Peer Review File

The evolutionary consequences of behavioural plasticity

Corresponding Author: Dr Carlos Botero

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
To the authors/editor:

General comments:

The author presents a model examining the evolutionary consequences of behavioral plasticity. While I find it extremely useful to formalize how behavioral plasticity may shape organisms' capacity to adjust to environmental changes across evolutionary timescales, I see scope for improving the present work as outlined in my comments below.

I hope that my comments will help the authors to strengthen their manuscript.

Major comments:
The background provided in the introduction is relatively thin. In my specific comments, I point to potentially relevant literature that the author may consider to incorporate. While the manuscript is generally well-written, the methods and results would benefit from clearer structure and more details (see specific comments). My biggest concern is with the model itself: As outlined in the specific comments, I am not sure whether I agree that this is really a model capturing the impact of behavioral plasticity. Irrespective of this point, the paper lacks a transparent description of the model's assumptions and limitations. I find that both the results and discussion are fairly strongly worded given the limitations of the underlying model.

Specific comments:

- Lines 43 following: Is the scope of this manuscript to only explore behavioral plasticity in response to stress? If so, could this be made more explicit elsewhere? If not, could this example be made more general?
- Lines 56 following: Could the authors explain why a shielding from negative consequences implies weaker selection gradients?
- Lines 59 following: Similarly, could the authors explain in more detail why behavioral plasticity leads to accelerated genetic evolution? Is the explanation just that persistence in suboptimal conditions leads to more chance for evolution (the authors allude to this but the explanation could be more explicit)? Or, is there more to it?
- Lines 78-97: In my opinion this paragraph provides unnecessary signposting. Briefly outlining what the reader can expect in this paper (i.e. a model of X implemented through Y) is valuable. But a detailed outline of the paper, including results, and discussion just blows up the introduction. I also feel that the introduction is quite 'thin' when it comes to discussing other potentially relevant work. I think the author could shorten this paragraph and use the space to add more background to the introduction. Here is some potentially relevant literature (Burton et al., 2022; Fawcett et al., 2013; Mathot et al., 2012; Ratikainen & Kokko, 2019; Stamps & Krishnan, 2017):
-
Burton, T., Ratikainen, I. I., & Einum, S. (2022). Environmental change and the rate of phenotypic plasticity. Global Change Biology, 28(18), 5337–5345. https://doi.org/10.1111/gcb.16291
Fawcett, T. W., Hamblin, S., & Giraldeau, L.-A. (2013). Exposing the behavioral gambit: The evolution of learning and decision rules. Behavioral Ecology, 24(1), 2–11. https://doi.org/10.1093/beheco/ars085
Mathot, K. J., Wright, J., Kempenaers, B., & Dingemanse, N. J. (2012). Adaptive strategies for managing uncertainty may explain personality-related differences in behavioural plasticity. Oikos, 121(7), 1009–1020. https://doi.org/10.1111/j.1600-0706.2012.20339.x
Ratikainen, I. I., & Kokko, H. (2019). The coevolution of lifespan and reversible plasticity. Nature Communications, 10(1), 1–

7. https://doi.org/10.1038/s41467-019-08502-9
Stamps, J. A., & Krishnan, V. V. (2017). Age-dependent changes in behavioural plasticity: Insights from Bayesian models of development. Animal Behaviour, 126, 53–67. https://doi.org/10.1016/j.anbehav.2017.01.013

- Line 94: The mention of advanced cognition comes out of the blue. There is no other mention of it elsewhere in the manuscript. Could the author remove it or integrate it into the manuscript?
- I expected a model which produces patterns of evolution in line with either the cognitive buffer or behavioral drive hypothesis based on basic assumptions fed into the model. Instead the author build a model that intrinsically assumes cognitive buffering. The abstract and end of the introduction could be rewritten to better reflect the actual modeling.
- Line 109: What is the justification for this model? What kind of assumptions does it make about behavioral buffering. How does this equation represent "general capacity to buffer selection through behaviour"? I struggle to relate to the dynamics: When b =1 and behavioral buffering is maximal, fitness will always be 1, irrespective of the phenotype-environment mismatch. How does this compare to behavioral buffering in real organisms? I expected a model in which organisms can adjust b (through trial and error) to change I and that b would be an evolvable parameter. I don't expect a model to mimic reality with all its complexity but I do find it important that the consequences of simplifying model assumptions are made explicit and, to some extent, justified.
- Line 127: Rewrite the sentence to "In this model, the evolutionary dynamics of insulation phenotypes were strongly correlated with the effectiveness of the behavioural buffer (Figure 1)."
- Lines 138-140: This information belongs in the discussion section not in the methods/results
- Lines 145-150: idem
- Lines 127-198: The results section would benefit from headers, which create more structure and make it easier for a reader to follow the type of result that is being described.
- Line 167: The author should explain what mutation-selection balance means.
- Lines 180-183: "The above analyses reveal that although behavioural plasticity can decelerate the evolution of morphology and physiology by reducing selection gradients, it can also accelerate it by increasing diversity in the reservoir of potentially adaptive variants before environments change." I struggle with these kinds of sentences. Is it not directly baked into model that higher b (which the author terms 'behavioral plasticity') reduces the selection gradient d? Regarding the second part of the sentence, I don't quite understand yet at what point and how standing genetic variation is quantified (I could not find it in the R code and the description in the manuscript is insufficient). Does the positive relationship between b and standing genetic variation, at high b levels, not merely stem from the fact that high b values eliminate selection and thus do not constrain the space of adaptive trait values for I? This would mean that this is not really a finding as much as a direct assumption of the model.
- I find the methods hard to follow. I would expect to see a clearer explanation of what happens in each step of the simulation. The information should describe what happens in the main loop of replicate.run in the R script. Some of the information is currently in the manuscript it is just spread out. Once that has been described sufficiently, I would describe the simulation chronology.
- Lines 202-204: "My analyses suggest that behavioural plasticity exerts dual and time-dependent effects on the tempo and mode of adaptive evolution: it weakens selection gradients, and it promotes the accumulation of genetic variation within populations." Based on my current understanding, these are not findings of the model but rather assumptions baked into the model (see above).
- Lines 204 following: "On short evolutionary timescales—or when opportunities for niche expansion are severely constrained—these effects lead to slower morphological and physiological evolution in an altered environment (Fig. 5)." Where in the model did the author investigate the effect of timescales? Figure 5 is just a schematic figure and does not seem to be based on modeling outcomes. The way the text is written suggests otherwise.
- In my opinion, the discussion is too strongly worded given the simplicity of the model. This is corroborated by the fact that the discussion lacks any mention of limitations of the current model.
- In the discussion, I would have found it interesting to read something about the interplay between phenotypic plasticity in morphological and behavioral traits and their combined impact on evolutionary dynamics.
- It would have been nice to also see the R-code that generates the various figures in the paper.

