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**: Thank you for this helpful comment. Eggs were dosed only once, immediately after collection and prior to incubation at the different temperatures. We clarified the timing and frequency of the corticosterone treatment in the revised manuscript: “Eggs were allocated to one of two developmental stress treatments: CORT-treated eggs were topically supplied with 5 µL of a CORT solution (10 pg/mL) once, immediately after collection and before incubation. Control eggs received an equal volume of 100% ethanol.” (LINES).

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**: Thank you for these detailed questions regarding the behavioural testing protocol. We clarified the text to ensure all aspects are explicit. Specifically:

- Age of hatchlings: Animals were between 106 and 136 days old when testing began. This information has now been added: “The lizards were between 106 and 136 days post-hatching (see Fig. S5 in *Supplementary Material*).” (LINES).

- Timing and frequency: Each numerical test was performed only once per lizard, we have clarified this in the revised manuscript: “We performed five numerical discrimination tests per lizard, each differing in the ratio between both choices of crickets: 1 VS 4, 1 VS 3, 2 VS 4, 2 VS 3, 3 VS 4 crickets. Each numerical discrimination test was made only once per lizard.” (LINES). Tests were conducted daily between 1000 and 1200 h across five days (17–22 May 2024) (LINES).

- Order of testing: The order of testing was counterbalanced across treatments. On each day of testing, an equal number of individuals from each treatment group were subjected to different numerical tests. To improve clarity, we rephrased and repositioned this information in the methods section as follows: “We counterbalanced the order of testing by ensuring that, each day, an equal number of individuals from each treatment were tested in each type of trial.” (LINES).

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**: Thank you for this thoughtful suggestion. We considered this possibility but find it unlikely, as the results of our discrimination index closely mirror first-choice behaviour. If hunger was the main factor influencing lizard behaviour, we would expect first choices to be indiscriminate, but not the following interactions. We have added a paragraph in the Discussion to address this point: “We found no evidence of numerical discrimination in any of the tests performed. One potential concern is that pre-test fasting could have caused indiscriminate foraging, masking numerical discrimination. This seems unlikely: if hunger alone were driving behaviour, we would expect random first choices but clear preferences during subsequent interactions, reflected by the interest score. No such pattern was observed, suggesting that hunger levels do not explain our results.” (LINES).

Minor comments

Line 133 – “by” should be replaced with “for”

Corrected

Line 170 – “follows” should be replaced with “followed”

Corrected

Line 177 – The b in based should be capitalized.

Corrected

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.

**RESPONSE**: We greatly appreciate the reviewer’s careful and insightful feedback. Our responses to the comments are detailed below.

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.

**RESPONSE**: We appreciate the reviewer’s observation. Our intention was not to directly test whether *L. guichenoti* uses the OFS or ANS, but rather to explore quantity discrimination in OFS-related context (i.e., small numbes). We agree that the way the abstract was initially written could give the impression that testing the use of OFS vs. ANS was a primary aim of the study. We have now revised the relevant sentences to clarify this issue: “Quantity discrimination affects a range of behaviours essential for fitness, including social interactions, navigation, and foraging and is widespread across animal taxa. However, in reptiles, evidence for this ability—particularly for small number contrasts—remains limited and mixed. Early-life conditions, such as hormonal exposure and temperature, can also shape brain development and cognitive performance, potentially affecting numerical abilities.” (LINES).

L54 replace “etc” by “e.g.”: “(e.g. in number, size, volume)”

Corrected

-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).

**RESPONSE**: We thank the reviewer for this helpful observation. Both Agrillo & Bisazza (2014) and Hyde (2011) are review papers that provide broad overviews of methodological approaches and theoretical frameworks within the field of numerical cognition. We used them specifically when referring to general concepts or methodological caveats, where a summary source is appropriate. However, we have revised the relevant sections to include additional primary sources where needed.

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”

**RESPONSE**: Thank you for this observation. We agree that OFS is typically limited to small quantities. However, even for small quantities, animals might rely on the ANS if continuous cues are available. To clarify this, we have rephrased the sentence in the manuscript as follows: “However, when discriminating between small quantities, the engagement of these systems also depends on contextual factors such as the availability of continuous variables (e.g., area, density, movement) and the nature of the task (Stancher et al. 2015).” (LINES).

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.

