COMMENTS TO THE EDITOR

Comment: I have received comments of two highly qualified referees, both of them found the study and question very interesting and acknowledge the elaborate experiments. They both made very detailed and in my opinion helpful comments that would help to improve the manuscript. The referees are sometimes not perfectly clear in understanding the manipulations you did and the resulting conclusions drawn, which might arise from the complexity of the system, the different perturbations you applied, and the indirectly acting forces (acting through changes in competition and population structure) that drive the population response. Separating out plastic or shifts in genotypes is not always conclusive. A cartoonlike figure that introduces the complex life course of the mites, which could include aspects of the manipulations you did might be helpful.

Reply:

Dear Editor, Dear Ulrich,

Thanks for helpful reviews and your own comments. Below we address each of them one-by-one; comments are in black text and our replies in blue. We have added several new figures to explain the life cycle and experimental design, which hopefully clarify our procedures. All in all, we think the revisions improved the manuscript and we hope you agree.

With best wishes,

Isabel & Jacques

COMMENTS TO REVIEWER 1

We would like to thank the reviewer for their comments, which have contributed to clarifying key aspects of the manuscript.

Comment: In this paper the authors present two experiments that have been done on bulb mite populations. In this species, well fed juvenile males are more likely to become "fighter" adults and nutritionally deprived juvenile males usually become smaller "scrambler" adults. In a nutshell, there is a threshold of "juvenile condition" that determine the type of adult morphs the juvenile males will become. We also learn that before becoming adults, some mites (males and females) can transform themselves into a transient specific morph called "deutonymph". Based on previous observations that male deutonymphs always become fighters, the authors explain that deutonymph probably come from individuals of good condition despite being placed in poor environmental conditions. The authors explain that the developmental bifurcation between fighters and scrambler which corresponds to a form of developmental plasticity can either be considered as a form of **mitigating** plasticity or of **anticipating** plasticity, both of which are supposed to be adaptive plasticity; also they never clearly mention that. No other alternative interpretations are considered.
Reply: We have clarified in the first and second paragraph of the Introduction that this manuscript is concerned with adaptive developmental plasticity. Also, we have added a sentence at the start of the second paragraph of the Introduction (Lines 44-49) explaining that different types of adaptive developmental plasticity can be identified and that we focus on two main types: anticipatory and mitigating developmental plasticity (Smallegange et al. 2019; Smallegange 2022).

Comment: The authors claim that they want to study how these two types of plasticity can "fuel the eco-evolutionary population change in response to perturbation", although it is not always what they mean in practice by that.
Reply: We have added further explanation. Specifically, we used the environmental threshold model to create hypotheses on how a change in fitness or a change in population size and structure impacts alternative male phenotype expression evolutionarily (former) or ecologically (latter), and thereby fuel eco-evolutionary population change. We now mention this in the third paragraph of the Introduction (Lines 72-94).

Comment: The authors have individually isolated 420 offspring produced by 3 females sampled from each population and followed them until maturity to record the number of individuals that became deutonymphs, tritonymphs (a new form that has not been mentioned before??) and the sex and morph of adults.
Reply: The tritonymph stage is the final instar stage in the bulb mite's life cycle. We now include a figure with the complete life cycle.

Comment: Before presenting these experiments, the authors make some predictions but unfortunately I found their explanation and the links with the experimental design pretty hard to follow. Here is what I manage to understand:

In the case of mitigating plasticity, I understood that if the idea is to avoid paying unaffordable physiological costs, I would predict the threshold to remain constant despite some changes in the juvenile density, given that the cost paid should depend on the juvenile condition at metamorphosis. But the authors predict that the threshold will increase, using a complex reasoning involving a series of more or less speculative causalities mixing plastic, demographic and evolutionary responses and mixing different time scales not clearly identifiable. For instance, they

argue that harvesting juveniles will lead to an early maturation which I interpret as a long-term evolutionary response. Although it seems pretty ambitious to expect to observe such evolution of the maturation reaction norms on such a short time scale. But beyond that, one could well imagine that the harvesting treatments will change the juvenile density, competition, probably also growth trajectory and condition and maybe also their maturation strategy, but I do not understand why this is supposed to modify the condition threshold if it is determined by physiological cost associated with the production of the fighter phenotype.
Reply: We respectfully disagree that the physiological costs to develop fighter legs depend on condition. We have previously shown that the condition-dependent enlarged fighter leg development in the bulb mite shows food-dependent allometric plasticity (Rhebergen et al. 2022). So, bigger males have bigger fighter legs. The key thing is that, if a male is in poor condition, it will not have the resources to develop fighter legs and will therefore refrain from doing so, but still survive. Indeed, our predictions on how our harvesting treatments affect alternative male phenotype expression are assumed to occur on evolutionary timescales. However, in the bulb mite, such evolutionary changes can occur within 5-10 generations (e.g., Tomkins et al. 2011; Smallegange & Deere 2014). Since their generation time can be as short as 11 days, such evolutionary changes can occur on relatively fast time scales. In fact, in the Methods, we mention that "The harvesting treatments were applied for a total of 217 days, which in a similar experiment was sufficient to observe an evolutionary shift in fighter expression (Smallegange & Deere 2014). (Lines 278-280)". But in the case that if the experiment was too short for a strong evolutionary response to occur, we can still distinguish between our anticipatory and mitigating hypotheses, as we outline in Lines 281-293.

Using the example of the selective harvesting of juveniles: this lower the fitness functions of fighters if we assume that the developmental plasticity is mitigating (because, all else being equal, fighters have longer development times than scramblers and thus suffer a higher mortality probability) (Figure 1). Using the ET model (see previous comments), which we now explain in the third paragraph of the Introduction (Lines 72-94), we can then predict how this change in fitness function evolves the threshold for alternative male phenotype expression (indeed, like you say: the maturation strategy). We hope that this extra explanation clarifies our predictions. Indeed, removing juveniles can affect other population processes (reducing competition), potentially slowing the evolutionary response to increased mortality rates.

Finally, we think some of the confusion is because of an assumption that the condition threshold is determined by physiological cost associated with the production of the fighter phenotype. This is not the case. We now clarify that we use the ET model, in which the threshold depends on condition. Please see the third paragraph of the Introduction (Lines 72-94) for a complete explanation.