I hope my review will help the authors to strengthen their manuscript.

I sign all of my reviews for accountability and transparency.

With kind regards,
Nicole Walasek

(Remarks on code availability)
The code seemed fine and was appropriately commented. I did not try to run it but merely read through the txt file. I had no concerns about it being able to run when looking through it. However, the code did not contain script to generate the figures from the manuscript which I found unfortunate.

Reviewer #2

(Remarks to the Author)
This paper explores a simple simulation model to understand how buffering of maladaptation affects the rate of evolution

and the survival of subpopulations in stressful environments. I think this is an area of theory in which there is much left to learn and that this scenario would interest a broad range of readers. However, I think that the modeling falls short of what I would want to see in order to recommend this paper, though it certainly could be revised to be an excellent contribution.

The Introduction is excellent and lays out the central question and its motivation concisely and clearly. My only quibble would be that the approach the paper will take is left quite vague in the Intro. This is OK, but I would like to have a little more about the methodological novelty here. Has no one modeled this question, or is there a difference in the approach taken here compared to prior work that can help explain why novel answers were found? It is fine if this topic is reserved for the Discussion (though it isn't much discussed there either), but some foreshadowing at least in the Intro would suit me.

The Results and Methods are presented together, with a later section giving more Methods. I think this is a fine approach in general but confused me at first reading, as the earlier section seemed to have too little detail to be a complete Methods section but also too much to be a concise overview of the model. We aren't told important things like that this is (I think) a model of discrete, non-overlapping generations in either place. I don't need to know the mutation rate (which varies anyway) before understanding basic aspects of the model like how demography works. The crucial information that population size is allowed to vary is mentioned only in the later Methods section, making the survival results in Figure 4 come out of left field for me. This should be communicated earlier, and the constant k included in the the equation on line 109 so that it actually gives absolute fitness. I am still left confused about some aspects like density dependence–is there any? It seems like excess individuals are discarded, but there isn't any effect of density on reproduction otherwise. This seems like a very odd choice as it doesn't permit populations to stabilize at sizes below carrying capacity. I suspect that this is why the prob. of survival plots in Fig. 4 are so steep.

Because of the lack of density dependence (or detail on how it works, if it does), it's hard to understand what role population size might be playing here. This also plays in to the choice of asexual reproduction. In the right context, it can be fine to ignore density dependence and to use this simpler, asexual form of reproduction. But, I'm not convinced that these questions benefit from these choices, because the level of genetic variation seems so critical. I would expect feedback between fitness, as buffered by behavior, and population size to be very important for understanding how buffering affects adaptation. Similarly, I would expect the ability of sexual populations to quickly generate extreme phenotypes to be critical to colonizing new and stressful environments. I therefore don't think that the model is ideally suited to investigating these questions, though it could readily be extended to do so.

I would also suggest that focusing the model on insulation isn't quite the best level of abstraction, as it makes the model read as narrow when it doesn't need to. There's nothing specific to insulation as far as I can see, and certain aspects of the model (symmetrical loss of fitness if one is over- or under-insulated) seem abstract anyway. Similarly, there's nothing about the buffering mechanism which is specific to behavior in my view. This is both good and bad; the paper could apply much more broadly, but also these results are perhaps less original when viewed in the context of general buffering mechanisms.

I would recommend recasting the model in more general, abstract terms, positioning it more thoroughly in terms of prior theory work, and modifying or extending it to allow for more reasonable demographic behavior, and possibly sexual reproduction. With these sorts of changes, I think this would be a very valuable paper.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
To the author / editor:

I really appreciate the respectful and thoughtful responses by the author. I found that they not only addressed my concerns but also were quite interesting to read and engage with.

I have no further comments.

Kind regards,
Nicole Walasek

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This is a very thorough revision that fixed many of my issues and improved on the earlier draft's strengths. The Intro has

gone from good to great in my estimation; I've read a lot of plasticity papers and actually found it quite stimulating in this new version. The model has been revised considerably in actual function and presentation. The addition of a clear model of density dependence is excellent, as is the presence now of recombination. The balance of abstraction and specificity of the model is also now more artfully presented and, to me, works. I really have no further issues, and recommend acceptance.

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

To the authors/editor:

General comments:

The author presents a model examining the evolutionary consequences of behavioral plasticity. While I find it extremely useful to formalize how behavioral plasticity may shape organisms' capacity to adjust to environmental changes across evolutionary timescales, I see scope for improving the present work as outlined in my comments below.