**RESPONSE**: We appreciate the reviewer’s insightful comment and have revised the sentence to make this distinction clearer: “In spontaneous choice tests, for instance, animals may rely on cumulative surface area (ANS-related) rather than numerical (OFS-associated) differences when selecting between two sets of food items (Agrillo and Bisazza 2014).” (LINES).

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.

**RESPONSE**: We thank the reviewer for this suggestion and have added a specific example to illustrate how stress responses can be transmitted to offspring: “For instance, stressing pregnant rats (*Rattus norvegicus*) suppresses neurogenesis in the dentate gyrus of the offspring, associated with impairments in spatial learning (Lemaire et al. 2000).” (LINES).

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?

**RESPONSE**: We thank the reviewer for this thoughtful comment. We agree that early-life conditions include a wide range of environmental factors that can influence cognitive development. In our study, we focused on incubation temperature and prenatal CORT because both are ecologically relevant sources of phenotypic variation in reptiles. Subtle differences in incubation temperature—even within natural ranges—have been shown to affect neurodevelopment and cognitive performance, while stress-related hormones such as CORT can influence both brain function and behaviour. Additionally, thermal conditions, particularly those near the edges of the optimal range, can activate physiological stress responses, making it important to understand how these factors interact. We have clarified these points in the revised introduction and modified the relevant sentence for clarity: “Developmental conditions can strongly influence cognition because the brain is particularly sensitive to environmental inputs in early life stages (Zhu et al. 2004). In ectotherms, the temperature at which embryos develop have significant effects on several traits (Noble et al., 2018). For example, incubation at warm temperatures increases neuron density and learning abilities in different species of reptiles (Amiel and Shine 2012; Clark et al. 2014; Amiel et al. 2014; Amiel et al. 2017; Dayananda and Webb 2017; Siviter et al. 2017; Abayarathna and Webb 2020). Importantly, thermal effects on cognition may arise directly or through parental effects (Crino et al. 2023). Thermal fluctuations, particularly those approaching the edges of an organism's optimal range can trigger the stress response, leading to the release of glucocorticoids (GCs) (Sapolsky et al. 2000), which may then be transmitted to offspring (Crino et al. 2023). Elevated prenatal GCs can, in turn, influence cognition by altering brain function (Szuran et al. 1994; Zhu et al. 2004; Crino et al. 2014; Farrell et al. 2015; Bebus et al. 2016; Lui et al. 2017). For instance, stressing pregnant rats (Rattus norvegicus) suppresses neurogenesis in the dentate gyrus of the offspring, associated with impairments in spatial learning (Lemaire et al. 2000). As such, the interaction between GCs and developmental temperature may be relevant in shaping reptile cognition, especially under the context of climate change (Crino et al. 2023). Most studies on the effects of GC exposure or elevated temperatures during development focus on learning and memory, but few studies have investigated how early-life conditions can affect quantity discrimination despite its importance on animals’ lives (but see Vila Pouca et al. 2019). Furthermore, the interactive effects of prenatal GCs and temperature on cognitive abilities remain largely unexplored.” (LINES).

L124-125 replace hyphens with an M dash

**RESPONSE**: We have corrected this here and throughout the text.

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.

**RESPONSE**: We appreciate the reviewer’s comment. We have included some information in this regard: “Specifically, we hypothesized that lizards incubated at higher temperatures would outperform those incubated at lower temperatures in the numerical discrimination task as previous studies have reported enhanced cognitive abilities in ectotherms incubated under warmer conditions (Amiel and Shine 2012; Clark et al. 2014; Amiel et al. 2014; Amiel et al. 2017; Vila Pouca et al. 2019).” (LINES).

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?

**RESPONSE**: We thank the reviewer for raising this point. Garden skinks can live over five years in the laboratory, and the individuals tested here were mostly first- or second-generation offspring from wild-caught parents. Inbreeding is therefore unlikely to be a concern in our colony. As most studies on early environment and cognition use laboratory-reared animals, we consider our approach appropriate and have not expanded on this in the manuscript.

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.

**RESPONSE**: Thanks for this comment, we address both points below.

- Eggs were incubated individually, not together with clutchmates. We have rephrased the methods to improve clarity: “Then, we treated eggs with CORT or a vehicle control (see Manipulating Early Thermal and CORT Environments details below) and placed each egg on individual cups (80 mL) with moist vermiculite (12 g water to 4 g vermiculite) and covered the cups with plastic wrap to retain moisture. Each egg was incubated individually in incubators (LATWIT 2X5D-R1160) programmed to two different thermal regimes (see Manipulating Early Thermal and CORT Environments details below) until hatching.” (LINES).

