

Editor's comments:

Both reviewers and I really like the 'premise' of this article. Unfortunately, both reviewers were also quite critical of a number of aspects of the paper, including the background, the definitions, caveats, and model itself. Given these concerns, I am sorry to say that I cannot support publication of this paper in Ecology Letters. I see two options:

First, the authors could revise their manuscript as best as they can and seek publication elsewhere. Especially if they were to take on several of the reviewer comments, this might be a relatively straightforward task.

Second, if the authors feel that they can rather fundamentally alter the shape and structure of their manuscript, we might be willing to consider a reworked version. I should say, however, that given the nature of the reviews, and the detailed advice about the concerns and possible ways forward, that this would be a rather significant reworking bordering on a new submission. Such a revision would need to rework the model section, broaden the scope, and really tackle many of the caveats and issues brought up by the reviewers.

Of course, I completely understand if the authors choose the first pathway, as the second pathway would be a lot of work and there is no guarantee that it would satisfy the reviewers. Nevertheless, there is important potential for the authors and topic, and I wanted to leave the 'door open' should the authors be willing to take on this task.

We appreciate the editor's honest assessment of the state of manuscript and the task of a revision for *Ecology Letters*. We agree this is an important topic and the editor and referee's comments have led us to completely redraft the manuscript with a broader focus (we estimate that only 10-20% of the originally submitted text remains in the revised manuscript). We believe the revised submission better serves the current state of this field and could help rapidly advance progress in research on environmental tracking.

Referee 1 comments:

The authors present an interesting (I'd call it perspective not review) manuscript that is focused on what they call "environmental tracking". With that, they really mean the ability of species to shift their phenologies in response to changing environmental conditions. This is clearly an important and timely topic, and I was excited to see someone tackling this. This said, the title & abstract did not prepare me for the content, and I felt a bit let down. The focus is much narrower than both suggest, and does not provide clear insights into a community context. The main content of the paper is focused on why some species track long term changes (e.g. in temperature) and others don't, with some speculation of how this might affect competition. The literature is not reviewed comprehensively, and largely focused on a few systems and big reviews, without really digging into the available literature. I know there isn't as much out there on this topic, but there is more than is given here. For a review, that's not enough, for a perspective, it's still a bit short.

We understand the reviewer's concerns and have worked to completely re-draft the manuscript to provide a more substantial and useful review of the field, while still aiming to be forward-looking. We provide more details on these changes below.

The main part of the paper is a model, and that is where I had the most issues. I'll explain the details below, but it really appears to be a not well developed toy model (albeit I may have missed how exactly it worked, see below), and highly system, condition specific. As a consequence, much of what it shows we already know, or its unclear how we can generalize it to other systems and scenarios. There are no analytical solutions or comprehensive simulations & exploration of the parameter space. The authors also ignore much of the theory we already have on this type of model (it may not be called "phenology", but otherwise very similar). In the end, there are so many restrictions on the model that the outcome is known without any need of simulating the dynamics. There is even little to now discussion of existing models specifically on consequences on phenological shifts, which almost feels like intentional omission, but it's not clear why. While highly relevant (some could predict similar outcomes) none of them are discussed or but in context to current models, so we don't really know what's new or different. As a consequence, I don't think that this is a good fit for Ecology Letters. I would suggest the authors either focus on a more comprehensive literature review and drop the model component, or really dig into developing the model and focusing on the exciting new questions that could be addressed with it, but that deserves it's own paper. I know this is not an encouraging review, but I like the inherent idea and there is clearly a need for this topic to be emphasized, so I'd like to see more of this.

We thank the reviewer for their candid assessment and agree the mix of the model with the overview of the field did not work well. As such, we have removed the model from the main text and now review its results briefly in a Box ('Adding tracking and non-stationarity to a common coexistence model') with its description in the supplement. We have also tried to highlight where this model fits in a suite of models that could provide inference on tracking in stationary and non-stationary systems in a new section, *Tracking in multi-species environments* (see sectionmutisppstart-sectionmutisppend), which covers theory from plasticity, priority effects and coexistence models similar to the one we present.

Overall, I faced some major confusion with the model and really got stuck on many aspects of it. So let me go more in detail:

- Resource in this system is specified to not be renewed and only gets depleted. This is a reasonable assumption for some systems, but not for others, so it should be clarified and emphasized to avoid confusion. Importantly, this sets the system up for positive priority effect, i.e. resources are always at maximum at start of the season, so early arriver will always have a benefit over later arrivers. Again, this is reasonable for certain systems, but not for others so it requires some more explanation and justification. It also prevents consumers from overshooting, i.e. there is not punishment for arriving too early (before the resource). Later on the authors confirm this expectation on early arrival advantage. It would be good to cite some literature on this (this is a common optimality problem and has been used in wide range of models). However, there should also be some detailed discussion on what systems match these specific conditions, and which don't (e.g. systems where resources don't start at max but build up over time, systems where later arrivers have an advantage etc.)