I hope that my comments will help the authors to strengthen their manuscript.

Thank you for your insights. The detailed feedback you provided has indeed helped me identify some key areas that needed more clarity and/or work.

Major comments:
The background provided in the introduction is relatively thin. In my specific comments, I point to potentially relevant literature that the author may consider to incorporate. While the manuscript is generally well-written, the methods and results would benefit from clearer structure and more details (see specific comments).

As detailed below, I have incorporated the suggested citations and taken this opportunity to better connect my work to earlier theoretical papers that are now classics in this field. Following your suggestion (and a similar one by Reviewer 2), I have also aggregated the Methods into a single section in the main text to provide more details and a clearer structure. Similarly, I added passages to the Results and Discussion describing potential limitations and next steps. Details on these changes are provided in my responses to your specific comments and in my responses to Reviewer 2.

My biggest concern is with the model itself: As outlined in the specific comments, I am not sure whether I agree that this is really a model capturing the impact of behavioral plasticity. Irrespective of this point, the paper lacks a transparent description of the model's assumptions and limitations. I find that both the results and discussion are fairly strongly worded given the limitations of the underlying model.

I appreciate the reviewer's request for a better description of the assumptions and limitations of the model. It was clearly needed and my hope is that the changes detailed below have fully addressed that concern. Having said that, I want to emphasize that this is absolutely a model that captures the downstream evolutionary consequences of behavioural plasticity! By simply operationalizing the idea of a behavioural buffer into the fitness function, the model is able to recreate the patterns predicted by both the cognitive and the behavioural drive hypotheses and clarifies the contexts in which acceleration and deceleration of genetic evolution are to be expected.

Specific comments:

- Lines 43 following: Is the scope of this manuscript to only explore behavioral plasticity in response to stress? If so, could this be made more explicit elsewhere? If not, could this example be made more general?

"Stress" is a very general term related to any kind of disruption to the normal state or environment of an organism. Nevertheless, in light of your comment, I have replaced the term "stress" for "altered environments" to avoid any confusion. The example you mention is simply used to illustrate why it has been so difficult to conceptualize and study the effects of behavioural plasticity on evolution: behaviour can be very idiosyncratic even within a single species. The generality of this point is explicitly made in the first sentence of that paragraph (changes in version 2 are presented in bold):

"*A key challenge in understanding the effects of behavioural plasticity on the evolutionary process is that behavioural responses to* ***altered environments*** *vary widely within and among species*"

… and at the end of the same paragraph:

"*Similarly striking qualitative differences in behavioural response are common among species that differ in morphology, habitat, or natural history, further complicating the derivation of general principles from individual case studies*."

Regarding the scope of this manuscript, please refer to:

- Title: "*The evolutionary consequences of behavioral plasticity*"
- Abstract: The goal of this manuscript is to "*examine how behavioral plasticity influences the tempo and mode of adaptive evolution*".
- Introduction: "*Here I integrate theoretical and empirical insights from animal behaviour, cognitive ecology and evolutionary biology to identify conditions under which behavioural plasticity* ***shapes*** *genetic adaptation* ***and influences*** *the evolution of other phenotypic* ***traits****.*"

To help readers better appreciate why this exercise is important, I also modified the last paragraph of the introduction as follows:

"*These findings help explain empirical discrepancies in the observed effects of behavioural plasticity* ***on evolutionary rates*** *and provide a better understanding of how behavioural flexibility* ***influences genetic adaption*** *in the natural world*"

- Lines 56 following: Could the authors explain why a shielding from negative consequences implies weaker selection gradients?

Thank you for letting me know this was not clear. I have reworded the paragraph to more fully introduce readers to the ideas behind the cognitive buffer hypothesis and the behavioural drive hypothesis. The new text clarifies the particular issue you mention as follows:

"*The cognitive buffer hypothesis states that behavioural flexibility shields individuals from the fitness cost of phenotypic mismatch*[20–23] *and therefore predicts a decrease in rates of genetic evolution due to weaker selection*[16]*. For example, the selective use of animal skins allowed early hominins with different levels of insulation to experience similar thermoregulation costs in cold environments, thereby weakening selection on traits like fur and body fat*[24]"

- Lines 59 following: Similarly, could the authors explain in more detail why behavioral plasticity leads to accelerated genetic evolution? Is the explanation just that persistence in suboptimal conditions leads to more chance for evolution (the authors allude to this but the explanation could be more explicit)? Or, is there more to it?

That is certainly the case. Behavioural drives are very well known and have been used to explain rapid evolution in a variety of systems, including high-profile examples like the diversification of the Galapagos Finches. The basic idea is that behavioural plasticity facilitates persistence in suboptimal environments (or the exploration of new ecological niches) and that in doing so, it provides more opportunities for optimizing selection to drive the evolution of new forms and species. There really is not much more to this idea! I have made the following changes to the paragraph to clarify these issues:

"*In contrast, the behavioural drive hypothesis*[15] *notes that behavioural buffering rarely nullifies selection, and therefore makes the opposite prediction that rates of genetic evolution should be faster in more behaviourally plastic lineages because improved abilities to occupy new habitats and ecological niches are likely to provide more frequent opportunities for optimizing selection to act upon the phenotype*[15,18,25–28]*.*"

For further context on this and the previous comment, I also added the following section to the end of that paragraph:

"*The conditions under which behavioural buffering shifts from constraining to accelerating genetic evolution are poorly understood and remain a key unresolved challenge in cognitive ecology.*"

- Lines 78-97: In my opinion this paragraph provides unnecessary signposting. Briefly outlining what the reader can expect in this paper (i.e. a model of X implemented through Y) is valuable. But a detailed outline of the paper, including results, and discussion just blows up the introduction.

