somewhat misleading. It is difficult to prove a negative (no effect of early conditions on quantity discrimination) based on the lack of evidence, especially when practically no studies have been conducted on this topic. I think a better approach would be to be more specific to quantity discrimination.

L403 Is this based on the Podarcis sicula study alone? The discussion starting at L325 should be better developed. Reading through that section it wasn’t clear what relationship is there between the lack of quantity discrimination and an insectivorous diet. What is the hypothesis? What about the insectivorous diet could be responsible for this? How does this fit into the biology and ecology of this skink species (see also comment L336-340)?

L403-L405 I think a more in depth discussion on this particular point would be very valuable to readers and it is warranted. In L386 it is suggested that other approaches or contexts may be more affected by early-life conditions. What are some examples of these other contexts? See also comment on L208 above about the use of frozen crickets as prey stimulus. What are some of the caveats and limitations of the experimental design and paradigm used here? What could have been done differently? Is this a good system and species to conduct cognition studies? Have you or other authors conducted previous studies using these two-choice tests and platforms? These are some ideas off the top of my head.