

Running title: Tracking & climate change

How temporal tracking shapes species and communities in stationary and non-stationary environments

E. M. Wolkovich¹ & M. J. Donahue²

¹ *Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4 (e.wolkovich@ubc.ca)*

² *Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, Kānā'eohe, HI 96744 (donahuem@hawaii.edu)*

Corresponding author: see¹ above; Ph: 604.827.5246 (no fax).

Authorship statement: EMW and MJD both conceived of the paper, performed modeling work and edited the paper, EMW wrote the paper and did the literature review, while MJD wrote the supplementary information on the model.

Data statement: Review, so no new primary data, but data from a comprehensive literature review will be archived in an appropriate public repository and the data DOI will be included at the end of the article.

Keywords: community assembly, global change, climate change, phenology, environmental variability

Article type: Reviews and Syntheses

Article information: Abstract: 169 words; Main text: 5,254; Figures: 4; Boxes: 4 (text in Box 1: 367; Box 2: 800; Box 3: 265, Box 4: 743); 115 references

Abstract

Climate change reshapes the environments of all species. Predicting species responses requires understanding how well species can track environmental change, and how such tracking shapes communities. Growing empirical evidence suggests that environmental temporal tracking—how an organism shifts the timing of major biological events in response to proximate environmental cues—is linked to species performance and is a structuring force of species and communities today. Here, we review the concept of tracking in empirical studies and through the lens of community ecology theory. We argue that much current theory for tracking ignores the importance of a multi-species context beyond trophic interactions. Yet community assembly theory predicts competition should drive variation in tracking and trade-offs with other traits. Existing community assembly theory can help us understand tracking in stationary and non-stationary systems and, thus, predict the species- and community-level consequences of climate change. But it will require greater efforts to integrate priority effects into modern coexistence theory and improved empirical estimates of tracking and the underlying environmental cues.

1 Main text

Anthropogenic climate change is causing widespread changes in species distributions, with many species shifting in both time and space (IPCC, 2014; Hoegh-Guldberg *et al.*, 2018). Species are moving to higher elevations and poleward (Chen *et al.*, 2011), shifting their recurring life history events (phenology) earlier (Wolkovich *et al.*, 2012; Cohen *et al.*, 2018), or both as climate warms (Amano *et al.*, 2014; Socolar *et al.*, 2017). These general trends, however, hide high variability across species. A large proportion of species are not shifting (Cook *et al.*, 2012; Amano *et al.*, 2014), which has raised concerns about whether these species may be more vulnerable to population declines with continued warming. Such concerns come in part from increasing research that links how well species track climate change—especially through temporal shifts—to changes in biomass, growth and other metrics related to performance (Cleland *et al.*, 2012). Tracking—a form of phenotypic flexibility—may then be a major component to understanding and predicting the fitness consequences of climate change, including population declines, with cascading effects on community and ecosystem structure (Menzel *et al.*, 2006; Parmesan, 2006).

The hypothesis that tracking improves fitness outcomes with climate change has gained significant traction in the ecological literature (e.g., Cleland *et al.*, 2012) and several areas of theory support it. Evolutionary models predict species that track will be favored in novel environmental conditions (Chevin *et al.*, 2010). Niche models of community assembly suggest that a warming climate should open up new temporal niche space and favor species that can exploit that space (Gotelli & Graves, 1996; Wolkovich & Cleland, 2011; Zettlemoyer *et al.*, 2019). However, empirical data on the costs (and benefits) of tracking are mixed (e.g., Block *et al.*, 2019). Further, there has been comparatively little work connecting tracking to community assembly theory, which shows temporal sequencing and environmental variability can alter the relative fitness and niche differences between species that determine coexistence, suggesting important ecological constraints to tracking.

This disconnect could be because most ecological theory was constructed for stationary systems (e.g., Chesson & Huntly, 1997). While major arenas of research such as ‘modern coexistence theory’ or population ecology now embrace environmental stochasticity, they generally still assume stationarity, where the underlying distribution of the environment is unchanged across time (i.e., constant mean and variance, Barabas *et al.*, 2018).

Climate change upends the assumption of stationarity. By causing increases in temperature, larger pulses of precipitation, shifts in snowpack, increased drought, and more storms (Stocker *et al.*, 2013), climate change has fundamentally shifted major attributes of the environment from stationary to non-stationary regimes (see Fig. 1 and Box: Environmental variability & change). This transition is reshaping ecological systems. New work has aimed to adapt coexistence theory to non-stationary environments (Chesson, 2017; Rudolf, 2019). Yet there is still little theoretical work on what such a transition may mean for communities and the species within them, including how community-level processes might feedback to modify species responses.