Thank you for sharing your perspective. Approaches to scientific writing vary among scholars, and it appears we simply hold different views on this matter.

I also feel that the introduction is quite 'thin' when it comes to discussing other potentially relevant work. I think the author could shorten this paragraph and use the space to add more background to the introduction. Here is some potentially relevant literature (Burton et al., 2022; Fawcett et al., 2013; Mathot et al., 2012; Ratikainen & Kokko, 2019; Stamps & Krishnan, 2017):

Thank you for these suggestions. They are indeed all excellent papers. I had originally chosen to give preference to citations that grounded my model on empirical work (only 70 citations are allowed in this journal) but I realize now that in doing so, I had neglected putting the model in the proper context of existing theoretical work. To address that gap, I now cite the sources you suggest plus classic work from Scheiner, West-Eberhardt, Witlock and others (see responses to R2). One more thing about the sources you suggest… all of them deal primarily with the evolution of plasticity itself rather than with its downstream consequences on the evolution of other aspects of phenotype (i.e., what my paper is really about). This commonality made me realize that I needed to be more explicit about the key difference between my goals and those of other recent models on plasticity (behavioural or otherwise). To that end, and although I cite the suggested sources in multiple sections of the manuscript, I'd like to call your attention to the following passage in particular:

"*Earlier theoretical work on phenotypic plasticity has provided important insights on its evolutionary consequences, but has not typically distinguished between behavioural and more structural forms of plasticity*[28,29,41–44]*. Likewise, the studies that have explicitly made such distinction have focused primarily on the evolution of plasticity itself rather than on its downstream consequences on the evolution of other components of the phenotype*[37,45,46]."

- Line 94: The mention of advanced cognition comes out of the blue. There is no other mention of it elsewhere in the manuscript. Could the author remove it or integrate it into the manuscript?

Apologies for that. I included this term because the cognitive buffer hypothesis is primarily invoked as a possible explanation for the evolution of "advanced cognition". Given space constraints, I chose to simply remove the offending term rather than explaining it.

- I expected a model which produces patterns of evolution in line with either the cognitive buffer or behavioral drive hypothesis based on basic assumptions fed into the model. Instead the author build a model that intrinsically assumes cognitive buffering. The abstract and end of the introduction could be rewritten to better reflect the actual modeling.

Thank you for making this important point. The reviewer correctly notes that behavioural buffering is not an emergent property of my model but rather a built-in mechanism used to operationalize our most recent understanding of how behavioural plasticity influences selection. I must nevertheless point out that while there is broad agreement on what behavioural (or "cognitive") buffering does to selection, there is still considerable disagreement and public debate on what should happen as a result. That is still a **major question** in the field of cognitive ecology and the currently high rate of environmental change has made it relevant to all of ecology and evolution. Details on how I have clarified these issues in the revised version of my manuscript are provided below.

- Line 109: What is the justification for this model? What kind of assumptions does it make about behavioral buffering. How does this equation represent "general capacity to buffer selection through behaviour"?

I explain in the introduction that "*One practical way to circumvent the variability in observed behaviours is to focus our comparisons on the general capacity to buffer selection through behaviour (hereafter 'behavioural buffering') rather than on the specifics of individual response*". The model I propose here simply operationalizes this idea (i.e., the cognitive buffer hypothesis) by explicitly incorporating a buffering capacity into the fitness function. Figure 1A (cited later in the same paragraph) unambiguously shows how varying *b* alters selection gradients. The ability to vary b in this simple formulation is what allows us to investigate later how different buffering levels influence the evolution of a second trait in a variety of scenarios. The relevant text in the new version has been modified as follows to make this clearer:

"*Absolute fitness was determined by the mismatch between an individual's genotype, I, and its current environmental target, E, scaled by a population-wide buffering parameter, b (i.e., the lineage's potential for behavioural plasticity). Specifically,*

$$W_i = e^{d \cdot |E - I_i| \cdot (1-b)},$$

where d determines the steepness of fitness decline with increasing phenotypic mismatch (i.e., lower values = stronger selection). ***In accordance with the cognitive buffer hypothesis, this formulation produces shallower fitness gradients with increasing behavioural plasticity:*** *when b = 0, phenotypic mismatches are strongly penalized,* ***but as*** *b approaches 1, variation in insulation levels becomes progressively less consequential (Fig. 1A).*"

I struggle to relate to the dynamics: When b =1 and behavioral buffering is maximal, fitness will always be 1, irrespective of the phenotype-environment mismatch. How does this compare to behavioral buffering in real organisms?

Imagine an environment with temperatures close to absolute zero, no oxygen, high UV radiation, and no gravity. As far as we know it, no life form on Earth (even tardigrades!) can operate and thrive under those conditions. Yet, somehow, we humans have done it… specifically, we have used our accumulated knowledge, material culture, and collaborative tools, to enable occupancy of low-Earth orbits for over 2 decades and have even set foot on the moon. Our simple phenotypes may be relatively weak but the buffering our brains provides them has allowed our bodies to explore space, the bottom of the ocean, visit the poles, live in the hottest and driest deserts… all in the absence of clear physical adaptations to exploit those environments. In these examples, the "behavioural buffer" is expressed in both physical form (i.e., in space suits/space station, scuba gear, air conditioning, aqueducts, etc.) and behavioural form (know-how, procedures, etc) but the result is close to 100% protection against selective forces that would otherwise kill us. Although it is true that in nature most behavioural buffers are unlikely to be 100% effective (as acknowledged in the passage about crows/ravens in my manuscript), I believe it is still very useful to consider the full range of possible buffering capabilities when developing a general theoretical framework that clarifies our thinking on these issues.