The reviewer is correct that the resource does not renew each season, we now write (see non-renew) "one dominant (non-renewing) pulse of a limiting resource each season," to help clarify this. As we have grossly cut the text devoted to this model we do not go into great details over this now, but we have added discussion about the need for more models that place a cost on early arrivals throughout the manuscript (see modelcosts1, modelcosts2, modelcosts3).

- *It took me a bit to think through the model formulation to understand how “timing” is incorporated here, and I’m not sure I’m still totally clear on it. Part of it stems from confusion about the two time scales, within vs between years and the notation was confusing to me which one is which. For instance, is $g(t)$ the germination for year t , or for time t within a year? The latter would suggest that there is some sort of distribution of germination events within a given year, while the first would indicate a single event. From the wording, I assumed that it is indeed a single event per year. Furthermore, it appears that there is not difference in relative timing per se (say relative to the resource), but instead timing effects with a given year are solely driven by how many germinate in a given year, out of the total. So it’s not a question of “when”, but “how many”*

The reviewer is again generally correct here and highlights an important point we did not discuss in our original submission: the reality that most phenological events are a mix of both ‘when’ and ‘how much.’ Our model abstracts the ‘when’ to focus more on the ‘how much’ and thus may be confusing to some readers. To address this we now discuss these intertwined issues (whenhow1start-whenhow1end) and return to them in our community modeling section (whenhow2start-whenhow2end) where we argue many of the current models to address these questions focus on only one or the other issue (‘when’ versus ‘how much’) and we highlight the need for more work combining these aspects, along with costs to mis-timed events.

- *Overall, this confusion makes it hard to evaluate what the model does. If we stick with the one germination event a year, let’s assume both species are identical for sake of argument. In that case, both species appear at the same time, but at different initial abundances, creating solely numerical priority effects.*

- *It also assumes that per-capita effects are unchanged, which is a specific assumption that is reasonable for some systems, but not many others. In addition, it would ignore the temporal dynamics, i.e. temporal overlap of competitors should be different, but without an explicit start time, it’s not. Again, all this is based on not having enough information to determine how the model really works, but based on the supplemental information I assume all populations start at same time within a given year just at different abundances.*

In the model, there is one germination event per year that occurs at the same time for both species. Therefore, the model does not incorporate an explicit early arriver advantage. Instead, variation in the timing of the resource pulse leads to differential germination of the competing species: when the timing matches the preferred timing of species 1, then species 1 has higher germination, and vice-versa. We model tracking as the ability of a species to change its preferred timing from year to year to match the timing of the resource pulse. This model, therefore, focuses on the “how much” of tracking rather than “when”. In addition to adjusting the language we use to describe the model in Box 3.4 and in the supplement, we have now clarified this distinction in our manuscript both in respect to the model through a new figure (see Fig. ?? and in general in respect to the diverse ways the literature currently studies environmental tracking (whenhow2start-whenhow2end).

This confusion is further increased by not providing information on how “non-stationary” is modeled in this system. If there is no real timing, does this mean it’s modeled as move from environmental to biological timing? So “shift” results in decrease in number of individuals if biological timing doesn’t shift but environmental timing shifts earlier?

We agree this was not sufficiently clear in the original draft. The reviewer’s understanding is correct. Non-stationarity is modeled as a shift in the mean timing of the resource pulse (see the top panels of Fig. ??, which we have now moved into the main text). For each species, this shift in the timing of the resource pulse (τ_P top left panel in Fig. ??) moves the resource pulse closer or further from the species’ fundamental timing, τ_I , (cf. Panel C of Fig ??). When τ_P is closer to a species τ_I , that species’ germination rate is higher; as the mean timing of the resource pulse changes, the mean germination rate of each species changes. In our model, a non-tracking species has a fixed τ_I , and a tracking species has the ability to adjust its τ_I to match (to varying degrees) the changing τ_P .