Comment: In the case of anticipating plasticity, although it is not explained in these words, I understood that harvesting the deutonymphs which are supposed to become fighters will give a signal to the other nymphs that it may be more advantageous to become a fighter event if you have relatively low condition given that the competition between fighters will be less harsh. Thus the condition threshold could plastically vary and become smaller, which is not expected to be the case when harvesting concerns the other types of juveniles (also one could also say that the proportion of fighter males will increase if only the deutonymphs survive and thus it could be more advantageous to become a scrambler even if you are in relatively good conditions).
Reply: Firstly, removing deutonymphs will change the structure of the population because we are removing individuals of good condition, and thus we are removing males that would most likely

have become fighters. This is an ecological response, which we clarified now in the fifth paragraph of the Introduction: lines 118-124. Alternative male phenotype expression in this species is not frequency dependent so more fighters in a population does not elicit an increase in scrambler expression (Deere & Smallegange 2014). We have added this to lines 103-104.

Comment: The study these two alternative hypotheses relies on the identification of the "condition threshold". But how "juvenile condition" is measured is not clearly defined: is it body size, growth rate, corpulence? Condition seems to be equivalent to body size given that the author measure the size of the different morphs but then how can one tell if two juveniles have different conditions given that their size is susceptible to change as they grow. Moreover, the authors seem to quietly abandon the idea of studying the sliding of this "body condition threshold" and seem to replace this idea by the measurement of "fighter expression" expressed sometime as "proportion of fighters". But this is very different from the above-mentioned condition threshold, given that for instance higher growth rate could well enable more juvenile to cross the threshold and metamorphose into a fighter without any modification of the threshold itself.
Reply: We define condition as "the resource budget available for the production and maintenance of adaptive traits" (start of the Introduction). Indeed, we here take body size as a proxy of condition, following our previous work where we measured a strong link between juvenile male body size and alternative male phenotype expression (Rhebergen et al. 2022). We now clearly state that, like others, we approximate condition by body size (Lines 258-260 [Methods]). Furthermore, in none of our experiments have we observed compensatory growth after nutritional deprivation in the bulb mite (Leigh & Smallegange 2014: Exp Appl Acarol (2014) 64:159–170. DOI 10.1007/s10493-014-9822-y).

The ET model states that any change in the threshold for alternative male phenotype expression will affect the proportion of individuals developing either phenotype because it is expected to track the intersection of the alternative phenotype fitness functions. This is why we can deduce evolutionary shifts in the threshold from evolutionary changes in fighter expression (the proportion of adult males that are fighters). We have added this information to the Methods (Lines 288-293).

It was not possible to directly measure the threshold itself as logistically we did not have enough people in the lab to conduct these time-consuming and labour-intensive assays (see e.g. Smallegange & Deere 2014). We now mention this in Lines 298-300.

Comment: Rather than displaying the mean scrambler and fighter size on different and differently formatted graphics which prevent any comparison (4B, 4C; 5A with a missing figure showing the fighter size in the second experiment), the author could rather show the distribution of the fighter and scrambler size and study how their joint distribution, which should reveal in its centre the famous threshold, is modified by their treatments.
Reply: We only measured the sizes of up to 5 individual deutonymphs, fighters, scramblers, and females at a census moment, and these numbers are not high enough to fit a distribution. Instead, we describe the distribution with its first moment, the mean, which we state in the Methods (Lines 259-260).

We made the y-axes of fighters and scramblers the same within each Figure but not for females or deutonymphs because they were significantly smaller and then the differences would hamper the visual difference over the time periods we wanted to highlight; setting the y-axes the same for these graphs would make this visual difference less. Note that these are now Figures 6 and 7, respectively.

Unfortunately, the threshold for male morph expression is measured in final instars, immediately prior to maturation, and not by comparing adult scrambler and fighter sizes. Because the assays to measure the threshold are time-consuming and labour-intensive (see e.g. Smallegange & Deere 2014), we were unable to do them. We now mention this in Lines 298-300. In fact, we did try, but were not able to follow the fate of enough individually isolated mites so could not measure the threshold.

We have added female size and fighter size to Fig. 5 (now Fig. 7).

Comment: It is also difficult to understand the rationale under the different harvesting treatments. More information is required on the population size and structure of the control populations and how it is affected by the harvesting treatments. How many juveniles, how many deutonymphs, how many male and female adults? How does it vary with time, etc.
Reply: Analysing treatment effects on the number of juveniles and male/female adults was not part of testing our hypotheses. However, we have included their analysis in the Online Appendix on Figshare (https://doi.org/10.6084/m9.figshare.22012589.v1) (Lines 329-331).

The number of deutonymphs across all treatments and time are in Fig. 3D (now5D).

Comment: Why did the authors harvested the same % of deutonymphs and juveniles rather than the same number? I guess that the deutonymphs represent only a small part of the juveniles and thus the demographic effect of harvesting 100% deutonymphs versus 100% juveniles are not at all comparable. Thus, how to rule out the idea that the effect of the treatment is simply a demographic effect rather than a selection of individuals according to their condition or their potential future?
Reply: Precisely to figure out if our observed responses were due to demographic or evolutionary effects, following the end of the long-term population experiment, we conducted a common garden life history assay to assess if any differences in body size and fighter expression between treatments were plastic or genetic. It turned out that we did not have enough statistical power and we discuss these results in the first paragraph of the Discussion.

Populations of different sizes meant that harvesting the same number of individuals across populations would disproportionally affect some populations if their total number were lower than another. Using the same percentage, relative to the population in question, avoided this problem and ensured an equivalent comparative removal across populations (i.e., treatment vs control).

Comment: I found the paper quite dense, and at some moments pretty difficult to understand especially for someone not familiar with the very interesting but somehow quite complex biological system. I think that an effort should be made to try to better explain what has been done. Understanding the links between the general framework, the theoretical predictions, the biological system and the experiment design is not easy. I thought that it would help to have detailed presentation of the biological system before presenting the general framework, or at least to better explain the links between the two.
For instance, there is a cognitive leap from lines 67 to 68.