I expected a model in which organisms can adjust b (through trial and error) to change I and that b would be an evolvable parameter. I don't expect a model to mimic reality with all its complexity but I do find it important that the consequences of simplifying model assumptions are made explicit and, to some extent, justified.

Thank you for letting me know what you expected. This is very helpful. Understanding the forces that drive the evolution of behavioural plasticity is indeed a very interesting question. However, the model I present here explores a different yet (I would argue) equally interesting question: how does existing behavioural plasticity shape the (genetic) evolution of other aspects of the phenotype? Since this is the first attempt to explicitly address this question, I believe it is important not to confound cause and effect. Nevertheless, please note that one of the changes I made to this new version of my manuscript is the inclusion of the following statement suggesting that the two approaches could be linked in future work:

"*Future extensions could relax these assumptions by allowing plasticity itself to evolve, incorporating energetic or opportunity costs of flexible behaviour, modelling effects on multiple traits with conflicting selective demands, or embedding the process in spatially structured or temporally autocorrelated environments. Such extensions could further refine our understanding of the limits of behavioural buffering.*"

- Line 127: Rewrite the sentence to "In this model, the evolutionary dynamics of insulation phenotypes were strongly correlated with the effectiveness of the behavioural buffer (Figure 1)."

The results section has been substantially edited to incorporate changes to the model that were suggested by R2 (results and main take-home messages remain unchanged). As a result, the sentence you highlighted here now reads:

"***After environmental changes were implemented in generation 500 (dashed vertical lines in Figs. 1B, 1D, and IF), the*** *evolutionary dynamics of insulation* ***became*** *strongly correlated with the effectiveness of the behavioural buffer*."

- Lines 138-140: This information belongs in the discussion section not in the methods/results
- Lines 145-150: idem

The unique structure of research articles in *Nature Communications* and the intended generality of their audience puts a premium on describing the work as if the author is telling a story. To do that, I have found it very helpful in the past to interpret findings as they are presented and to explain how one finding may have motivated additional downstream analyses. I stand by these writing choices.

- Lines 127-198: The results section would benefit from headers, which create more structure and make it easier for a reader to follow the type of result that is being described.

Thank you for this suggestion. Headers have now been added.

- Line 167: The author should explain what mutation-selection balance means.

This passage is no longer part of the manuscript given the edits that accommodated the changes in model architecture.

- Lines 180-183: "The above analyses reveal that although behavioural plasticity can decelerate the evolution of morphology and physiology by reducing selection gradients, it can also accelerate it by increasing diversity in the reservoir of potentially adaptive variants before environments change." I struggle with these kinds of sentences. Is it not directly baked into model that higher b (which the author terms 'behavioral plasticity') reduces the selection gradient d? Regarding the second part of the sentence, I don't quite understand yet at what point and how standing genetic variation is quantified (I could not find it in the R code and the description in the manuscript is insufficient). Does the positive relationship between b and standing genetic variation, at high b levels, not merely stem from the fact that high b values eliminate selection and thus do not constrain the space of adaptive trait values for I? This would mean that this is not really a finding as much as a direct assumption of the model.

Lots to unpack here! You are correct in noting that behavioural plasticity (b) enters explicitly into the model by scaling the degree to which phenotypic mismatch translates into fitness loss. This operationalization follows directly from the verbal logic of the cognitive buffer hypothesis. In that sense, the buffering effect of b is not an emergent feature of the model, but a deliberate translation of a hypothesis into a formal fitness function. This, by the way, is part of the novelty of my approach.

As for the difference between assumptions and results, it is important to distinguish between *how* plasticity is implemented and *what follows from that implementation over evolutionary time*. While higher *b* necessarily reduces instantaneous selection gradients acting on the focal trait, the downstream evolutionary consequences are not trivially assumed in my model. In particular, the model does not prescribe a priori how changes in selection strength interact with mutation, drift, density dependence, population persistence, and demographic stochasticity across multiple ecological contexts and timescales. Slower short-term evolutionary change, the accumulation of standing genetic variation, and the emergence of a non-monotonic "Goldilocks" relationship between plasticity and long-term evolutionary rates are all **observed outcomes** rather than baked-in assumptions. All of these outcomes are emergent properties, and as I note in the next paragraph, not all of them are immediately obvious.

It is true that very high values of *b* weaken stabilizing selection and thereby permit the accumulation of broader genotypic variance. That mechanism is well understood in theory and is the basis for my prediction for the cognitive buffer hypothesis. What is not known, however, is that such variance will necessarily translate into accelerated adaptation following environmental change. Indeed, my simulations show that excessive buffering can *constrain* subsequent adaptive evolution by weakening selection gradients to the point of no selection. Intermediate buffering, however, enables both persistence under suboptimal conditions and relatively rapid post-change optimizing selection. The key result is therefore not that plasticity "eliminates selection," but that the balance between buffering and selection strength produces qualitatively different evolutionary regimes! That finding may now seem somewhat obvious (thanks to the model itself!) but it has certainly not been part of the academic discussion that has been taking place over the last two decades.

In that context, I'd like to point out that debates over whether phenotypic plasticity (generally defined) accelerates or decelerates evolution have persisted for decades precisely because intuitive verbal arguments can be made in both directions! The value of the present model lies in formalizing an influential verbal hypothesis and demonstrating that it can help us understand why the evolutionary consequences of behavioural flexibility appear to be contingent, scale-dependent, and non-linear. The fact that you now find the mechanisms you mention quite transparent should therefore not be viewed as a weakness of the model, but as evidence that it is doing exactly what it needs to do: provide conceptual clarity about when, why, and over what timescales behavioural plasticity will speed up or slow down genetic evolution!