So to summarize I took away these following assumptions:

(1) Simulations of within season population dynamics start at the same time, there is no temporal offset of population dynamics, and no “escape” from competition in time. So there is no explicit temporal niche modelled

There is no explicit intra-annual temporal niche, but there is a temporal niche inter-annually, which is critical to our aim to model tracking. We have worked to clarify this distinction this in the current version Fig. ?. A model with both intra-annual and inter-annual temporal niche dynamics would be an excellent area for future work, but we believe it could be difficult to tackle this at the same time as adding tracking and non-stationarity to a single model. We now clarify, starting on whenhownewmod, that this is an important area for future work.

S(2) temporal differences only affect starting densities not temporal dynamics or per capita effects. In other words, this is a model where phenological shift only affect reproduction (and thus numerical), not interaction effects.

Temporal dynamics affect species interactions through density. Interactions terms are not explicitly altered. We have attempted to clarify this by better contrasting models of the type we presented with those where species parameters are directly tied to the environment ?, see S2start-S2end, which read:

Building a changing environment into such models thus requires knowing how environmental shifts filter through to species-level parameters (?). For example, ? added the temporal environment to competition models by defining interaction strength as dependent on the temporal distance between species. In other models, the environment is more specifically defined.

S(3) There is always an early arriver advantage: which ever is closest to environment timing, has higher proportion of seeds emerging and will win (assuming all else equal)

Yes, arriving early or closest to the environmental start time is always an advantage; we now stress that this is the case across many models related to environmental tracking (whenhow2start-whenhow2end, and Fig. ??).

S(4) Season ends when resource is depleted to lowest R^ . So season is not ended by environmental conditions but resource availability, and just a function of competitor densities.*

Each season ends when the R^* value of the better resource competitor is met, which is generally driven by species but can also occur due to abiotic loss in some seasons. As this model is no longer a focus of the paper, we have not addressed these details in the main text, but have

clarified them in the supplement.

S(5) Germination function with difference in environmental vs. biological timing is non- linear.

Yes. We have aimed to clarify this in the supplement, but as this model is no longer a primary focus of the paper, we have not addressed these details in the main text.

S(6) Tracking parameter is difference between fixed vs. moving biological timing. Given this set of assumption, the generality of the model is strongly limited to a few systems/scenarios (very specific plant system), limiting the general inference that can be obtained from it.

The addition of the tracking parameter (which can vary from 0 to 1, which is from no tracking to ‘perfect’ tracking) shifts how well a species matches to its environment each year (where a high match yields to more offspring). This is a modification of a common coexistence model, used by Chesson widely (??), which is commonly applied to plant systems but extends in the simple form we use here to coral reef fish, forest trees and many other organisms (see ?) so we do not believe its inference is strongly limited.

We have worked to further highlight where this model fits within the current literature, which measures a mix of tracking of climate data to fundamental tracking (often studied in the trophic mismatch literature). We hope the changes throughout the manuscript relating to this will help clarify the extent to which this, and many other models, may apply.

Overall, I gained very little from the model, and as far as I can tell nothing new emerged from the model that we did not already know from other systems (e.g. much of this reminds me of stage-specific multi-parasitoid competition systems where the life stage of the host resembles the environment here, or a simple inter-annual model where reproduction varies across years, and this may or may not be correlated across species). In addition, given the specific conditions, the conclusion that non-stationary environments will change coexistence outcome has to be true given the model formulation, and could be easily inferred from recent Rudolf 2019 model (which shows how phenological shifts alter coexistence conditions). As far as I can tell, the only novel aspect here is the tracking aspect, which I quite liked. But, the way it’s implemented, it’s not dynamic but forced on the system and simply shifts the initial relative numerical abundance of species, so again the outcome could be inferred from a simple L-V type competition model. So I’m still struggling to understand why this model is necessary and what new insights we gained. Otherwise it just adds confusion, so maybe it would be best shown in a simple verbal or graphical model. In fact, I would strongly favor the graphical option, since that would be clearer, and outcome can easily be predicted without simulation the system from the many existing models we already have. I still think the tracking approach is very interesting, but hasn’t been fully developed to ask more detailed question on how tracking will affect long-term dynamics, and rigorously explores when and how it influences long-term dynamics. I think this deserve its own fully developed manuscript, and sticking it in here is really selling it short of its potential. This would also allow the authors to examine how many of the unresolved questions they list later on influence the outcome, e.g. what are consequences of tracking if changes in environmental conditions alter multiple aspects (e.g germination & per-capita effect) etc.

We appreciate the reviewer’s concerns about how useful the model is in this paper and have thus moved it to a box where we focus mainly on its outcomes via figures (after much discussion we have decided to keep the full model description in the supplement as we cannot derive its

findings from a verbal or graphical model).