Reply: In response to this and, your previous comments, we have clarified what theoretical model we use to derive our hypotheses (the environmental threshold model) and added text to further clarify our hypotheses. We have also added a figure with the complex life cycle of the bulb mite (Fig. 2). We have also added text in between lines 67 and 68 to explain the rationale to create predictions on how changes to the fitness functions and condition distribution of anticipatory or mitigating developmental plasticity fuel eco-evolutionary population change (third paragraph of the Introduction: Lines 72-94).

Comments on the figures

Figure 1

Comment: Please explain how fitness curves have been measured. It could also help to provide the empirical data on which this figure is based on. The "fitness" of each type of male depends on its body condition (body size?) prior to metamorphosis, but shouldn't it also depend on the adult density and especially on the relative proportion of the different type of males which is not directly linked to the "condition" axis. Thus, it is a bit difficult to understand what the fitness lines really represent. Or it may be useful to explain that these fitness crossing lines represent the fitness of each type of morph measured in a similar specific fixed environmental condition.
Reply: We apologise for the confusion. The fitness functions in this figure are theoretical predictions derived from the environmental threshold model. We have now clarified this in the Introduction (third paragraph: Lines 72-94) and Figure legend. We assume that the fitness of the male mites does not depend on the relative proportion of males of each morph in the population because alternative male phenotype expression in this species is not frequency dependent so more fighters in a population does not elicit an increase in scrambler expression (Deere & Smallegange 2014). We have added this to lines 103-105.

Comment: How it is possible to know the fitness of a fighter of weak condition given that they do not exist. -> you could distinguish in the graph the fitness functions that have been measured from those that are extrapolations (dotted lines).
Reply: Indeed, it is experimentally very difficult to measure these fitness functions and we have not yet succeeded to do so for the bulb mite. As we mention elsewhere too, the fitness functions in this figure are theoretical predictions derived from the environmental threshold model. We have now clarified this in the Introduction (third paragraph: Lines 72-94) and Figure legend of Figure 1.

Comment: Condition is not clearly defined. Condition of what? If it is body size, then this should be explained. It is not clear what the distribution of condition is in the population. If the condition=size, then is this distribution equivalent to the population size structure that includes all the different stages/ages.
Reply: We define condition in the first sentence of the Introduction: the resource budget available for the production and maintenance of adaptive traits (Hill 2011; Casasa et al. 2020; Nijhout & McKenna 2018). We now clearly state that, like others, we approximate condition by body size (Lines 49-52 [Introduction] and Line 259 [Methods]). Indeed, you are correct that the condition distribution is equivalent to population size-structure. We clarify this now in the third paragraph of the Introduction (Lines 72-94).

Comment: 515 -> "deutonymph harvesting will reduce the size of the condition distribution" I am not sure to understand this sentence. Do you mean reduce the population size? I found it difficult to understand what the blue and red lines represent. Is it the size distribution in the populations, including all the different individuals? If this is the case, I do not understand how the population size (grey area), can increase after a harvesting event (D, F), especially after juvenile harvesting given that deutonymphs seem to be quite rare among the other juveniles. Could you explain what you mean by "after" compared to "before". As it is, "after" can be understood as after harvesting but this is probably not the case. But it is probably a "long-term after" especially if you expect some demographic and evolutionary response. It may help to clearly distinguish the two time-scales: how the population will look like just after harvesting and what could be the long-term demographic responses.

Reply: Indeed, the size of the condition distribution is equivalent to population size. We have added this to the Figure legend and now also explain this better in the Introduction (third paragraph: Lines 72-94). Regarding why population size can increase after harvesting, as it says in the Figure legend, because scramblers mature earlier in life (Smallegange 2011), live longer than fighters (Radwan & Bogacz 2000) and sire more offspring than fighters (van den Beuken et al. 2019), an increase in scrambler expression can increase population size (Smallegange & Deere 2014), and thus increase the size of the condition distribution (d,f).

Finally, regarding the time scales, as we've explained in our responses to your previous comments, evolutionary responses in the threshold for male morph expression can occur within only a few generations, so are not very long-term. We have added this to the Figure legend and in the Introduction.

Comment: Then it could be nice to compare these predicted changes in the size distribution to the observed size distributions in your population at different time scales. I must admit that it is difficult to follow the reasoning given that it is quite complex and relies on mulpliple causalities. Juvenile harvesting -> males will mature earlier (through a plastic or evolutionary response?) -> more scramblers (in proportion or total number?) -> "this will increase scrambler fitness" (I do not understand why, and fitness compared to what?) -> evolution of a higher threshold (not clear why).

Reply: Indeed, we conducted this comparison by analysing population size, which is the total area under the size distribution (see section 'Population size…' in the Methods and Results). About the causal reasoning: we have clarified this better in the Introduction (Lines 172-190): according to the ET model and assuming that the male polyphenism is mitigating, the stress mechanism of the developmental system can fuel a mitigating response to increased deutonymph and juvenile mortality because males can mature early as a scrambler to escape the juvenile stage quickly (Ernande et al. 2004), fuelling the evolution of scrambler expression. Thus, if scrambler fitness increases relative to that of fighters, the threshold for fighter expression will evolve to decrease, both in response to deutonymph harvesting and in response to the selective harvesting of other juveniles.

Comment: Then it is written that the scramblers mature earlier, live longer and produce more offspring than fighters (in which conditions?), and thus the population size will increase. But if scrambler are so performant, why do fighters exist? And what about the density dependence? Is the link between these performant scramblers and a change in population size empirically verified?

Reply: Like we said before, there is no effect of male morph frequency (Deere & Smallegange 2014) or of density (Radwan 1995), which we have added to the Introduction (Lines 103-105). It was our prediction that scrambler/fighter expression affects population size, and we find evidence for this in this manuscript (see Results and Discussion).

Comment: In a nutshell I found the reasoning difficult to follow, probably because this reasoning is based on many causal relationships, different time scales (just after harvesting, a fter demographic response, after possible genetic evolution) and responses of different natures (evolution and plasticity) in a system where there is necessarily some complex demographic feedback, and it is difficult to understand how they can act (a figure could help present how the different possible causalities intertwine).
Reply: We apologise for all the confusion. I hope our revisions and extra explanation of how we derived our hypotheses from the environmental threshold model, as well as further references to the bulb mite empirical system and a figure with the mite life cycle (Fig. 2) has increased clarity.