Here, we review the concept of tracking as commonly used in the empirical climate change impacts literature and in related ecological theory. We provide definitions that distinguish between fundamental and environmental tracking, highlighting the distinction between measuring

tracking and its fitness outcomes in empirical systems. Then, after briefly reviewing evolutionary theory that predicts variation in tracking across species and environments, we examine how well community assembly theory—especially priority effects and modern coexistence theory—can be extended to test the current paradigm that climate change should favor species that track.

1.1 Defining & measuring tracking

Defining tracking

Tracking is a commonly used word in the climate change literature, especially relating to phenology (e.g., Menzel *et al.*, 2006; Parmesan, 2006; Cleland *et al.*, 2012; Deacy *et al.*, 2018). Yet there are few, if any, definitions of it. Conceptual and theoretical treatments of tracking often relate how well an organism matches the timing of a life history event to the ideal timing for that event, what we refer to as ‘fundamental tracking.’ In contrast, empirical studies of tracking often focus on estimating a change in the timing of an event relative to a measured environmental variable, with the aim of measuring what we refer to as ‘environmental tracking’ (Fig. 2)—the change in timing of a major biological event due to an organism’s cue system given change in the environment (though most studies lack the required knowledge of the underlying cue system, Chmura *et al.*, 2019). Both these definitions are readily applied to phenology—the timing of recurring life history events—though they can also apply to non-recurring life history events (e.g., seed germination), or events not normally defined as part of life history.

Fundamental tracking rests on an assumption that there is a timing (an ‘ideal timing’) that yields maximum fitness, and event timings moving away from this ideal result in reduced fitness. This is a foundational concept of the trophic mis-match literature (Visser & Gienapp, 2019), which often assumes the peak timing of a resource defines the ideal timing for phenological events dependent on that resource (e.g. egg laying dates dependent on caterpillar abundance, Visser & Both, 2005). For most events, however, fitness outcomes are likely dependent on a suite of interacting forces (e.g., Reed *et al.*, 2013)—for example, egg laying dates for migratory birds may depend both on the timing of peak prey abundance and the need to leave nesting grounds before winter. Additionally, the timing of events is often connected to the investment process of an event (e.g., number of eggs in a clutch, Inouye *et al.*, 2019) in current, and sometimes past and future, years. Whatever their full underlying causes, fitness consequences of suboptimal timing should drive the evolution of organismal cues to predict and match the optimal timing (the degree of this match defines cue reliability, Fig. 2). Environmental tracking combines the outcome (timing) of these cues with environmental variation.

Environmental tracking depends on the intersection of the environment’s variability—which aspects of the environment vary, how (e.g., temporally each year, spatially at x scale) and how much—and an organism’s response to the environment via its proximate cues. If the varying components of the environment are not in the organism’s set of cues, then the organism does not ‘track’ per this definition (although covariation with other environmental components might give the appearance of tracking). Environmental tracking at the individual-level is a purely plastic response to environmental variation (in line with current findings on most climate change

responses, Bonamour *et al.*, 2019), with the plasticity itself an outcome of selection (Chevin *et al.*, 2010) through its connection to fundamental tracking (Fig. 2). At the population-level, tracking may also incorporate evolutionary change—change in genotype frequencies—depending on both the timescales of study and the species’ generation time (this evolutionary response can be predicted as the difference between the environmental sensitivity of phenotypic selection and an organism’s plasticity, $|B - b|$ in Chevin *et al.*, 2010).

Measuring tracking

Measuring ‘tracking’ and comparing variation in it across species, space and time is a rapidly growing area of ecological research (e.g., Cook *et al.*, 2012; Fu *et al.*, 2015; Thackeray *et al.*, 2016; Cohen *et al.*, 2018). Studies that directly quantify fundamental tracking are uncommon (but see Visser *et al.*, 2006; Charmantier *et al.*, 2008), given in part the difficulty of estimating fitness, though many studies in the synchrony literature attempt to link consumer change to resource change, with an assumption that the measured resource is the dominant determinant of ideal timing for the consumer (though this may rarely be true, see Singer & Parmesan, 2010; Johansson & Jonzen, 2012; Reed *et al.*, 2013). Instead, most studies focus on estimates closer to environmental tracking. Some studies estimate simply change in days over time (e.g., Parmesan, 2007; Kharouba *et al.*, 2018), though most studies now estimate shifts as responses per unit temperature (for example, multiple meta-analyses show plants’ spring phenology shifts with spring or annual temperatures 4-6 days/°C on average across species, Richardson *et al.*, 2006; Wolkovich *et al.*, 2012; Thackeray *et al.*, 2016) or precipitation (Inouye *et al.*, 2002; Craine *et al.*, 2012).