Re your confusion about how SGV was computed, please consider that in the first version of my manuscript I say that:

"*Standing genetic variation in I also increased with behavioural buffering (Fig. 2).* ***Specifically, within-population variance at generation 500 (s^2_{G500})*** *rose sharply at low b…*"

After incorporating the revisions to the model architecture suggested by Reviewer 2, the timepoint for comparison has been changed to generation 1500 but other details remained unchanged. In the new version, I tried to clarify the definition by restating it as follows:

"*Standing genetic variation, as measured by the within-population variance in I at generation 1500 (s^2_{G1500}),…*"

Finally, please note that the code for computing these variances was already provided in the first version of the supplementary software file, in the statement:

these.results$GenotypicVarianceBaseline <- var(my.pop[,'Genotype'])

… although I understand that this may have not been entirely obvious because the code to produce figure 2 was not shared in my original submission. The new version of this file contains this instruction in line 177, which takes place immediately after finishing “phase 2” (i.e., the 1000 generations allowed for evolution in response to a change in environmental target from 0.5 to 0.3), and the code for Figure 2 now transparently shows where the data are coming from.

- I find the methods hard to follow. I would expect to see a clearer explanation of what happens in each step of the simulation. The information should describe what happens in the main loop of replicate.run in the R script. Some of the information is currently in the manuscript it is just spread out. Once that has been described sufficiently, I would describe the simulation chronology.

Apologies for that. I have revised the manuscript to include a full description of the model, its assumptions and main steps right at the beginning of the Results section. The changes are too extensive to quote here, but I do note that, per your suggestions, a description of every step in the main loop of the simulation code is now available in a single location.

- Lines 202-204: “My analyses suggest that behavioural plasticity exerts dual and time-dependent effects on the tempo and mode of adaptive evolution: it weakens selection gradients, and it promotes the accumulation of genetic variation within populations.” Based on my current understanding, these are not findings of the model but rather assumptions baked into the model (see above).

Please see my response to comment on Lines 180-183 (above).

- Lines 204 following: “On short evolutionary timescales—or when opportunities for niche expansion are severely constrained—these effects lead to slower morphological and physiological evolution in an altered environment (Fig. 5).” Where in the model did the author investigate the effect of timescales? Figure 5 is just a schematic figure and does not seem to be based on modeling outcomes. The way the text is written suggests otherwise.

As described in the main text, the experiments simulating a small diaspora that is exposed to a dramatically different environment intend to approximate the process of allopatric divergence in macro-evolutionary timescales (in contrast with the first section, which was concerned with gradual in situ macro-evolution). The reason this passage was placed in the Discussion is because it is integrating the totality of my findings into a single framework to give us a global, cohesive understanding of the likely effects of behavioural plasticity. I appreciate, however, that I can be more explicit in my distinction between results and interpretation, so in response to the reviewer’s comment I have revised the text as follows:

“*On short evolutionary timescales, or when opportunities for niche expansion are limited, my results suggest that these effects are likely to translate into reduced rates of morphological and physiological evolution (Fig. 5). However, over longer timescales, or when opportunities for ecological diversification abound, they suggest instead that the differential abilities to persist under suboptimal conditions will produce a Goldilocks pattern in which the pace of evolutionary change peaks at intermediate levels of behavioural plasticity (Fig. 5).*”

- In my opinion, the discussion is too strongly worded given the simplicity of the model. This is corroborated by the fact that the discussion lacks any mention of limitations of the current model.

I have included a new paragraph in the Discussion to cover the model's main limitations. The new text reads:

"*The model developed here is intentionally simple, a choice that entails both limitations and advantages. Some of its limitations include the assumption that behavioural plasticity is cost-free and non-evolving, and the examination of evolutionary consequences on a single quantitative trait under stabilizing selection. Future extensions could relax these assumptions by allowing plasticity itself to evolve, incorporating energetic or opportunity costs of flexible behaviour, modelling effects on multiple traits with conflicting selective demands, or embedding the process in spatially structured or temporally autocorrelated environments. Such extensions could further refine our understanding of the limits of behavioural buffering.*"

- In the discussion, I would have found it interesting to read something about the interplay between phenotypic plasticity in morphological and behavioral traits and their combined impact on evolutionary dynamics.

I agree that could be interesting too. However, please note that this suggestion is beyond the scope of this manuscript.

- It would have been nice to also see the R-code that generates the various figures in the paper.

I added R-code to reproduce all figures and reported stats at the end of the supplementary software file.

I hope my review will help the authors to strengthen their manuscript.

I sign all of my reviews for accountability and transparency.

With kind regards,
Nicole Walasek

##
###############
Reviewer #2 (Remarks to the Author):

This paper explores a simple simulation model to understand how buffering of maladaptation affects the rate of evolution and the survival of subpopulations in stressful environments. I think this is an area of theory in which there is much left to learn and that this scenario would interest a broad range of readers. However, I think that the modeling falls short of what I would want to see in order to recommend this paper, though it certainly could be revised to be an excellent contribution.

Thank you for your positive comments and for the detailed and useful suggestions provided below. I have fully addressed your primary concerns about the model by incorporating explicit density dependence via Beverton-Holt dynamics and enabling sexual reproduction. These changes have strengthened the modelling framework without substantially altering the qualitative nature of my findings. Details on the changes made and their consequences are provided below…

The Introduction is excellent and lays out the central question and its motivation concisely and clearly.