Comment: I think that it would help to represent the different times scales, and to represent similar data derived from the experiments (for instance the size distribution in the populations and their change after harvesting and during the experiment).
Reply: As we now hopefully explain better (please see also our replies above): the demographic and evolutionary time scales of changes in the threshold for male morph expression and population size and structure are almost concurrent. This is also the definition of eco-evolutionary change and this is what makes the bulb mite system an excellent system to study eco-evolutionary population change. Please also see our replies above regarding our analysis of the treatment effects on population size.

Comment: Given that this figure is supposed to help make clear predictions, it would be useful to add some predictions that are directly linked to the traits that will be analysed (body size rather than the threshold).

- ⇨ In case of mitigating plasticity, we expect to observe an increase in total population size after harvesting compared to controls, a smaller number but bigger fighters after both deutonymph and juvenile harvesting.
- ⇨ In case of anticipatory plasticity, we predict to observe a smaller population size in harvested populations compared to controls, the same proportion of fighters compared to controls but smaller fighters only in the deutonymph harvesting treatments.

The figures of the results should be shaped to present the data in a way to facilitate the comparison with the above-mentioned predictions which is not the case right now.
Reply: We now in the Introduction (third paragraph: Lines 72-94) explain better how our hypotheses are derived from the environmental threshold model and what the main response variables are and why (proportion of males that are fighters and population size). Like we mentioned earlier, for logistical reasons, we could not run enough life history assays to measure the threshold for male morph expression and thus used the proportion of males that are fighters as the response variable. The ET model does not give predictions on whether fighters (or scramblers) on average increase (or decrease) in size.

Figure 2.
Comment: Given that the number of populations is rather limited, it would be nice to show the population dynamic of each population and to add some information on the size distribution or of the cumulative number of the different groups of individuals. Cumulative bar plots could be used for instance to show the temporal changes of the population size and structure. You could also add an arrow each time harvesting takes place. And it would also be nice to show the different periods on the graphs.
Reply: Please see our reply to your comments above about not being able to show the full size distribution. We have added grey lines in Fig. 2 (now Fig. 4) to denote time periods and clearly

show week numbers. We have added cumulative plots of the number of individuals in each life stage to the Online Appendix (Figure S1).

Comment: As it is now, the figure 2 does not show the between-population variability of total population size or of the population structure, both demographic traits being important to grasp to understand what is happening.
Reply: Please see our reply to your previous comment.

Comment: Please also explain why the first part of the population dynamics are missing in for the JD100 and J-D50 treatments.
Reply: As we mention in the Methods (Lines 233-235), the juvenile harvesting populations were started 8 weeks after the other treatments, but before harvesting commenced (Fig. 3), for logistical reasons. At the point where these populations matched the dynamics of the other treatments we started harvesting.

Comment: It is a good idea to use different colors for the different treatments but it would be nice to use the same colors in all the figures.
Reply: Because we merge some treatment levels in the analyses, we cannot use the same colours used in the time series Figure in the other Figures. Also, we prefer to use black, white and grey lines with different types of dashes so colour-blind people understand the figures too. Vice versa, we cannot impose the same grey/black colouring and shading to the time series Figure because we merged treatment levels.

Figures 3 and 4.
Comment: I found it difficult to follow what is going on because the way the results are presented changes from one panel to another. If I do understand the rationale behind the idea of grouping together some treatments, depending on the result of the statistical contrasts, I do not think that it is a good idea to present the data as it is done on the figure because 1) it requires too much brain effort to understand what is going on each time one switches from one panel to another and 2) it is not possible anymore to directly compare the results to your predictions. For instance, we need to be able to compare the different treatments to the control treatment which is now grouped together with other harvesting treatments.
Thus I suggest representing the result of the models with the interaction and thus to have an independent estimate for each of the 5 treatments and to use the same colors as the one used in figure 2. Figure 3 A and B could be fused together.
Reply: We appreciate the point you make but in this way, the results in the figure match the results of the statistical analyses. One can still compare the results with our predictions, and we do so in the Discussion. If a control treatment is grouped with another treatment, this means that they are not statistically different. However, the control treatment then does differ from the other treatments shown in the panel. We have now added to the Figure legends that treatment levels that did not significantly differ from each other are grouped.

For your convenience, however, we have added to the online appendix the interaction plots for all panels in Figures 3 and 4 (now Figures 6 and 7).

We disagree that Fig. 3A and B can be fused together because they have different x-axes.

Comment: Moreover, to better represent the real amplitude of the between period and between treatment's variations, I think that it would be more relevant to scale the y axis of the graphs from

zero. Plotting the 95%Ci rather than se could also help to easily see which are the treatments that significantly differ and this could help draw lines between the graphs and the results of the statistical analysis.
Reply: We appreciate your point of starting the y-axis at 0, but when the highest number is 500 in case of population size, this means that the averages get bunched up at the top and you won't be able to see them very well or distinguish between the treatment levels. We do, however, now mention in the Figure legend that the y-axis does not start at zero.
Furthermore, as we explained above, the results in the figure match the results of the statistical analyses and we therefore respectfully refrain from plotting all means of all treatment levels with 95%Cis as, in our opinion, this gives rather busy panels. We have added to the Figure legends that treatment levels that did not significantly differ from each other are grouped.

Other comments
I allow myself to write down some other comments that came to me while reading your manuscript in case it could be useful to you.

Title
Comment: I am not sure that the "leave a male polyphenic signature" in the title is clear enough. I would rather recommend a title like "The short- and long-term effects of selective juvenile harvesting on the demography and developmental trajectory in the male polyphenic bulb mite"..
Reply: Because we did not empirically show that there was a long-term evolutionary effect of our treatments on male morph expression, we'd like to not state that in the title, as it suggests that we show these effects. Furthermore, with this title we'd like to emphasize the role of development, and not so much the harvesting per se.