Outside of the model, I generally liked the idea of getting a much better understanding of what species track environments, and which ones don't. I completely agree with the authors that we know way to little about this, and more research needs to be done. This said, the manuscript here did not feel like a review, but a "food for thought" short opinion paper. If this is truly supposed to be a review, I would expect a more thorough and quantitative analyses of the literature, since much literature was missed, and largely restricted to plant systems. So my main complaint here would be that it felt like it was just touching the surface and did not provide enough depth (i.e. go into exiting studies).

We agree that in focusing on the model, we had little room for a more thorough review of the literature. Our revised manuscript draws on literature from vertebrates, corals, plants, arthropods and more to provide a fuller sweep of the literature. In our focus on what may drive variation in tracking, we provide a literature review focused on understanding tracking within a syndrome or traits (see `traittradeoffstart-traittradeoffend` and Box 'Trait trade-offs with tracking').

While we understand the desire for a more quantitative review, many have recently tried this and ended up focusing more on methodological issues than ecological predictors (??). Indeed, the first author of this manuscript designed the statistical approaches in ? and knows first-hand how difficult it is to accurately measure 'tracking' across studies currently. We feel the critical needs for this field now are (1) a greater use and development of theory to provide testable predictions and (2) better definitions and guidance on how to define and measure tracking so quantitative reviews will be possible in the future. Given these needs, we have written a manuscript that we see as most useful to the field now.

Finally, there was very little coverage over theory on phenological shift. The authors mention Rudolf 2019 in passing, without discussing any similarities, differences that are clearly there. Similarly, they never mention other phenology models, like Nakazawa & Doi, 2012, Revilla et al 2014 etc. Even the simple graphical temporal niche approach that the authors introduced themselves (Wolkowich & Cleland 2011) is not discussed (but brings up interesting question about "single" vs multiple resources approaches).

This is an excellent point. We have worked to better frame where ? fits within many ways of introducing phenology into current multi-species models (see `S2start-S2end`) and we have worked to build in more references to the trophic mismatch literature, where is what Nakazawa & Doi, 2012, Revilla et al. 2014 focus on (e.g., `tmm1`, `tmm2`). Additionally, we cite these references on `S2end`.

Specific comments:

P 3 L30ff: the notion that earlier spring should favor earlier phenologies relies on the assumption that the "niche" is empty i.e. no other species are earlier. So this is applies to very specific systems (i.e. resources are not available before that time point, so temporary resources) and should be clarified.

We have adjusted this text to now read (starting on `r1sc1start`):

Considering tracking as a form of plasticity, evolutionary models predict species that

track will be favored in novel environmental conditions. Similarly, some models of community assembly suggest that a warming climate should open up new temporal niche space and favor species that can exploit that space (???).

And we provide a longer discussion of alternative models throughout the manuscript now, especially in the new section, *Tracking in multi-species environments*.

L34-35, there are some studies (and should be cited here), e.g. Block et al 2019 Oikos. Showing that phenological plasticity is a poor predictor of performance.

In the current version of the manuscript this sentence no longer exists, but we have added a reference to the study in the same paragraph (starting on r1sc2start), “Yet not all studies find the purported link (e.g., ?), and there has been comparatively little work to improve predictions by formally connecting tracking to foundational ecological theory.”

L 50ff: there has been progress, e.g. Rudolf 2019 specifically incorporates non-stationary systems and variability to examine how it influences coexistence and communities (since it’s focused on phenology it seems like a highly relevant citation here). In fact this citation would be great to support the claim that it matters, instead of simply stating that nobody looked at it (which is incorrect).

We have added this citation where requested.

Equaton 9: “n” is undefined. Along the same line, what determines the end of a growing season?

We apologize for these omissions, n is the number of species, and each season ends when the R^* value of the better resource competitor is met. As this model is no longer a focus of the paper, we have not addressed these details in the main text but they continue to be provided in the supplement.

P13 L 9: this prediction hinges on the assumption of early arriver advantage and single resource competition etc. So as it stands, this is one of the predictions, not the only one.

Agreed, these predictions fail when there are costs to tracking too closely, we now discuss these costs in detail in multiple places in the modeling section of the manuscript (see modelcosts1, modelcosts2, modelcosts3).

P15L10ff: what about species that are just very plastic, i.e. can adjust to cope with various environmental conditions, and thus take an alternative strategy to shifting. There has been increasing discussion of phenotypic plasticity vs let’s call it “environmental” plasticity, i.e. species that can perform equally well at cold and warm temperatures. So environmental generalists.

We now discuss plasticity in depth on sectionmutispstart-endplastic.