Communicating complex ideas to a general audience is often a careful balancing act. Your comment is a highly appreciated counterpoint to Reviewer 1's characterization of the same section as 'thin'.

My only quibble would be that the approach the paper will take is left quite vague in the Intro. This is OK, but I would like to have a little more about the methodological novelty here. Has no one modeled this question, or is there a difference in the approach taken here compared to prior work that can help explain why novel answers were found? It is fine if this topic is reserved for the Discussion (though it isn't much discussed there either), but some foreshadowing at least in the Intro would suit me.

Thank you for pointing this out. I added the following paragraph to highlight the main differences:

"*Earlier theoretical work on phenotypic plasticity has provided important insights on its evolutionary consequences, but has not typically distinguished between behavioural and more structural forms of plasticity*[28,29,41–44]*. Likewise, the studies that have explicitly made such distinction have focused primarily on the evolution of plasticity itself rather than on its downstream consequences on the evolution of other components of the phenotype*[37,45,46]*.*"

As for why novel answers were found… the simple answer is that this model takes a very different approach to modelling plasticity: the operationalization of our current understanding of the explicit mechanism by which behavioural flexibility affects selection.

The Results and Methods are presented together, with a later section giving more Methods. I think this is a fine approach in general but confused me at first reading, as the earlier section seemed to have too little detail to be a complete Methods section but also too much to be a concise overview of the model. We aren't told important things like that this is (I think) a model of discrete, non-overlapping generations in either place. I don't need to know the mutation rate (which varies anyway) before understanding basic aspects of the model like how demography works. The crucial information that population size is allowed to vary is mentioned only in the later Methods section, making the survival results in Figure 4 come out of left field for me. This should be communicated earlier, and the constant k included in the the equation on line 109 so that it actually gives absolute fitness.

Apologies for that. I can see how the earlier structure was confusing, and I appreciate that key details from my approach had not been described. To address these deficiencies, I completely restructured the paper, providing a full description of the methods, assumptions and limitations in the first lines of the Results section. Relevant new text includes:

"***Modelling Framework.*** *To investigate the evolutionary consequences of behavioural plasticity, I developed an individual-based simulation model that tracks the evolution of a quantitative morphological trait, I, in a sexually reproducing population with non-overlapping generations and density-dependent regulation. Selection on I is stabilizing, such that fitness is maximized when trait values match local environmental conditions.*"

Further revisions made in response to this comment are detailed below, including the addition of sections explaining how density dependence and sexual reproduction were incorporated into the model.

I am still left confused about some aspects like density dependence–is there any? It seems like excess individuals are discarded, but there isn't any effect of density on reproduction otherwise. This seems like a very odd choice as it doesn't permit populations to stabilize at sizes below carrying capacity. I suspect that this is why the prob. of survival plots in Fig. 4 are so steep.

You are absolutely right… I can and should do better on that front. The algorithm I used to enforce a carrying capacity in the first version of my model was admittedly not the most elegant implementation of density dependence. In this new version of my model, I have replaced it with Beverton-Holt dynamics. The process is now described in the main text as follows:

"*Population growth was regulated by Beverton–Holt density dependence*[51]*. Specifically, effective fitness, W_i', was computed as*

$$W_i' = \frac{W_i}{(1+\alpha * n)},$$

where n is the current population size and α governs the strength of density-dependence. α was generally tuned by assuming a baseline reproductive output of $R_0 = 2$ offspring and a carrying capacity of N=5000 individuals (i.e., $\alpha = (R_0 - 1)/N = 0.0002$)."

Because of the lack of density dependence (or detail on how it works, if it does), it's hard to understand what role population size might be playing here. This also plays in to the choice of asexual reproduction. In the right context, it can be fine to ignore density dependence and to use this simpler, asexual form of reproduction. But, I'm not convinced that these questions benefit from these choices, because the level of genetic variation seems so critical. I would expect feedback between fitness, as buffered by behavior, and population size to be very important for understanding how buffering affects adaptation. Similarly, I would expect the ability of sexual populations to quickly generate extreme phenotypes to be critical to colonizing new and stressful environments. I therefore don't think that the model is ideally suited to investigating these questions, though it could readily be extended to do so.

Thank you as well for these insightful comments. Your intuition is largely correct. I implemented sexual reproduction through the following algorithm (description taken from the revised manuscript):

"*The algorithm for sexual reproduction began by determining the number of offspring produced by each adult in the population, drawing these numbers from a Poisson distribution with a mean equal to the individual's density-regulated fitness (W_i'). For each offspring, a second parent was randomly drawn from the population, with selection probabilities proportional to effective fitness. Following standard quantitative genetic approaches for polygenic traits*[52,53]*, offspring genotypes were subsequently determined as the mid-parent value plus a small mutational change drawn from a normal distribution with mean zero and a standard deviation of μ=0.05 (unless otherwise stated). This modelling approach captures additive genetic inheritance without explicitly modelling individual loci, approximating with a single mutational change value the cumulative effect of many small and low-probability events across loci.*"

As you predicted, the introduction of sex increases the rate at which new phenotypic variants are introduced to the population and alters evolutionary dynamics. While the main findings remain unchanged, there are some notable differences in the findings from the new approach:

- Standing genetic variation now increases linearly with behavioural buffering (Fig. 2). For reference, this increase was best described by a cubic spline in the first version of the model.