Abstract & Introduction
Comment: I think that the abstract and introduction could be improved to clarify things. Use shorter sentences when possible. It would help to use more empirical examples to illustrate certain theoretical concepts and ideas that are not always easy to understand. Although some efforts have been done to explain the hypothesis and prediction, notably using the figure 1, I had difficulties understanding the presented predictions (see comments on figure 1). It is difficult to understand what represents and on what are based the fitness curves. The condition distribution is also not clear. Is it equivalent to size distribution within a population?
Reply: Following your comments we have added further explanation in the Introduction. Specifically, we used the environmental threshold model to create hypotheses on how a change in fitness or a change in population size and structure impacts alternative male phenotype expression evolutionarily (former) or ecologically (latter), and thereby fuel eco-evolutionary population change. We now mention this in the third paragraph of the Introduction (Lines 72-94). As already mentioned before and now also explicitly highlighted in the Introduction (third paragraph), the condition distribution equals the size distribution. We have also shortened long sentences in the abstract and aimed to clarify the text better.

- L2-3 First sentence long and complex. -> Shorten / split?
  Done
- L6 Do you consider that there are only these two types of plasticity?
  We have rephrased this for clarity
- What do you mean by "underlie different ecoevolutionary population dynamics"?
  This is now explained in more detail

- L7-10 the transition to your system, to the selective harvesting and male polyphenism without explanation is a bit abrupt. -> give more information on the system and the question that arise before presenting the experiment.
  We have refrained from doing so to keep word count manageable
- 30 "large major *individuals"*
  These males are referred to as majors or minors in the field of sexual selection
- 33 if the difference are typical, tell us what the difference are.
  Added: because they follow different developmental trajectories.
- 34 "Difference in individual development…"-> what do you mean by "difference"? Do you already refer to plastic of genetic modification of the condition threshold? Please clarify what will affect the population structure size and growth.
  We have amended the sentence to reflect this.
- 35-38 Could you illustrate with some empirical examples?
  The references provided give examples. It's beyond the scope of this study to discuss those here.
- 39 It could help to present what you mean by "the perturbation of developmental trajectories" upstream.
  Text has been removed in revision.
- 41 "perturbation *of* individual development"?
  Text has been removed in revision.
- 42 could you illustrate this idea with an example?
  That is what follows: the minors and majors.
- 46 It could help to present the two types of developmental plasticity using some examples before remobilizing these concepts here. Or maybe first explain that you consider that DP can be either anticipatory of mitigating and then present the two alternatives.
  Done.
- 47 Mechanism *is* the evolutionary result -> *can be interpreted as?*
  Done.
- 54-56 I do not understand the difference with anticipating plasticity, which also constructs phenotypes that are successful at reproducing. Please clarify the difference between the two types of plasticity. You could illustrate each type with an example.
  We summarise the differences at the end of the second paragraph of the Introduction.
- 63 Please clarify. I do not understand what you mean by "a plasticity that is regulated by the dynamic of population density…" If the plasticity is the existence of a condition threshold, you mean that this threshold is itself plastic and will vary depending on the population dynamic and level food competition? I would have rather considered that these demographic factors will affect the conditions of the individuals which will determine their position relative to the condition threshold which itself remains constant. As it is written right now, it is not clear what you mean to explain.
  We have clarified the sentence to indicate that the threshold for alternative male phenotype expression is regulated, as you indeed concluded.
- 64 Why do you write that the condition distribution informs on population size? I do not understand the link between the two.
  We have added text to explain this. Please see above.
- 65-67 This is not clear for me. What do you mean by "fuel the ecoevolutionary population change". Can you be more specific. I am afraid that this formulation can have several interpretations.
  We have added text to explain this. Please see above.
- 67-68 large cognitive leap. It may help to present the biological system in more details earlier.

We have added text to explain this. Please see above.

- 69 “impact eco-evolutionary population responses” -> as above, please explain what you mean by providing for example what are the precise demographic traits that you ambition to measure and on which timescale.
  We have added text to explain this. Please see above.
- 73 “good condition juveniles” -> given that this is central to your experiment and reasoning it could be useful to better explain what you mean by that. How do you know that a juvenile in your population has a large resource budget? Body size or corpulence could be a proxy but I guess that body size increase with juvenile age, thus how can you tell if a small juvenile is a nutrition-deprived individual or a well fed but just young individual?
  - ➔ it may help to have a figure showing the lifecycle of your mite and the different stages, etc.

  We take body size as a proxy for condition, which we added to the Methods (second paragraph under ‘Experimental procedure’. We have added a life cycle figure (Fig. 2).
- 74 The growth rate could indeed be a proxy of the juvenile condition given its age, but can you measure that in your populations/experiments? But the link between growth rate and condition at metamorphosis is not straightforward given that a condition deprived individual could reach metamorphosis with the same condition as a well-fed individual. It could just take longer to reach this stage.
  We here refer to the paper by Rhebergen et al. (2022) who showed that, as it says in the text, fighter males metamorphose from good-condition juveniles that have a large resource budget.
  74 nutrition-deprived juveniles that grow slow or are of worse condition when metamorphosing into an
  Unfortunately, it is not clear from the comment what we are required to address
  75 dult instead express the scrambler phenotype that does not have such leg modifications
  Unfortunately, it is not clear from the comment what we are required to address
- 76 Could you cite between brackets what could be these unfavourable environmental conditions (higher density?).?
  Done (low food)
- 81 “carries development costs…” -> it is maybe more cautious to write something like “linked to/associated with a reduced size…”, given that formally demonstrating that it is a “development cost” may be a bit challenging.
  We showed that deutonymph expression has costs, which are reduced egg production and reduced size at metamorphosis (Smallegange & Deere 2015).
- 83-87. May be first explain that Deutonymphs always mature as fighters and then give your interpretation.
  Thank you for the suggestion, but we disagree as that changes the causal reasoning.
- 91 remove “thus”?
  We have kept the wording as removing it would change the causal reasoning.
- 94 what do you mean by “juvenile performance”?
  The survival and growth of juveniles
- 94-100 difficult to follow. You could clarify when you consider that the threshold will “evolutionary change” or when it may be plastically modified depending on the environmental conditions.
  We have amended this for clarity; see previous comments
- 105 Could you try to rephrase this more simply?
  Tried to do so.
- 107 You refer here to the genetic evolution of the maturation threshold?
  minor -> scambler?