All species-rich studies of phenology-climate relationships find high variation (Cook *et al.*, 2012; Thackeray *et al.*, 2016), including some species that do not track or track poorly (i.e., high noise surrounding observed statistical relationships). Researchers have worked to link such variation to the underlying cues (e.g., Cook *et al.*, 2012), species traits (e.g., Cohen *et al.*, 2018) and trophic level (e.g., Thackeray *et al.*, 2016). These approaches hint at the three majors explanations for why some species do not appear to track climate or appear to track poorly: (1) environmental tracking is either not possible or optimal for all species or in all environments (Simons, 2011), (2) researchers have measured the wrong environmental variable (i.e., a variable species do not track, Chmura *et al.*, 2019), and (3) statistical artifacts that make it difficult to measure tracking robustly (see Box ‘Challenges & opportunities in measuring tracking’).

These confounding factors may make many current estimates of interspecific variation in tracking less reliable than they appear. This in turn makes robust quantitative analyses across species difficult (Brown *et al.*, 2016; Kharouba *et al.*, 2018), yielding a muddy picture of which species, when, and where, do and do not track. Given this difficulty, we believe clear testable predictions from ecological theory are especially critical to guide the field forward (Smaldino & McElreath, 2016).

1.2 Tracking in single-species environments

Predicting variation in environmental tracking in stationary systems

Evolutionary models predict strong selection for tracking in heterogeneous environments where

there are predictable cues for the ideal timing of events and the underlying genetics to develop a heritable cue system (Piersma & Drent, 2003; Reed *et al.*, 2010). Tracking is likely strongly heritable, given that many cue systems are themselves heritable (e.g., van Asch *et al.*, 2007; Wilczek *et al.*, 2010). The strength of selection is determined by the costs and benefits of cues (Donahue *et al.*, 2015). The costs include the machinery an organism uses to monitor its environment (e.g., accumulated temperature or daylength), while the benefits are the increases in fitness gained from better timing (e.g., how much tissue is saved by avoiding a coldsnap). Selection, however, can be lower than expected from reaction norms predicted by simple evolutionary models for many reasons, including trade-offs with tracking (Singer & Parmesan, 2010; Johansson & Jonzen, 2012), gene flow from other environments that may continually push a population away from its local optimum (Lenormand, 2002), limits due to standing genetic variation (Franks *et al.*, 2007; Ghalambor *et al.*, 2015), or deeper evolutionary history that may produce co-evolved traits making it difficult for selection to act solely on tracking (Ackerly, 2009).

Apparently poor cues may occur for organisms in environments where there is both a low cost and low benefit to the cue(s). Whereas expensive cues, such as complex multivariate systems, are possible given a high pay-off. Most in-depth empirical studies of species' phenological cues find evidence for complex multivariate systems that appear adapted to handle unusual—though not completely uncommon—years (Chaine & Regniere, 2017). This suggests that multivariate cues may better couple environmental tracking to fundamental tracking, while simple cues are more likely to trigger growth or reproduction at a suboptimal time.

Tracking should generally not be favored in unpredictable environments (e.g. when early season climate cannot be used to predict later season climate), or environments where species otherwise face high uncertainty in the timing of investment decisions. Instead theory suggests the optimal strategy may often be to bet-hedge (Venable, 2007; Donaldson-Matasci *et al.*, 2012; de Casas *et al.*, 2015) via a high diversity of timings or a conservative timing. Because bet-hedging, by definition, maximizes geometric-mean fitness in the long-run, its short-term outcomes can appear maladaptive. How often observed 'maladaptations,' which may easily include species that do not track or appear to track poorly, are actually the outcome of bet-hedging is difficult to estimate, as robustly assessing bet-hedging requires studies of fitness over longer timescales than many current field experiments (Simons, 2011). Environmental variation often includes both predictable and less predictable aspects. In such cases theory predicts organisms may evolve tracking that is a mixed strategy between bet-hedging and plasticity (Wong & Ackerly, 2005).

Evolutionary theory, which integrates over the sometimes hidden costs and benefits of particular cue systems and considers environmental predictability, thus provide multiple reasons species may not track or track weakly. This suggests that—at least in stationary systems—we should expect a number of species that do not track.