- The rate at which mutations accrue is now much more consequential than in the initial model (see new Fig. 3E). Nevertheless, it is still true that behavioural buffering enables the accumulation of genetic diversity beyond what is possible from mutation alone (see Fig. 3F). It is also now evident that the positive effects of behavioural plasticity on the accumulation of genetic variance are strongest when selection gradients are more intense (see new Fig. 3C). The relevant passages in the new text are:

"As expected, stronger selection (smaller d) accelerated genetic evolution and reduced standing variation (Figs. 3A–B). Similarly, the pace of genetic adaptation continued to be negatively associated with behavioural buffering across d values, but the decline was considerably steeper in simulations with stronger selection (Fig. 3A). Likewise, the positive association between behavioural plasticity and genetic variance persisted across d values, but overall levels of variation decreased when selection was more intense (Fig. 3B). Close inspection of this pattern indicates that strength of selection had a pronounced effect at low b but minimal or no effect at high b. Consequently, the rate at which standing genetic variation rises with b is steeper when selection is stronger (Fig. 3C), underscoring the critical role that behavioural plasticity can play in enabling the accumulation of genetic variants that would otherwise be removed by selection."

… and for mutation:

"*The effects of mutation rate on the evolution of insulation were similarly affected by behavioural buffering (Fig. 3D-F). As expected, higher rates of mutation accumulation generally fuelled faster adaptation, but this effect weakened in lineages with stronger behavioural buffers (Fig. 3D). Similarly, standing genetic variation increased conspicuously with mutation rate (Fig. 3E) and continued to be dependent on b. Here too, the increase in standing genetic variation with b was more pronounced at higher mutation rates (Fig. 3F), suggesting that behavioural plasticity potentiates the effects of mutation by facilitating the accumulation of genetic variants produced during periods of stability."*

- One last difference. As you predicted, transitions from total collapse to long-term survival are now more gradual in my diaspora (macro-evolutionary) simulations (see new Fig. 4). Nevertheless, the general pattern is still largely a stepwise change, indicating that relatively small changes in cognitive ability (or behavioural buffering) could have profound implications for population survival and diversification potential. The relevant text here is:

"*As expected, population survival was more likely when environmental change was less intense (Fig. 4A), selection was weaker (Fig. 4B), or mutations accumulated at a faster rate (Fig. 4C). With the possible exception of Fig. 4B, these effects were visibly weaker than those of behavioural plasticity, further suggesting that behaviour can play a critical role in* *protecting populations against collapse during rapid environmental change. Furthermore, the stepwise shifts observed in Fig. 4 from population collapse to long-term persistence indicate that some relatively small increases in cognitive ability can substantially improve a lineage's capacity for colonization, diversification, and long-term viability."*

I would also suggest that focusing the model on insulation isn't quite the best level of abstraction, as it makes the model read as narrow when it doesn't need to. There's nothing specific to insulation as far as I can see, and certain aspects of the model (symmetrical loss of fitness if one is over- or under-insulated) seem abstract anyway.

Thank you for noticing that the model is broadly framed and that its findings are not limited to the evolution of insulation. It has been my intention from the beginning to make it as general as possible. I am nevertheless of the opinion that having a concrete example can help readers conceptualize the issues and develop their intuition. Yet, your comment made me realize that my initial wording did not convey generality as explicitly as I'd like. To address that gap, I added the following statement the model description:

"*For illustration purposes, I hereby frame the model in terms of thermal insulation: thick coats are favoured in cold habitats and impose costs in warm ones, whereas thin coats exhibit the opposite pattern. However, readers should note that this model is intentionally general and applies to many other traits under selection toward an environmentally defined optimum.*"

Similarly, there's nothing about the buffering mechanism which is specific to behavior in my view. This is both good and bad; the paper could apply much more broadly, but also these results are perhaps less original when viewed in the context of general buffering mechanisms.

This is a great insight, thank you sharing it. I have incorporated it into the manuscript and have suggested some likely buffering agents that could affect evolution in similar ways. The relevant text is:

"*There are also, however, some important conceptual benefits to the model's generality. Given that behavioural plasticity was broadly abstracted as a general buffering agent that helps reduce the fitness costs of environmental mismatch, the model's insights can potentially be extended far beyond behaviour alone. For example, rapid physiological adjustments (e.g., acclimation), human-mediated interventions (such as conservation actions), and even ecological interactions*[44] *(e.g., microbiome shifts or symbioses) can presumably buffer the costs of phenotypic mismatch in similar ways, and could therefore result in similar evolutionary dynamics."*

I would recommend recasting the model in more general, abstract terms, positioning it more thoroughly in terms of prior theory work, and modifying or extending it to allow for more reasonable demographic behavior, and possibly sexual reproduction. With these sorts of changes, I think this would be a very valuable paper.

I want to thank the reviewer once more for their constructive critiques and insightful suggestions. Most of the issues raised in this last summary paragraph have been addressed in my responses to earlier comments so I will focus here on the ones I have not already covered.

Prior theory work. I expanded the introduction to better differentiate my model from earlier theoretical studies on plasticity (both behavioural and broadly defined). I have added both sources suggested by Reviewer 1 and what I consider to be some of the classics in the field. The literature on this topic is vast, and difficult decisions had to be made re what should be included and what was less relevant. Please let me know if I missed something vital.

Model simplification. Given that changes to reproduction and density dependence algorithms involved re-running all simulations, I took the opportunity to further simplify the model. Some of these changes were practical (e.g., I vectorized the code to help it run much faster) but other were structural. Specifically, rather than allowing populations to evolve twice in response to an environmental change, I formally established a 500 generation burn-in period at E = 0.5 (i.e., the mid-value of starting genotypes) to allow populations to establish natural levels of standing genetic variation prior to introducing the first environmental change. This structural change simplified the model description and eliminated a paragraph from the results, giving me some extra space to accommodate the changes requested by reviewers.