We refer to the mechanism of male morph expression where, in our system, majors are fighters and minors are scramblers.

- 108 "the *evolution* of scrambler expression" -> could you consider using another word to avoid confusion with genetic evolution?
  Changed. It is the threshold that evolves and the threshold is a polygenic trait (added to the Introduction).
- 108 "Mean fighter expression": please explain what you mean. How do you measure it?Is it the proportion of juveniles that become fighter, the total number?
  Proportion of males that are fighters. Clarified in the Introduction.
- 109 reasoning difficult to follow given that you could also imagine that harvesting-> less juveniles -> less competition -> more resources-> higher growth rates and increased juvenile conditions -> more fighters.
  Tried to clarify
- 113 Do the scrambler mature earlier or at the smaller size?
  Earlier, but not necessarily at a smaller size
- 113-114 It is written as if it was true in all environmental conditions. But I imagine that the scrambler's performance is regulated by the density, the proportion of fighters, etc. Otherwise, why do fighters exist if the scramblers mature earlier, live longer and produce more offspring.
  Please see our replies above
  116 ((-> (
  Fixed
- In a nutshell I found your reasoning difficult to follow given the intertwining of time scales and potential demographic feedback loops.
  We hope the revisions have increased clarity.

Methods.

- More details are required to better understand the experimental design (a figure could help).
  We have added a figure with the experimental design (Fig. 3).
- It is not clear why the authors choose to analyse time as discrete three periods rather than as a continuous covariable (using gam for instance).
  Because we are interested in long-term (instead of transient) temporal changes in the response variables over time within each treatment group, as is stated in the Methods.
- Why did you choose to harvest the same percentage rather than the same number of juveniles versus deutonymph to keep the demographic impact constant?
  Populations of different sizes meant that harvesting the same number of individuals across populations would disproportionally affect some populations if their total number were lower than another. Using the same percentage, relative to the population in question, avoided this problem and ensured an equivalent comparative removal across populations (i.e., treatment vs control).
- 124 please explain what the tritonymph are and give some information on the duration of these stages (and the other stages also… through a figure for instance) and the mean proportion in your populations.
  Figure added (Fig. 2)
- 124 an ->and
  Fixed
- 126 unfavourable conditions -> also high density?
  Only if high density means high competition and low food
- 131 -> do you expect to have a lot a genetic variability in your populations?

Added to the Methods: Individuals across populations are highly inbred with the exception of scrambler males in stressed environments: see Stewart et al. 2019 doi: 10.1186/s12862-019-1385-4

- 136 Please indicate here that harvesting has been done on a weekly basis (the information is otherwise provided on line 167).
  Done
- 150 What do you mean by "from the source population with founder mites"?
  Clarified: Each population was established with 50 randomly selected mites from the source population, which we refer to as founder mites.
- 156. explain what the time periods are before mentioning them. Why did you choose to split the experiment into three time periods? Why not just show the time trajectories of the measured traits to show how they continuously vary with time?
  Please see our previous comments
- 157 "up to five" -> how many on average?
  On average: 4.05 $\pm$ 1.46 SD
  161-163 could you illustrate these measurements with some pictures?
  Yes, these are now in the new Figure 2
- 165 "dynamics to stabilize" -> What do you mean? Dynamics of what? Do you mean population size and age/size/morph structure?
  Changed to population numbers
- 172-173 Isn't it the condition threshold that is supposed to remain stable rather than "fighter expression" which I read as the number of fighters produced, which itself could increase if there is a higher juvenile growth rate and juvenile condition due to a lower number of juveniles due to the harvesting?
  Fighter expression is the proportion of adult males that are fighters and this is determined by where threshold is located within the condition distribution. This is now better clarified in the Introduction.
- 175-177 Please clarify why you expect such evolution given that the two types of plasticity are supposed to be adaptative. Plasticity may be sufficient?
  Because our harvesting treatments change the selection pressures. Clarified in the Introduction.
- 188-189 The effect on which traits?
  Size, age and morph at maturity. Added to the text.
- 200 Please justify the three periods. Why three, why the limits, why not consider time as a continuous variable?
  Please see our replies above.
- 205 Did you use Poisson models? Did you include the number of females as a covariable?
- Why not analyse the fighter expression as a proportion using a binomial model (number of fighter males versus scramblers)?
  As it says in the Methods, we used Poisson models to analyse the number of deutonymphs and on total population size. Using an offset term, as we did, is preferred over using binomial distributions.
- 207 You analyse the size of the different types of individuals and the number of deutonymphs. What are your predictions on these traits as a function of the different treatments and different types of plasticity?
  These are stated in the Introduction and discussed in the Discussion
- 211 In the analysis of number of deutonymphs, did you add the total density as a covariable, given that in the introduction it is explained that more deutonymphs are expected when the conditions are harsh (higher density, higher competition?).

We never stated that deutonymph expression is sensitive to population density. It is sensitive to food availability and we kept that constant.

- 215 In the life history assay why didn't you measure the age at maturity rather than mean adult size given that one of your predictions was that this trait would evolve?
  Not enough person power to monitor the mites daily in all life history assays.

Results.

- The models and their simplifications should be grouped together into tables to clearly present what has been done.
  These model simplifications are fairly straightforward and adding another Table makes this a very long manuscript.
- 235 Given that the interaction is marginally significant, it would be nice to have an unconstrained view (in the figure) of what is going on in your data, showing the prediction of the models with the interaction. The interaction may become significant if you pool together some treatments as you have done later.
  Please see previous comments
- 235 When you present some results could you also give the magnitude of the effects: "Population size was *on average 3%* highest when…"
  The main point is not how high these population sizes are but under which experimental treatments they are high(er) or low(er). We believe that the graphs are representative enough to highlight this as they show the actual numbers.
- 139)) ->)
  This is correct
- 240 need a test for the difference between periods. Given that you choose to split the time period into three periods rather than analyzing time as a continuous covariable, you should maybe say "population size different significantly among periods" rather than "increased over time".
  Changed.
- 242 verify the chisq symbols.
  They are correct
- 246 What is this test? Given that there is a significant interaction between periods and harvesting treatments, shouldn't you have to do this test independently for each period?
  We disagree as that would increase the type I error of finding a significant result when there is none.
- To better understand the statistical analysis you could also refer to a table where you report all the different statistical models that you have done.
  All the relevant statistics are in the Results section. As mentioned above adding an additional Table would make this a very long manuscript.
- 246 You say that "the proportion of fighters was always low when 100% of deutonymphs were harvested" but this is not the case because it is always higher than 80% which is not low... You could maybe rephrase with something like "on average, the proportion for fighters varied between ~85% and 93%, and we found that during the first two periods, it was slightly but significantly lower (-5%) in the D100 treatment (~85%) compared to the other treatments (~90%)."
  Adjusted.
- 249 The mean proportion was lower during the third period does not mean as it is stated that the "proportion significantly decreased toward the end of the experiment". It could well increase during the third period. Formulations that suggest that time has been measured continuously should be avoided. (cf. 251, 259, 262, 268, 279 etc.).