Predicting variation in environmental tracking in non-stationary systems

A major open area of research is adapting life history theory to the non-stationary environments introduced by climate change. Critical to predictions is whether cue systems maintain

their reliability with change; i.e., whether they continue to yield high fundamental tracking (Bonamour *et al.*, 2019). Consider a simple case in which an organism’s cues evolved based on a correlation between peak prey abundance and daylength: in a stationary environment the daylength cue may be fairly reliable (generally predicting peak prey abundance based on daylength, with some interannual variation), but would become unreliable, and lead to fitness declines, if warming continually advances peak prey abundance. Multivariate cues are often argued to be more reliable because they can capture multiple attributes of the environment (Dore *et al.*, 2018; Bonamour *et al.*, 2019), but they may be equally vulnerable to failure if non-stationarity decouples the cues from fundamental tracking (Bonamour *et al.*, 2019) and thus optimal fitness is no longer associated with the cue system. Predicting the outcome of non-stationarity from the stationary environment requires that researchers know: (1) the full cue system of an organism, (2) how it relates to fundamental tracking, and (3) how both that cue system and the underlying fundamental model shift with non-stationarity.

In recent years, plasticity theory has developed to provide insights on non-stationarity (or ‘sustained environmental change,’ see Chevin *et al.*, 2010). 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. At the individual level, environmental tracking is a plastic response, and thus we might predict greater individual tracking in non-stationary environments. This outcome, however, assumes there are no costs related to plasticity (Ghalambor *et al.*, 2007; Tufto, 2015). If there are costs associated with tracking, as discussed above in stationary systems, then species may evolve lower tracking (Auld *et al.*, 2010).

1.3 Tracking in multi-species environments

Life history theory often ignores other (non-focal) species or abstracts them as an aspect of the environment. While the trophic mis-match literature has addressed this gap for trophic interactions (Visser & Both, 2005; Visser & Gienapp, 2019), there is little consideration of competitive coexistence, yet this perspective is critical to understanding environmental tracking (Metcalf *et al.*, 2015). Considering how selection in multi-species environments is structured by competition highlights that tracking cannot be considered as a singular trait, but must be evaluated as part of a trait syndrome (or mosaic of traits, Ghalambor *et al.*, 2007) and should ultimately produce communities of species where tracking trades-off with other traits.

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) and thus may need to be more stress tolerant. 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 (Heberling *et al.*, 2019) or species with shallow roots starting growth sooner in an

alpine meadow system, while species with deeper roots begin growth later (Zhu *et al.*, 2016). In such cases, tracking is akin to a competition-colonization trade-off (Amarasekare, 2003), 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 (Dorji *et al.*, 2013; Lasky *et al.*, 2016; Zhu *et al.*, 2016). Further, many studies have found a correlation between higher tracking and ‘early-ness’ each season, which has been linked to resource acquisition traits associated with lower competitive abilities (Wolkovich & Cleland, 2014, see Box ‘Trait trade-offs with tracking’).

Understanding these trade-offs is clearly critical, but the short-term dynamics of a changing environment with plastic species is additionally important and highlights how little ecological theory we have for tracking. While evolutionary theory sometimes predicts the fitness outcomes of a new environment, non-stationarity in the climate today means understanding the trajectory may be most relevant—and bridges across evolutionary and ecological timescales. Evolutionary models show how plasticity may limit standing variation and thus reduce fitness in novel environments (Ghalambor *et al.*, 2007; Fournier-Level *et al.*, 2016; Fox *et al.*, 2019). But such findings may not hold if ecological dynamics reshape the environment as systems transition from stationary to non-stationary.

Including tracking in multi-species community assembly models

Predicting how tracking may determine which species are winners and losers with climate change requires integrating non-stationary environments into models of community assembly. Recent advances in coexistence models, sometimes called ‘modern coexistence theory,’ recognize that both mechanisms independent of fluctuations in the environment (e.g., R^* and other classical niche differences) and mechanisms dependent on fluctuations in the environment (relative non-linearity and storage effect) can lead to coexistence (Chesson & Huntly, 1997; Chesson, 2000). These models, which underlie much of current community ecology research (Mayfield & Levine, 2010; Barabas *et al.*, 2018; Ellner *et al.*, 2019), provide a framework to begin to integrate tracking and non-stationarity into community ecology theory.