- Some egg and hatchling mortality occurred. However, this experiment was part of a larger project involving thousands of eggs, and detailed survival data will be reported separately. For the present study, we note in the manuscript that mortality did not exceed 13% in any treatment: “Egg and hatchling mortality was not higher than 13% in any of the treatments (Leibold *et al*., In prep.).” (LINES).

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.

**RESPONSE**: Thank you for noticing this. Egg length and mass were recorded as part of a larger project but were not used in the analyses presented here. We have removed this sentence from the methods to avoid confusion. We have done the same with hatchlings mass and length.

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

**RESPONSE**: Thank you for this suggestion. We respectfully believe that the current order provides a clearer structure. In this section, we first introduce both manipulations (CORT and temperature) and then justify each in turn. We therefore prefer to retain the current order, as moving L180–182 would disrupt this logical flow.

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

**RESPONSE**: Thank you for this helpful comment. Eggs were dosed only once, immediately after collection and prior to incubation at the different temperatures. We clarified the timing and frequency of the corticosterone treatment in the revised manuscript: “Eggs were allocated to one of two developmental stress treatments: CORT-treated eggs were topically supplied with 5 µL of a CORT solution (10 pg/mL) once, immediately after collection and before incubation. Control eggs received an equal volume of 100% ethanol.” (LINES).

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?

**RESPONSE**: Thank you for raising this important point. As noted, sesame oil is commonly used as a vehicle for transdermal CORT application in live lizards. However, when working with eggs, ethanol is more common (e.g. Carter et al., 2016, *Physiol. Behav*.; Carter *et al*., 2018, *J. Exp. Biol.*). In our experiment, we adapted the concentration and volume of ethanol to be appropriate for our model species. We have now cited relevant studies using ethanol as a vehicle in the revised manuscript to clarify and support our methodological choice.

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?

**RESPONSE**: We appreciate this comment. In a previous experiment, we found that this dose increased yolk CORT concentrations by more than two standard deviations above control levels (Crino et al., 2024). We have clarified this in the Methods: “This CORT dose increased mean yolk CORT levels 2 standard deviations higher than those of control eggs in previous studies (Crino et al., 2024).” (LINES). We therefore consider this manipulation to be 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.

**RESPONSE**: Thank you for raising this point. We accounted for potential pseudoreplication by including clutch identity as a random factor in all models, which controls for variation among clutches regardless of their size. Excluding small clutches (<4 eggs) would have unnecessarily reduced sample size and statistical power without providing additional control beyond the random effect structure. We note that all data, including clutch sizes, are publicly available with this manuscript, so these details can be verified directly.

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?

**RESPONSE**: Thank you for this comment. We acknowledge that the sample sizes could be reported better. We have now included a paragraph at the beginning of the results where we report all this information. “We recorded and analysed a total of 400 videos from 80 lizards (n = 20 per treatment, see Fig. 1). The lizards were between 106 and 136 days post-hatching (see Fig. S5 in *Supplementary Material*). Final sample sizes were composed by 43 females and 37 males, with no less than 6 females per treatment and no more than 13.” (LINES).

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

**RESPONSE**: We appreciate the reviewer’s attention to detail, we have rephrased the sentences to avoid confusion: “We performed five numerical discrimination tests per lizard, each differing in the ratio between both choices of crickets: 1 VS 4, 1 VS 3, 2 VS 4, 2 VS 3, 3 VS 4 crickets. Each numerical discrimination test was made only once per lizard.” (LINES).

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.

**RESPONSE**: Thank you for raising this point. We agree that prey movement can influence lizard choices, which is precisely why we used frozen crickets — to ensure they remained stationary during the tests. We considered alternatives such as cooling live crickets or removing legs, but these approaches either are unlikely to eliminate movement, could alter prey perception, or raise ethical concerns. Using smaller containers would also not fully solve the issue of prey movement. To minimize potential biases while maintaining naturalistic perception, we habituated lizards to the thawed crickets prior to testing, which we believe allowed us to assess numerical discrimination without confounding effects of movement.

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?

**RESPONSE**: Thanks for the attention to detail. We have clarified this in the manuscript: “After acclimatization to the arenas and before the numerical tests, lizards were habituated to eating thawed frozen crickets for one month before the experiments.” (LINES).

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.