Done

- 262 Isn't a statistical test missing to support this sentence?
  Changed.
- 273 Explain what is this test. Shouldn't you also show that there is no significant interaction between D50-JD100 and period?
  This is part of the model simplification explained in the Methods.
- I do not understand how it is possible to have the higher number of deutonymph in the treatment were 100% of the deutonymphs are removed each week.
  We discuss this in the second paragraph of the Discussion.
- You provide many measurements on the size of the morphs but there isn't a clear hypothesis or predictions regarding these traits.
  Please see our revised Introduction.
- 286 You provide here the mean size. It would be nice to provide the mean values of each of the trait that you study.
  Means of treatments, or combined if they were non-significantly different, are given in the figures.
- 289 Refer here to figure 5B.
  Done
- 292-298 You present non-significant results. This approach can be understood but is in contradiction with the previous approach which consisted of simplifying the models and grouping treatments together even when the effects were marginally significant…
  Yes it is but we found these non-significant results to be important to our story, which is why we discuss these in the manuscript.

Discussion

- Discussion may have to be updated to take into account the potential above-mentioned suggestions of modifications.
  Done
- Nothing really is mentioned about the amplitude of the highlighted effects. I think that this should be mentioned and discussed.
- More could be said on the potential mechanism on which anticipatory plasticity may rely on? How does it work on other systems?
  We have revised our Introduction to provide more explanation of the different types of developmental plasticity.
- You could also discuss the limits of your experimental setup, the traits that you have missed that could be interesting to track down, the type of experiments that remains to be done to clarify things.
  We have taken this into consideration.
- ? 302

Unfortunately, it is not clear from the comment what we are required to address

REVIEWER 2

We would like to thank the reviewer for their comments, which have contributed to clarifying key aspects of the manuscript.

Introduction

Line 64: It is not clear what "condition distribution" is in this context. Is it the distribution of fighter/scrambler morph proportion, food availability, or the presence of metabolic wastes in the food?

Reply: The condition distribution is equivalent to population size-structure. We clarify this now in the third paragraph of the Introduction (Lines 72-94).

Authors may add a few lines explaining what kind of population responses are expected in anticipatory or mitigation developmental plasticity after line 67 which will help the readers to understand the expectations in each case.

Reply: Yes, this is a good suggestion and one that the other reviewer made too. We have added further explanation. Specifically, we used the environmental threshold model to create hypotheses on how a change in fitness or a change in population size and structure impacts alternative male phenotype expression evolutionarily (former) or ecologically (latter), and thereby fuel eco-evolutionary population change. We now mention this in the third paragraph of the Introduction (Lines 72-94).

Line 83-84: The working assumption that only juveniles of the good condition develop into deutonymph and all male deutonymphs molt into fighter males seem a bit farfetched given the sample size in the Deere et al. 2015 study was just 11 for deutonymph males, and most of the deutonymphs molted into females.

Reply: Although the number of males that emerged from a deutonymph was 11 in Deere et al. (2015) it was 75 in Smallegange & Coulson (2011), so we are confident to use this working assumption.

Methods:

Line 124: Typo in the spelling of and.

Reply: Corrected.

Line 135 onwards: The following sentences are not clear as to in which proportion the other juvenile stages from the J-D 100 and J-D 50 were removed. Is it the proportion of deutonymphs in the D100 and D50 treatment and 100 or 50 percent of those in the J-D 100 and J-D 50 treatment for that particular generation or it is 100 or 50 percent of the total deutonymphs present in the population from which the juveniles are removed?

Reply: We acknowledge that the description of how we removed juveniles from the control populations are not clear. Only protonymphs and tritonymphs were removed from the control populations. This was decided as: 1) larvae either develop into tritonymphs or deutonymphs and so protonymph removal would reflect individuals that did not develop into deutonymphs from larvae and 2) deutonymphs develop into tritonymphs before developing into an adult. To decide on the number of juveniles (protonymphs and tritonymphs) to remove, we first calculated the proportion of deutonymphs that were removed relative to the combined total of protonymphs and tritonymphs for a population. This proportion value was then applied to the combined total of protonymphs and tritonymphs in the control populations (J-D100 and J-D50) to calculate the number of juveniles to remove. The number of protonymphs and tritonymphs removed was calculated randomly. We have now added more text in the methods, second paragraph under

‘Experimental procedure’ section, starting in Line 212, to clarify this. We have also included a data file in the online appendix indicating the total number of protonymphs and tritonymphs harvested each week from the J-D100 and J-D50 populations.

Even in the datasheet in the tab R population counts, it is not clear in what proportion other juvenile stages are removed. For example, in column 150-151 the total number of juveniles removed is neither a reflection of 50 percent of the current population’s deutonymph number or 50 percent of the deutonymphs from the OA05 populations. Same is true for the other columns.
Reply: See comment above

Additionally, I didn’t see a column for the larvae that were removed in the OAC1 and OAC05, so it is not clear whether the larvae were removed or not.
Reply: See comment above

Line 129: It would be important to state what proportion of deutonymphs are seen in the regular maintenance of these populations, thus the readers could get an idea of the relative magnitude of juveniles being harvested from the population.
Reply: Good idea. We have added in the ‘Study system’ section (Line 194) that, in our source populations, mites rarely develop into a deutonymph and we previously estimated the percentage of males developing into one after removing food and imposing a dry period at 3% (Smallegange & Coulson 2011).