How the environment is defined in most community models falls into two broad categories. In some models the environment is expressed as variation in species’ parameters. For example, in an early formalization of the lottery model (Chesson & Warner, 1981), the environment appears as interannual variation in birth and death rates. In later generalizations of competitive coexistence in temporally-varying environments, including the storage effect model (Chesson & Huntly, 1997), the environment is formalized as the ‘species response to the environment’ (E_i), where environment variation is translated into the common currency of species’ low density per capita growth rates. Building a changing environment into such models thus requires knowing how environmental shifts filter through to species-level parameters (Tuljapurkar *et al.*, 2009) to impact fundamental tracking. For example, storage effect models predict shifts in communities when environmental change alters the long-term covariance between the environment and competition (i.e., decreasing $cov(E_i, C_i)$), leading to a decrease in the storage effect as a means of competitive coexistence. In another example, Rudolf (2019) added the temporal environment to competition models by defining interaction strength as dependent on the temporal distance between species. This is somewhat similar to models that include the environment effectively

through different levels of asynchrony (e.g., Nakazawa & Doi, 2012; Revilla *et al.*, 2014).

In other models, the environment is more specifically defined as a resource (e.g., seed germination models where an explicit resource pulse each year initiates germination) and models something close to fundamental tracking. Models that explicitly include the environment provide a major opportunity to predict how tracking and non-stationarity determine future communities. As an example, we modeled a shift to earlier growing seasons using a common coexistence model where the environment is defined as a limiting resource that determines the start of growth each year. The timing of the resource relative to each species' ideal timing determines the species-specific germination fraction each year, allowing us to include fundamental tracking. The shift to earlier seasons favored species that could track and narrowed the region of coexistence maintained by a trade-off between tracking and competitive ability (via R^* , see Fig. 3 and Box: 'Adding tracking and non-stationarity to a common coexistence model'). Like all models, it rests on a number of assumptions, including that species' cues remain as reliable in the non-stationary environment, but shows how non-stationarity could benefit trackers.

Most current models (including the previous example) examine the environment from only one of two relevant perspectives: they represent the environment through its effects on fitness (e.g., the storage effect model), or they represent the environment as used for species' cues (e.g., many models of plasticity). Combining these two perspectives, which connect to fundamental and environmental tracking (respectively), may be especially critical to understanding the costs, benefits and community outcomes of tracking in non-stationary environments.

Layered onto these different views of the environment is how species responses are defined. In general, species responses to the environment can be broadly grouped into models that explicitly define when species start an event (e.g., spawning or germination) versus those that model the magnitude of response (e.g., the number of propagules or seeds, as in Box: 'Adding tracking and non-stationarity to a common coexistence model'). Models that explicitly include when a species starts an event are often focused on situations where order of arrival is critical. For example, models of priority effects through niche pre-emption highlight the advantage tracking may provide when it allows species to be early: early arrivals receive a head-start advantage, by gaining priority access to resources (the environment) they can draw down the resources available to later arrivals (Fukami, 2015). Such models predict early-arriving species to out-compete other species, unless there is a cost to being too early or there are trade-offs with other species' traits (see Fig. 4).

Other models canalize species' responses to the environment into production and investment. Most models of inter-annual competition (most explicit examples of 'modern coexistence theory,' e.g., Chesson *et al.*, 2004; Angert *et al.*, 2009) fall into this camp. Species produce (via investment in offspring, tissue etc.) differentially depending on the environment each year and outcomes are mediated through density. While these models superficially may seem disconnected from timing, they highlight how event timing often relates to production and, thus, investment across years. Further, they almost always model the environment as a distribution (see Fig. 4), which provides the opportunity for the environment to alter the competitive environment each year and, thus, structure coexistence.

A model where species vary both when they start an event and how much they invest would

capture the important attributes of fundamental tracking—combining head-start advantages from being early with production variation based on the resource environment. To our knowledge, however, most models approach these questions separately, though models of bet-hedging come closest (Gourbiere & Menu, 2009; Tufto, 2015). A model that explicitly includes the linked decisions of when to time an event and how much offspring/tissue to produce during the event could provide insights on the relative importance of each aspect of this process. Such a model could be adapted to address multiple questions of tracking, including how these decisions (‘when’ and ‘how much’) may trade-off and which other traits may be most strongly linked to tracking, as well as explicitly modeling the costs and benefits of tracking in stationary systems—a critical precursor to extending it to non-stationary systems.