**RESPONSE**: We appreciate the reviewer’s suggestion and understand the concern. Our approach aimed to control for continuous variables such as length and total are, as in previous studies (see Stancher *et al*., 2015, *Anim. Cogn.*). However, while many studies assume that visual adjustments are sufficient to control for these cues, they do not always verify this directly. In our case, we chose to assess whether our method was effective by measuring the actual length of the stimuli post hoc using video recordings. Since the crickets used were of similar size and shape, we only measured length and assumed that it is a reliable proxy for total area in this context. Therefore, and based on our analysis, we believe that continuous variables were effectively controlled in most numerical contrasts, except for the largest ratios (1 vs. 4 and 1 vs. 3), where minor differences remained. Therefore, we believe that a) we should report our approach in the methods, and b) that total area was controlled for in parallel with total length.

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).

**RESPONSE**: We appreciate the reviewer’s perspective, and we understand the reasoning. However, we consider this analysis a methodological control rather than a core experimental result. For this reason, we included this information in the Methods section. However, to make sure that this is not interpreted as part of the results, we have replaced “Our results” with “This analysis” (LINES).

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).

**RESPONSE**: We thank the reviewer for this thoughtful comment. The prey orientation test was conducted using the same lizards as those in the quantity discrimination trials, and we have now clarified this in the manuscript: “In addition, after the experiments, we subjected the lizards to another set of trials to test whether the orientation of the crickets could affect lizards’ choice, but we found no effect of orientation (see *Supplementary Material: Prey orientation test*).” (LINES).

We agree with the reviewer that orientation and length may be correlated, and we acknowledge that our current design does not disentangle the use of these variables. However, since we found no preference for either orientation, we believe this indicates that neither orientation nor any associated visual cues influenced lizards’ choices in this context which was the goal of this additional 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.

**RESPONSE**: Thank you for this suggestion, we have moved the sentence up in the text (LINES).

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.

**RESPONSE**: Thank you for this comment. We have included a paragraph at the beginning of the results with sample sizes per group and sex, and information on the age of the lizards. “We recorded and analysed a total of 400 videos from 80 lizards (n = 20 per treatment, see Fig. 1). The lizards were between 106 and 136 days post-hatching (see Fig. S5 in *Supplementary Material*). Final sample sizes were composed by 43 females and 37 males, with no less than 6 females per treatment and no more than 13.” (LINES).

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.

RESPONSE: Thank you for this suggestion, we have included this information in the text: “Every trial began when we left the platform with the crickets in the enclosure, and lasted for 90 minutes until we took the platform away.” (LINES).

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.

**RESPONSE**: We thank the reviewer for their careful reading and helpful comments. We have clarified the description of the Interest variable in the revised manuscript: “Interaction was defined as touching the Petri dish with the snout or having the front half of the body on the platform. For each option, and regardless of latency, we recorded the time the lizard spent interacting with it for five minutes after the first interaction. The timer was not paused if the lizard walked away, so the recorded time reflects total engagement with each option during the first five minutes following initial contact. We calculated *Interest* by subtracting the time spent interacting with the smaller food option from the time spent interacting with the larger one.” (LINES).

L241 Indicate how many cases were these.

**RESPONSE**: There were only 11 cases in total where the lizards did not interact with any choice (n = 4) or ate the crickets from one of the options (n = 7 observations) (see LINES).

L244 Same here, how many lizards were excluded?

**RESPONSE**: No individuals were excluded from the analyses; information added in LINE.

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?

**RESPONSE**: Thank you for your comment. Each lizard received one cricket per day; we have included this information in the manuscript (LINES). While we cannot provide the exact percentage of body weight represented by each cricket, previous studies have shown that this feeding regime is appropriate to maintain motivation for foraging in these lizards (see Recio *et al*., 2025, *Anim. Behav.*; Recio *et al*., 2025, *Behav. Ecol.*)

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

**RESPONSE**: No trials were excluded; however, we believe it is important to report how we controlled for motivation and our planned criteria for exclusion, even though no trials ultimately met those criteria.

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

RESPONSE: Thanks for the attention to detail. We have included this information now (see LINES).

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.

**RESPONSE**: Thanks for this suggestion, we have done as recommended.

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

**RESPONSE**: We have added a brief explanation for these decisions in the revised manuscript: “Age was centered to the mean to facilitate interpretation of model estimates, and posterior predictions were averaged across sexes to provide estimates independent of the reference category.”.

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

Corrected

L297 Do any of these estimates have units?