Same page, next paragraph (sorry, having not continuous line numbers across pages makes this a bit frustrating). Good examples here would be species where phenologies are correlated across season/life stages. In some cases, phenologies in spring are determined by what happens in fall, or what happens later in summer may depend on how individuals perform during earlier life

stages in spring (e.g. changed developmental rates alter later phenologies etc.) In same context, Yang & Cencer 2019 Ecology examine “seasonal windows of opportunity”, which fits nicely in the context here. They took rigorous approach in finding what constraints those windows, which would also determine how shifts in them would change the optimal window.

We thank the reviewer for these examples, we now cite ? on yang2020 and here and throughout the manuscript have worked to discuss in-depth the constraints on tracking (e.g., constrain1start-constrain1end and constrain2).

Our apologies about the line numbers, these are the line numbers provided by *Ecology Letters*; we now provide continuous line numbers and refer to those here (though this may mean that after *Ecology Letters* adds their own line numbers we end up with two contrasting sets of line numbers, which we apologize for).

P16 L 34-35: very cool, I’ll have to look up change-point and hinge models, never heard of them!

P 16 L38ff: some recent approaches suggest using whole phenological distributions can strongly increase power as well (e.g. single species: Steer et al 2019 Methods E&E, or for species interactions Carter et al 2018 Ecol Letters)

Good point! We have added these references to lines addistribrefs and we now also discuss the issue of these events as a distribution on whenhow1start-endistrib.

Referee 2 comments:

1. Need clearer motivation in the introduction (section 1, “main text”).

a. What is the specific definition of tracking applied here? Tracking a set of abiotic conditions? Does it extend to tracking biotic conditions? Is there a way to quantify the relevant set of conditions, and therefore an organism’s ability to track them?

This is an excellent point and we now provide an extended definition of environmental tracking (see section, ‘Defining environmental tracking,’ define1start-define1end) and a new figure (Fig. ??) to help highlight the complexity of defining this. We believe this is an important addition to the literature, where ‘tracking’ is often used but rarely defined.

b. The second paragraph of the introduction suggests that “a shift toward earlier spring should favor earlier species, especially those that can environmentally track ever-earlier seasons” I’m not sure I follow the logic here; it seems like earlier spring conditions could just as easily limit the success of early spring species in particular. It’s not that the proposed hypothesis is never true, but it also doesn’t seem that it is likely to be necessarily or generally true, at least based on the argument presented. Is this intended as a straw hypothesis?

We have restructured the introduction to try to address this concern. We do not mean for it to be a straw hypothesis, but instead one that has gained traction in the literature and has some basic support from simple models, as we now state.

c. It seems like the assumption of stationarity has never been true, and ecological theory has always been a bit uneasy about this. Though maybe because so much of ecological theory is generally explanatory rather than specifically predictive, these deviations haven’t been too trou-

blesome. Perhaps the question then is more about how much worse the situation is with rapid climate change.

We agree with the author’s point. Most systems can appear either stationary or non-stationary depending on the scale and temporal period. We have tried to clarify this on nsstart-nsend, where we state:

Understanding the impacts of climate change further requires recognizing that most systems can be considered stationary or non-stationary depending on the timescale and period of study. Thus, predicting the consequences of current non-stationarity in ecological systems benefits from identifying the type and scale of non-stationarity, relative to long-term trends.

2. Environmental variability and change (1.1)

a. L21-24: *Is there good evidence of historical stationarity? The distinction between stationary vs. non-stationary environments seems to be scale-dependent, and thus somewhat subjective. Is that a problem?*

We believe there is good evidence for climatic stationarity, at least on ecologically relevant (and certainly researcher-relevant) timescales, and have adjusted text on lines nsstart-nsend and nsfutstart-nsfutend to clarify this.

3. Environmental tracking in time (1.2) a. *Chmura et al (2019) suggest that relatively little is actually known about the mechanistic/cueing bases of differences in phenological shifts, either because most studies don’t consider cues per se, or because they very rarely assess alternative mechanisms. If this is true, how does this affect the framework described in this section?*

This is a great point and we believe an area where this paper can offer some guidance. In our overhaul of the paper we now address this in two new sections *Defining environmental tracking* and *Measuring environmental tracking* on definedstart-endmeasure.

b. *The trade-off between plasticity (“tracking”) and bet-hedging has been examined in studies by Chevin, Lande, Ghalambor and others. Do those studies provide a useful perspective here?*

Yes, we now review the plasticity literature on lines sectionmutispstart-endplastic, including references to these studies. For example, plasticquotestart-plasticquoteend, we write:

Another area of life-history theory, that focused on plasticity, may be primed to provide insights on non-stationarity (or ‘sustained environmental change,’ see ?). Considering phenology as a trait (as we and others do, e.g., ???), environmental tracking is one type of plasticity. Researchers could thus more broadly understand environmental tracking through modeling an organism’s reaction norms (??) and understanding how cues and suites of cues—across environments—determine how fundamentally plastic an organism may be in its tracking. For example, multivariate cues should yield higher plasticity in this framework. From here, models of the role of plasticity in novel environments provide an important bridge to understanding the outcomes of non-stationarity, generally predicting non-stationarity should favor highly plastic species. This outcome, however, assumes there are no costs related to plasticity (??).

c. *The long-term value of plasticity vs. bet-hedging may not be apparent in relatively short field studies, since the relevant measure of fitness could require more time to assess.*

Agreed, we now discuss bet-hedging in more detail on lines bh1start-bh1end, and state, “Assessing bet-hedging in many systems, however, requires studies of fitness over longer timescales than many current field experiments.” See also bh3.

4. *Interspecific variation in tracking (1.3)*

a. *I’m a little concerned about the slant of this first sentence, which seems to suggest that tracking is both universally important and positive. Modeling studies seem to suggest that under some circumstances, more plastic responses could be maladaptive. The section goes on to identify some very interesting potential trade-offs with competitive ability, but the broader point is that it doesn’t seem to be entirely clear that “tracking” per se is universally favored even absent a competition trade-off. Perhaps this goes back to our limited ability to quantify “tracking” ability, and the implicit assumption that we can assess an organism’s ability to find optimal conditions. In most systems, it seems like we don’t have enough data to quantify tracking ability. In the absence of this, we can assess plasticity to specific cues, but whereas tracking may implicitly imply adaptive plasticity, plasticity is not always adaptive.*

Agreed, the opening sentence now reads (starting on newopenmeasure), “Attempting to measure environmental tracking and compare variation in it across species, space and time is a rapidly growing area of ecological research (????, e.g.).” Further, we have overhauled the manuscript and now address many of these concerns throughout the new sections *Defining environmental tracking*, *Measuring environmental tracking* and *Understanding variation in environmental tracking* from definelstart to plasticquoteend.

b. *L55-56. Because many climatic cues are correlated, and also correlated with other cues (photoperiod, biotic, etc), the observation that temperature models can explain more than 90% of variation in phenology probably shouldn’t be assumed as evidence of causation. Temperature in particular can be a very complex cue, and the determination of mechanistic causation is difficult, as described by Chmura et al (2019).*

Agreed, we have deleted this note and have worked to stress the complexity of potential cues throughout the manuscript, including multiple references to ? and related work.

5. *Model description and simulations (1.4.1)*

a. *This model conceptualizes “tracking” ability as a variable between 0 and 1 which describes an organism’s ability to adjust its biological start time to the (optimal?) environmental start time in a given year. This leaves aside some messy but potentially interesting issues of mechanism and constraint, including any explicit consideration of cues or environmental conditions. I’m not sure how I feel about this approach. This could be an effective way to focus on the issue of “tracking” per se, but also risks being too far abstracted from reality to provide a meaningfully realistic model. For example, how should we conceptualize “tracking” ability if the optimal start time becomes worse over time? Or if there is a disconnect between cues and conditions (i.e., an optimal tracking of cues leads to a poor tracking of conditions)?*

This was a concern of both reviewers and it highlighted for us the complexity in modeling phenology. This model definitely is a step removed from costs of tracking and we have worked to highlight this. Our new section on *Tracking in multi-species environments* reviews relevant modeling from the plasticity literature, priority effects and our modelling approach, among oth-

ers. We hope it provides a much more useful and broader view of the challenges in modeling ‘tracking’ and the broad relevant literature for this issue.

6. *Tracking in stationary environments (1.4.2) a. If I’m understanding this model correctly, there is the assumption of some kind of intrinsic circannual rhythm (represented by the fixed biological start time) which is then modified by cues (abstractly represented as “tracking”) to yield an effective or realized start time. This seems different than my understanding of circannual rhythms and zeitgebers in a potentially important way, where the current model would assume that even in the absence of any cues (or with a tracking ability of 0), an organism would consistently start on the same calendar day each year. This seems like a modeling decision that should be explained and justified. Are there studies to indicate that this is a reasonable model?*

This model is a form of one commonly applied to plant systems but extends in the simple form we use here to coral reef fish, forest trees and many other organisms (see ?). Part of why this model can be applied broadly is in its abstraction, which is also why it may be difficult to link neatly to phenological events. The addition of the tracking parameter (which can vary from 0 to 1, which is from no tracking to ‘perfect’ tracking) shifts how well a species matches to its environment each year (where a high match yields to more offspring).