Extending models to non-stationary systems is crucial to testing how environmental tracking relates to fundamental tracking and species persistence with climate change, and research has already begun to tackle this non-trivial challenge (Chesson, 2017; Legault & Melbourne, 2019; Rudolf, 2019). Most work to date, however, focuses on conclusions from systems that are initialized as non-stationary, ignoring the transition between stationary and non-stationary environments. Yet we expect this transition may be critical because communities formed in stationary environments (or periods with lower non-stationarity) are effectively filtered and assembled by that environmental regime and thus produce the baseline of variation and assembly dynamics for a shifting environment. While analytical solutions for systems transitioning from stationary to non-stationary may take time to develop (Chesson, 2017), simulation work can provide an immediate intuition and framework to address this challenge (for an example, see Box: Adding tracking and non-stationarity to a common coexistence model).

Outcomes for such community assembly models also depend on how effectively closed communities are. Dispersal of species or individuals with traits that make them better matched to the non-stationary environment would lead to new communities that may persist or be continually re-assembled as long as the environment remains non-stationary. Indeed, this logic underlies the argument that invasive species may be superior trackers benefiting from how climate change has altered growing seasons (Willis *et al.*, 2010; Wolkovich & Cleland, 2011). Evolutionary responses could also rescue species with low plasticity. Long-term population (e.g., Colautti *et al.*, 2017) and resurrection studies (Wilczek *et al.*, 2014; Yousey *et al.*, 2018), as well as field experiments (Colautti *et al.*, 2017; Exposito-Alonso *et al.*, 2019), have repeatedly shown species can shift to earlier flowering times, higher thermal tolerances or related genetically-controlled traits that confer higher fitness in warmer climates. Yet these studies also highlight that responses can be lagged (e.g., Wilczek *et al.*, 2014), associated with reduced population viability (e.g., Colautti *et al.*, 2017), and that other factors may constrain adaptive responses.

1.4 Future directions

Growing empirical research highlights that environmental tracking is linked to species performance and, thus, may be critical to understanding the forces that assemble communities and determine species persistence, especially as anthropogenic climate change reshapes the environment. We have outlined above how current community ecology theory could make advances

through models that combine effects of variation in timing and production amounts and models that include the environment as impacting species' cues, as well as species' fitness. Such models would explicitly allow the potential costs and benefits of tracking depending on how closely environmental tracking matches fundamental tracking. But to best test and develop such models we need a greater understanding of how the environment is changing, more robust estimates of environmental tracking and how it fits within a mosaic of correlated traits that determine individual fitness.

How is an organism's environment changing?

Currently, much research has focused on one major shift in the climate system (rising temperatures), but research on multivariate environmental shifts is growing and will be critical to understanding how climate change affects an organism's whole environment (e.g., Chevin & Lande, 2015). We can help guide these efforts by identifying environmental shifts that are often linked (e.g., Wadgymar *et al.*, 2018). For example, warming temperatures may drive earlier seasons and higher evaporative water loss. Researchers can also aim to more consistently and fully characterize the environmental distributions of their systems that appear to drive species performance and interactions: the environment of the years of study should be clearly reported and compared against long-term and recent climate for each system.

More interdisciplinary research with climate science could also speed a fuller understanding of what shifts are and are not expected with climate change, and what climate variables are inherently correlated. Such correlations make estimating cues and other biological parameters from long-term data especially precarious (Tansey *et al.*, 2017). But these correlations are equally critical in considering how species may view their environment and whether environmental change will couple or uncouple links between proximate cues and fundamental tracking (Bonamour *et al.*, 2019).

Understanding and measuring 'tracking'

Understanding how the environment is changing represents just one step towards robust measures of environmental tracking. Shifting environmental regimes must then be filtered through species cues, and impacts on growth and survival. Studies should clarify their definition of tracking, how the environment is defined, how an event relates to fitness, and how well—or not—the underlying cue system is understood (see Box: 'Challenges & opportunities in measuring tracking'). Currently, many studies examine fundamental and environmental tracking simultaneously (e.g., Visser *et al.*, 2006; Charmantier *et al.*, 2008; Cleland *et al.*, 2012; Yang & Censer, 2020), 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 results across studies. We expect progress will come from a balance between measures of fundamental tracking (measuring both event date variation and fitness), estimating an organism's system of cues (generally through controlled experiments followed by tests in the field), and measuring the change in an event date relative to environmental variation that is due to cues (environmental tracking). Clear statements of what is known, not known and what is measured will help.

What major traits trade-off with tracking?