**RESPONSE**: The units of the raw estimates depend on model’s link function reported in the data analysis section (log-odds for Choice, log for Latency). We report them on this scale because the values for age and sex are mainly included to control for their effects, rather than being the main variables of interest, so we believe that back-transforming is not necessary for interpretation in the context of our study.

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.

**RESPONSE**: We have clarified this by specifying the variables we were referring to.

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.

**RESPONSE**: Our hypothesis is that diet influences decision-making in spontaneous numerical discrimination tests, at least with small quantities. Studies using vegetables show that lizards can discriminate small food quantities, whereas insectivorous lizards typically do not. Importantly, this does not reflect an inability to discriminate, as *Podarcis siculus* can discriminate between the same quantities when trained (Miletto-Petrazzini *et al*., 2017, *Biol. Lett*.; Miletto-Petrazzini *et al*., 2018, *Front. Psychol*.). Instead, it suggests that these lizards do not rely on numerical information in foraging contexts, possibly because the actual probability of obtaining a prey item (e.g., a cricket) is similar regardless of the number present. Therefore, insectivorous lizards may not use numerical cues when making foraging decisions between options differing in item number, at least for small quantities. We have modified this part of the discussion to better reflect our hypothesis: “The lack of spontaneous numerical discrimination in L. guichenoti when presented with small quantities of food may be related to their insectivorous diet. Studies in which reptiles successfully discriminated small food quantities typically involved vegetables as stimuli (Gazzola et al., 2018; Szabo et al., 2021). In contrast, insectivorous reptiles tested in spontaneous choice tasks with fewer than four prey items generally failed to discriminate (Miletto Petrazzini et al., 2017). For example, Podarcis siculus did not distinguish between small numbers (never more than four) of Musca domestica larvae in a spontaneous choice test, although the same species was capable of quantity discrimination when prey size was the primary cue (Miletto Petrazzini et al., 2017) or when tested using a training paradigm (Miletto Petrazzini et al., 2018). These results suggest that the inability to discriminate in spontaneous tests is not due to a lack of cognitive capacity, but rather to decision-making strategies shaped by diet. In insectivorous lizards, the actual probability of obtaining a single prey item may be similar regardless of whether one or several items are present. Consequently, even though insectivorous lizards can discriminate small quantities under other conditions (see Recio et al., 2021), they may not use numerical cues in foraging decisions with small prey, as selecting larger numbers may not be advantageous.” (LINES).

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

**RESPONSE**: Because the studies in the previous sentence report discrimination: “In studies where reptiles **discriminated** between different amounts of food, the relevant stimuli involved either vegetables (Gazzola et al. 2018; Szabo et al. 2021) or large numbers or alive prey (e.g., 5 vs. 10) (Recio et al. 2021).” We have modified this paragraph to improve overall clarity (see comment above).

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.

**RESPONSE**: We thank the reviewer for this useful suggestion; we agree that including species-specific ecological context can strengthen the discussion and have revised the text accordingly: “*Lampropholis guichenoti* is described as a generalist insectivore that locates prey through a combination of active foraging and sit-and-wait behaviour (Lunney et al., 1989; Mo & Mo, 2021). Although handling time may vary with prey size, field observations suggest relatively slow feeding rates. For example, Martin (2015) reported that when feeding ant alates, individuals consumed an average of only two alates every 15 minutes even when the prey was abundant. In such situations, the difference in payoff between small prey group sizes (e.g., 1 vs. 4) may be constrained by capture probability and effort, reducing the selective advantage of relying on numerical cues in foraging decisions. ” (LINES).

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.

**RESPONSE**: Thanks for this suggestion. We have emphasised our goal in the introduction: “Our aim was to understand if this species could use the OFS to discriminate between small quantities of prey in a spontaneous choice test.” (LINES).

L346 Check “lizards decision” for grammar

Corrected

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.

**RESPONSE**: We appreciate the reviewer’s comment and understand the concern. While the control for continuous variables was only partially successful in some contrasts, we believe it is important to report the steps we took as part of the experimental design and to be transparent about their effectiveness. Although the method did not eliminate all differences — particularly in the 1 vs 3 and 1 vs 4 conditions — the design was largely effective in the remaining tests. Please refer to comment L213–214 above for further details.

L360 Is this conclusion based on “latency” alone?

**RESPONSE**: Yes, this conclusion is based on latency, which we considered a reasonable proxy for decision-making, as it reflects the time taken by the lizards to make a choice. We have now clarified this in the manuscript by using terms such as "decision time" and "time to make a decision" to better convey the interpretation of latency in this context.