Given both reviewers’ concerns regarding the model we have now moved it to a box (‘Adding tracking and non-stationarity to a common coexistence model’), which precludes an in-depth discussion of its exact potential conceptualizations. We have, however, worked to stress the varied interpretations of ‘tracking’ throughout the text (e.g., `define1start-define1end`), including across different modeling approaches (`startmodels-sectionmutispend`).

7. *Tracking in non-stationary environments (1.4.3) a. I get that this is not intended to be a realistic climate change scenario, but wasn’t able to understand the details of how the non-stationary environment was created without the SI. The key thing that seems clear is that the non-stationary environment favored earlier start times. It wasn’t clear if the optimal start time was actually advancing gradually over time, or if it was just changed in a single step. If I understand it correctly, this model doesn’t allow for any evolutionary responses.*

This was also a concern of both reviewers and, given that part of our aim was to show a model transitioning from stationarity to non-stationarity, we are sorry we failed at this. We have now moved the relevant Figure from the supplement to the text (Fig. ??).

8. *Model conclusions (1.4.4) a. The observation that tracking is favored seems to be almost an assumption of the model, rather than a conclusion. Could it be otherwise in this model?*

Tracking in this model is favored in the same way that a lower R^* is favored in the model. We have worked to clarify this as much as possible, while still keeping the text related to this model within the limits for a Box. We have worked to focus more of the main text on contrasting models and additional approaches (`startmodels-sectionmutispend`).

b. *The idea that “tracking” should be considered as a part of larger “trait syndrome” seems appealing, though I’m not entirely sure what it means. What are the other parts of this syndrome? My concern is that the idea of “tracking” ability per se is not sufficiently defined or justified to develop in this way, abstracted from cues and physiological mechanism.*

This is a great point and we have re-drafted the manuscript to dig in deeper on defining and measuring tracking and we return to that definition (and its often multivariate scope) in the new section on *Tracking in multi-species environments* (see sectionmutispstart-sectionmutispappend).

c. Despite this, I actually like the idea of a trade-off between tracking ability and competitive ability; it seems intuitively appealing, if not clearly defined. I think I'd like some additional justification that the idea of "tracking ability" is a meaningful one in nature, and that there are empirical reasons (not just based on theory) to think that it trades off with competitive ability. As a counterpoint, it seems like phenological traits are just as likely to be used as a tool in competition, where an organism may benefit by showing an earlier phenology in the presence of competitors (due to pre-emption, or asymmetric competition, e.g. for light), even when it would do better to have a later start in the absence of competitors. This requires a more careful definition of "tracking" is an organisms that deviates from its optimal timing in an abiotic-only context showing good tracking or poor tracking? What if a deviation from the abiotic optimum is favored under competition? What if competitive ability depends on the relatively phenological/ontogenetic stages of the competitors? Instead of thinking of ways in which phenological tracking and competitive ability trade-off, I'm left wondering more about the complex ways in which they could interact.

We agree and appreciate the reviewer pushing us to better define tracking. Our manuscript is in many ways re-written around this aim with new section on defining and measuring tracking. We provide some examples of empirical trade-offs (traittradeoffstart-traittradeoffend, which read:

As tracking often relates to the timing of a resource pulse, traits related to resource acquisition are likely contenders for a trade-off. Species with traits that make them poor resource competitors may need to track the environment closely to take advantage of transient periods of available resources, but will risk tissue loss to harsh environmental conditions more prevalent early in the season (e.g., frost or snow). In contrast, species with traits that make them superior resource competitors may perform well even if they track environments less closely, because their resource acquisition is not strongly constrained by competitors. Examples include under-canopy species leafing out earlier to gain access to light (?) or species with shallow roots starting growth sooner in an alpine meadow system, while species with deeper roots begin growth later (?). In such cases, tracking is akin to a competition-colonization trade-off (?), where species that track well gain priority access to resources and, thus, may co-exist with superior competitors. Research to date supports this, with several studies linking higher tracking to traits associated with being poor competitors (???). Further, many studies have found a correlation between higher tracking and 'earliness' each season, which has been linked to resource acquisition traits associated with lower competitive abilities (? , see Box 'Trait trade-offs with tracking').

As mentioned, we also provide further detail in the Box 'Trait trade-offs with tracking,' however, we feel much more is needed here and thus focus on it further in our future directions section (fdtradeoffsstart-fdtradeoffsend).