Basic theory of plasticity and competition suggest that environmental tracking must trade-off with other traits to allow multi-species communities. Yet to date empirical work has mainly documented tracking, linked it to performance, or focused on how it varies between native and non-native species (Willis *et al.*, 2010; Wolkovich *et al.*, 2013; Zettlemoyer *et al.*, 2019). Research has highlighted some traits that co-vary with tracking (e.g., Kharouba *et al.*, 2014; Lasky *et al.*, 2016; Zhu *et al.*, 2016), but to tie this work to models requires more research on traits that link clearly to theory, and a fuller understanding of how tracking and other traits jointly contribute to performance under varying environments.

Traits that link to resource competition, as we focus on here (as others have as well, see Rudolf, 2019), may be especially fruitful for greater research, but should not be the only ones considered. For example, traits related to stress, predator tolerance and avoidance may also play a role, but have been effectively unstudied. As empirical research in this area grows, models can aid progress in understanding the outcomes of these trade-offs for community assembly.

1.5 Stationarity in the future

While most environments today are climatically non-stationary and have been for decades, the climate will return to a more stationarity form in the future. There are many possible pathways to climatic stabilization, but almost all require first the stabilization of greenhouse gases—the subject of much policy and political debate. Once greenhouse gas emissions stabilize climate will not quickly snap back to a new stationary phase. Instead systems will slowly approach a new climatic stationarity depending on how they are effected by the earth’s multiple thermal reservoirs, and, in turn, how quickly those reservoirs stabilize. The timescale of this approach is generally expected to be on the scale of centuries, but could be much longer in certain oceanic systems (Collins *et al.*, 2013). Thus, ecologists are—and will remain for the foreseeable future—in a research area structured by climatic non-stationarity.

As paleobiologists and evolutionary biologists often point out, climatic nonstationarity is a common part of the earth’s history (Jansson & Dynesius, 2002)—even if stationary periods—be they cold or warm (glacial and interglacial periods), or dry or wet (e.g., megadroughts or pluvials)—are more common. Indeed, while much of this work has examined how species survive for millions of years given large oscillations in climate (Provan & Bennett, 2008), the periods that provide the most dramatic community reshuffling are periods shifting from stationary to non-stationary climate regimes (Vrba, 1980, 1985). Such stories of the past are now fundamentally happening today, and ecology is challenged to understand how transitions between stationary and non-stationary environments are reshaping the species and communities we have today and will in the altered climates of our future.

2 Acknowledgments

We thank I. Breckheimer, D. Buonaiuto, E. Cleland, J. Davies, G. Legault, A. Phillimore and three anonymous reviewers for helpful comments that improved the manuscript, K. Slimon for

423 help with trade-offs review, and the NSERC Discovery Award (RGPIN-05038 to EMW) and
424 Canada Research Chair in Temporal Ecology (EMW) programs that provided funding.

3 Boxes

3.1 Box: Environmental variability & change

Decades of ecological research highlight how temporally variable environments shape species and their communities at multiple scales (Sale, 1977; Chesson & Huntly, 1997). In seasonal landscapes, the environment limits periods for growth each year (e.g., by temperature, snowpack or drought); within-year variability in the environment (e.g., daily, hourly or finer resolution temperatures or rainfall amounts) compounds into inter-annual variability that shapes the distribution of the start and end of growing seasons. For long stretches of history this variability has been effectively stationary; that is, the underlying probability distribution that describes the start (or end) of the season (e.g., the date of the last major frost) does not change, even though the date may be dramatically different from one year to the next.

In other time periods, variability has been non-stationary in one or multiple dimensions. For example, climate in the northern hemisphere includes long warming and then cooling periods (i.e., increasing then decreasing means of the probability distribution) at the start of the Holocene, when the earth was coming out of the last glacial maximum. Anthropogenic climate change is a similar non-stationary process, with warming evident around the globe and knock-on effects for other climate metrics, such as heat extremes and the size of precipitation events.

Understanding non-stationarity in ecological systems requires first identifying which aspects of the environment have shifted—and how they have shifted with respect to one another—as the underlying distributions transition from stationary to non-stationary (Fig. 1). For example, with climate change, warming has increased mean temperatures over time, with minimum temperatures generally increasing more than maximum—this results in an underlying distribution for daily temperature where the mean is increasing through time while the within-day variance is decreasing (Stocker *et al.*, 2013; Screen, 2014). Additionally, climate change has decoupled historical relationships between precipitation and temperature in some systems (e.g., Cook & Wolkovich, 2016; Wadgyman *et al.*, 2018). 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.

3.2 Box: Challenges & opportunities in measuring tracking

Increasingly, research has outlined statistical difficulties in measuring tracking, which may underlie many observations of species that do not track. These issues mostly relate to the challenges of observing event distributions (Steer *et al.*, 2019; Carter *et al.*, 2018) using discrete (calendar) time, low power, and the complexity of climate data.