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.

**RESPONSE**: The reviewer is correct that no trials were excluded. However, we wanted to clarify that we controlled for motivation in multiple ways. We have revised the sentence for clarity: “Additionally, we observed that all lizards consumed the cricket immediately after the test, showing similar levels of motivation.”.

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.

**RESPONSE**: We thank the reviewer for this comment. While investigating age- and sex-related differences in cognition is interesting, it is beyond the scope of the current study. We note, however, that previous work in reptiles has shown weak effects of sex on cognitive abilities in reptiles (e.g., Szabo *et al*., 2019, *Behav. Process.*), while age seems to be more significant (e.g. Carazo *et al*., 2014, *Proc. R. Soc. B Biol. Sci.*). Nevertheless, further studies would be needed to explore how these factors interact with early developmental environment. We have included those references and a short sentence in the manuscript: “Previous work suggests weak effects of sex on cognition in reptiles (Szabo et al., 2019), while age may have a stronger influence (Carazo et al., 2014). Nevertheless, further studies are needed to explore how age and sex interact with early-life conditions in shaping behavioral responses.”.

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.

**RESPONSE**: We do agree with the reviewer that, given the lack of evidence for quantity discrimination, we should be cautious in generalizing conclusions about broader cognitive processes. However, we believe that the finding that decision time (Latency) was not affected by prenatal temperature or CORT exposure is robust within our experimental design. We have now clarified this issue: “Our findings suggest that L. guichenoti does not rely on numerical information for foraging decisions, which limits the potential for early-life conditions to influence this specific ability. However, our results show that decision time does not vary due to prenatal temperature or CORT exposure, suggesting that this aspect of decision-making can be robust to early environmental conditions, at least in the context of foraging behaviour” (LINES).

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)?

**RESPONSE**: Thanks for you comment. We have modified this part of the discussion to make it clearer (see the response to comment L336-340 above).

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.

**RESPONSE**: We thank the reviewer for these thoughtful reflections. We agree that expanding on these points helps clarify the context and limits of our conclusions. We addressed each of the questions raised below.

- Alternative contexts. In response to the question about alternative contexts in which early-life conditions might have a greater effect, we meant that, if numerical discrimination is tested using different stimuli or a different approach such a learning paradigm, these experiments may yield different results both on numerical discrimination itself and on the effects of early-life conditions on this cognitive ability. We have clarified this in the text: “For instance, testing numerical discrimination using learning paradigms yielded different results than spontaneous choice tests in *P. siculus* (Miletto Petrazzini et al., 2017; 2018). In fact, the only study we found that examined how developmental conditions influence numerical abilities showed that Port Jackson sharks (*Heterodontus portusjacksoni*) incubated at warmer temperatures learned a numerical rule faster than cold-incubated individuals (Vila Pouca et al. 2019), but discrimination in a spontaneous test was not investigated. Given the potent effects of early environment on cognition and brain development (Szuran et al. 1994; Zhu et al. 2004; Amiel and Shine 2012; Clark et al. 2014; Amiel et al. 2014; Lui et al. 2017; Amiel et al. 2017), it is crucial to explore how early environmental conditions shape numerical abilities across taxa employing different methodologies.” (LINES).

- Frozen crickets. Regarding the use of frozen crickets, as noted in our earlier response (L208), this was a deliberate design choice to eliminate motion cues and isolate numerical information. We believe it was necessary to avoid confounding effects of movement, and have briefly acknowledged this trade-off in the revised discussion.

- Experimental caveats. Like in other experimental approaches, spontaneous numerical discrimination tests rely on certain assumptions—particularly that animals are motivated to maximize food intake and will prefer the larger quantity. This assumption may not always hold, and we now include a brief note about this limitation and how learning-based paradigms could complement the current approach: “For example, employing learning paradigms where animals must learn to follow a specific numerical rule can yield different results.” (LINES).

- *L. guichenoti.* as a model. Prior studies have demonstrated this species’ learning ability and behavioral flexibility (e.g. Recio *et al*., 2025, *Anim. Behav.*; Recio *et al*., 2025, *Behav. Ecol.*), supporting its potential for cognition research.

- Platforms used. While the platform employed in this study has not been used previously, similar paradigms have been widely used in studies of numerical cognition across taxa. Therefore, we do not consider the platform to be an important factor in our results.