More importantly, is the proportion of juveniles big enough to induce a stress response in the populations to express the mitigating developmental plasticity? I ask because as I see in the datasheet, the proportion of deutonymphs removed from the population is small as compared to the total juveniles present in the population at that time. I am sceptical about how this small percentage of removal of juveniles from the population is going to induce a plastic response in the population. Authors may add a line or two on why they think this proportion may be consequential.
Reply: Indeed, we anticipated deutonymph numbers to be low, but we also suspected that our harvesting regimes would still induce population responses. For example, in previous experiments, we also imposed harvesting regimes within which we removed only a few individuals on a regular basis and found these selections to have significant impacts on the threshold for male adult phenotype expression and population size-structure (Smallegange & Deere 2014; Smallegange & Ens 2018). We have added this information to the ‘Experimental procedure’ section (Line 212).

Line 137&139: Typo: it should be ‘were’ instead of ‘where’.
Reply: Corrected.

Results
Line 233: The authors say “Mean total population size had stabilized across all treatments at the start of harvesting (Fig. 2)”. However, the mean total population size does not seem to have stabilized since the population size is still increasing for D-100, J-D-50 and control populations before the harvest began. Visually, it seems there is ~15-20% growth after the harvest started which is non-trivial. If population sizes are not stable or have not reached the equilibrium population size before the harvest began, there will be an interaction between the population growth rate and the harvested proportion of juveniles to influence the equilibrium population size, which may not represent the consequences of just harvesting on the population size. The authors

need to address this issue in Discussion and Methods or find a way to control for the effects of such growth in their analyses.
Reply: We removed the 'stabilising' text in the Results section and state that mean total population size increased over time since we started harvesting. We suspect that the increase in population size that we observed over the course of the experiment is associated with the observed decrease in mean body size over time in response to the strong, density-dependent food conditions, because smaller organisms typically reproduce faster and can reach higher carrying capacities than larger ones (we discuss this in the third paragraph of the Discussion (Lines 74-92)). Furthermore, in the Discussion (Lines 458-460) we have added that the 'delayed' response of reduced fighter expression with increasing time period could be due to that fact that our proportional harvesting treatments became more severe as population size increased over the course of the experiment. The increase in deutonymph numbers could likewise be explained as a numerical response (Lines 492-494).

Line 246: It is unclear why the proportion of fighters differs between the time periods. In the D-100 treatment, the proportion of fighter does not change at all time periods but in the D-50 treatment the proportion of fighters declines only in the last time period when the population size seem to have increased for D-50 (Fig.2).
Reply: Yes, this is an important observation. Given this and your previous comment, we have added this to the Discussion that the 'delayed' response of reduced fighter expression with increasing time period could be due to that fact that our proportional harvesting treatments became more severe as population size increased over the course of the experiment.

Does the population size increase significantly in the last time period to influence the expression of the fighter morphs in D-50?
Reply: Yes, it does and in response to your comments we have added this to the Discussion (please see above).

Line 291 onwards: I will suggest to not make inferences when the statistical differences are not significant, especially when the sample size for scrambler males is really low.
Reply: This text has been removed.

I think one of the important interactions that is not plotted is the total population size for all five treatments in different time periods (1, 2 and 3 which were used in analysis). It would be an important factor to know how decreased proportion of fighters would increase the population size. If the decrease in fighter proportion increases the population size then shouldn't the J-D 100 population should also have an enhancement in population size like D-50. This plot can be added to supplements if the authors think it makes the main article lengthy.
Reply: The interaction between harvesting treatment (all five levels) and time period was non-significant (see Results), which is why we did not plot it. But even though the total population size in J-D100 is the lowest, it still significantly increased over the course of the experiment, like D50, as you say. Perhaps because, overall, its population size was lower than in D50, the proportion of fighters in the population was slightly higher.

Discussion
It is important to highlight how an increase in population density could influence the expression of fighter morphs. Previous studies have shown that an increase in population density could lead to the suppression of fighter morphs, and as the population size is the highest in the D-100 treatment

the exuding pheromones from this population could lead to decline in the overall expression of fighter, since the population size is the highest in the D100 treatment.
Reply: Please note that for this species of bulb mite, *Rhizoglyphus robini*, male adult phenotype expression does not depend on the frequency of each male phenotype in the population (Deere & Smallegange 2014) nor on mite density (Radwan 1995). We have added this to the Introduction (Lines 103-105).

Another thing that I think is missing in the discussion is how the removal of deutonymph leads to removal of a higher number of potential females which could also affect the overall skew in the population sex ratio and how that would have affected the population size differently for D100 and D50 treatment. The removal of potential females could affect the egg output in the population and perhaps lead to a lower intensity of density-dependent selection as opposed to a condition when the female number is not altered leading to an increase population size.
Reply: This is a good point and we have added that to the Discussion (Lines 499-501).

Line 350 onwards: The explanation regarding the expression of more deutonymphs due to reduced olfactory cues seems unreasonable, because if the lack of deutonymphs affects the concentration of olfactory cues then the number of deutonymph should not be changing over the time periods for D100.
Reply: We have removed this text.

Line 380: I will suggest the authors to replace the term “carrying capacities” with “equilibrium population sizes” because when populations approach equilibrium, at that point due to higher population growth rates, the populations tend to overshoot the maximum sustainable population size.
Reply: Good point. We have adjusted the text.

Figures

Figure 3: Keeping the Y-axis same for 3A and 3B would be helpful for easy comparison.
Reply: Done.

Figure 4: Keeping the Y-axis same for 3 B, C and D would be helpful to compare the body sizes between males and females. Also, plot the Fig. 4 A, C, and D similar to 4 B, so that the consistency allows readers to compare the mean shift in body size for all the treatments and how that could have affected the mean population size or vice versa.
Reply: We made y-axes of fighters and scramblers the same within Figures 4 and 5 but not for females or deutonymphs because they were significantly smaller and then the differences would not be so obvious. Note that these are now Figures 6 and 7. We did not follow up on the last suggestion because we aimed to match the figures to the stats.