9. Future research (1.4.5) a. While I agree that improved predictions of climate change would be valuable, it isn't clear how these improved (i.e., more complex) climate predictions would benefit this model in particular. This model already seems quite far abstracted from cues and mechanism. More generally, I actually get the feeling that climatic projections are constantly improving

though improved climatological models (especially better local or regional scale models), but our ability to predict ecological outcomes (coexistence or otherwise) is not typically limited by the detail, complexity or resolution of these climatological projections.

We agree. We have placed our modeling results in a box to focus on these bigger issues and we now address the climate projections as needing to focus more on how climate change will impact how we measure tracking, see understandquotestart-understandquoteend of the subsection *Understanding and measuring ‘tracking’*, which includes:

Understanding how the environment is changing represents just one step along the towards robust measures of environmental tracking. Shifting environmental regimes must then be filtered through species cues. As robust estimates of these cues are currently available for few if any species, we suggest several major improvements on current methods to help interpret current trends and make comparisons more feasible.

Studies should clarify their definition of tracking, how the environment is defined and how well, or not, the underlying cue system is understood for study species. Currently, many studies examine fundamental and environmental tracking at once (e.g., ?), which is clearly helpful in advancing the field. However, the more researchers can clarify when and how they are addressing fundamental tracking versus environmental tracking the more easily we can compare across studies. Next, and relatedly, studies should define their environment: are they considering primarily the abiotic environment or measuring an environment fundamentally shaped by other species? This difference connects to fundamental versus realized niches and whether systems are primarily top-down (resources and the environment may be strongly shaped by other species) or bottom-up controlled. Finally, all researchers working on environmental tracking need to embrace their inner-physiologist, or collaborate with one. For many organisms there is often a related (perhaps sometimes distantly) species that has been studied for which cues underlie the timing of the life history event. Researchers should draw on this literature to bracket which environmental variables may represent true tracking and which may be proxies, and to highlight uncertainty. We expect progress will come from balance between measures of fundamental tracking, estimating an organism’s system of cues and measuring environmental tracking. Clear statements of what is and is not known and measured will help.

Embrace the sometimes contradictory pulls of conducting experiments to identify mechanistic cues and the multivariate climate of the real world. Clearly, we need more experiments to identify which specific aspects of the environment different species cue to and how these cues are filtered by their actual environmental regime (as outlined above). Suites of experiments that build from identifying cues, to understanding how they act when correlated are a major gap for most organisms.

As an aside, while climate models have accurately predicted general trends and anomalies (see Hausfather *et al.* 2020, Evaluating the performance of past climate model projections in *Geophysical Research Letters*, they have had limited success in predicting many extremes. Further, their reliance on model ‘tuning’ has made it difficult to understand mechanistically what underlies important divergences between models (see Knutti *et al.* 2017, Beyond equilibrium climate sensitivity in *Nature Geoscience*). So these models are amazing, but not all agree they are

always improving.

b. I would also be interested to know more about potential trade-offs between “tracking” and other traits, but would want to know first whether “tracking” ability is a meaningful construct. In this model, tracking is mathematically defined as inversely correlated with the difference between the intrinsic timing and the (optimal?) “environmental (abiotic)” start time. My sense is that there are very few systems where either the intrinsic start time is a realistic concept, or where the (optimal?) environmental (abiotic) start time has been well- characterized. If there are good examples of systems that support these concepts, they should be described. If this model is intended to provide more of any abstract framework, I would suggest that these caveats of definition and characterization should be much more prominent, and assessing these issues would probably be valuable future directions.

We agree this model is an abstraction (as all are) and we can see it was perhaps not the most useful one here, thus we have moved the model to a box and focused more of the text on defining, measuring and building depth across multiple areas of community ecology theory to better understand tracking.

c. Despite my concerns about the framework of this paper, I do think the question of how climate change will shape coexistence mechanisms is an interesting one. I’m not entirely convinced that this model sheds much light on this issue, but would be glad to be convinced otherwise.

We thank the reviewer for their comments, which helped us re-envision this paper. We hope the new version addresses some of the concerns and will provide a path forward for this field of research.

10. Boxes

a. The three boxes in this manuscript touch upon some of the issues that concern me about this manuscript, albeit too briefly. It seems clear that the authors have thought about some of these issues. Why not examine some of these complexities more centrally in this manuscript?

Agreed. We have moved some of the text from the Box ‘What underlies variability in species tracking?’ into the main text in the new sections *Defining environmental tracking*, *Measuring environmental tracking* and *Understanding variation in environmental tracking*.