Our units of observation may be non-stationary—estimates of mean days shifted per decade depend on the climate of the decade(s) studied, which is not consistent in many systems (McCabe *et al.*, 2012). Estimates based on a relevant climate variable can sometimes ameliorate

this problem, but may be equally vulnerable to non-stationarity in units (e.g., Sagarin, 2001). For example, processes that depend on thermal sums reported as days/°C will generally appear to decline with warming, as the thermal sum of an average day has increased in most regions with climate change. Relatedly, estimates of long-term change using simple linear regression depend on the climate at the start of the time-series (with greater changes seen from time-series that started in unusually cold decades, such as the 1950s for much of North America).

Even ‘long’ time-series may be too short for robust analyses of trends (Bolmgren *et al.*, 2013). Authors should be especially cautious if they find only large effects appear significant (e.g., CaraDonna *et al.*, 2014), which is a well-known statistical bias associated with p-values (Loken & Gelman, 2017). Additionally, effect sizes that are higher when climate variability is higher (for example, in temperate habitats temperature is highly variable in the spring and autumn compared to summer) may be more related to variation in statistical power than to biology.

Many statistical issues can be addressed by improved statistical approaches (e.g., Gienapp *et al.*, 2005; Pearse *et al.*, 2017), though such approaches may uncomfortably highlight how uncertain many current estimates are (Brown *et al.*, 2016) or reveal lower effect sizes. Impacts of start-years for long-term time-series can be muted by applying change-point or hinge models (e.g., Kharouba *et al.*, 2018). We suggest mixed models should be used more widely alongside randomization and/or data-simulation approaches (e.g., Bolmgren *et al.*, 2013), and we need models that can discriminate among confounding factors. For example, growing evidence suggests a potential fundamental trade-off where early species track, grow fast and die young, while later species track less, grow slowly and live longer—this might suggest later species bet-hedge more given their longer investment window. Or it could be an artifact where early species use simpler cues, and, thus, their tracking is measured more accurately given current methods.

Even without statistical issues, translating event date and climate data into estimates of tracking requires a firm biological understanding of an organism’s cues, which we rarely have (Chmura *et al.*, 2019). Currently, ‘tracking’ is often measured as the relationship between the dates of an event and a simple abiotic metric. Such measures, however, are almost always proxies for a more complicated underlying physiology where simple cues—such as warm temperatures or snowpack—can be modified by other cues, such as photoperiod, drought or light spectra (Bagnall, 1993; Stinchcombe *et al.*, 2004). Teasing out these other cues requires nuanced approaches to observational data with explicit assumptions (Tansey *et al.*, 2017) or controlled experiments (Wilczek *et al.*, 2009; Caffarra *et al.*, 2011).

Modeling multivariate cues is inherently difficult (Chuine *et al.*, 2016), especially since one cue may dominate in many conditions. For example, woody plant leafout responds to warm spring temperatures, but also to cool winter temperatures to prevent leafout in mid-winter warm snaps. Often this cool-temperature effect may be masked by sufficiently cool conditions. With warming from climate change, however, this additional trigger may become critical (and potentially lead many phenological models to fail spectacularly in the future, see Chuine *et al.*, 2016). Tracking in species with longer generation times may be especially complicated, as species may track low frequency climate signals and make investment choices on far longer timescales than species with shorter lifespans (Morris *et al.*, 2008). In some semi-arid systems, species time growth to pulses of rain that occur with cooler temperatures, indicating the start of the rainy season, and

not a rare summer rainfall event in the middle of months of drought (Wainwright & Cleland, 2013).

Addressing these issues is possible if we embrace our inner physiologists, or collaborate with one, to develop models that explicitly include species' cues, and consider the framework under which we expect cue systems have evolved (e.g., bet-hedging). We then must interrogate these models to understand when they work and where they fail (see Johansson & Bolmgren, 2019, for an example). This approach can help embrace the contradictory pulls of conducting experiments to identify mechanistic cues and understanding how they are filtered through the multivariate climate of the real world (see Wilczek *et al.*, 2010, 2009).

3.3 Box: Trait trade-offs with tracking

Research on temporal tracking and traits has increased greatly in recent years, with a major uptick in studies after 2010 (see SI Fig. S1). Most papers examining tracking and other traits across species have focused on plants (20/30), followed by birds and Lepidoptera (both 4/30), plankton and aphids (both 1/30). The most studied trait was how early or late a phenophase occurred (e.g., date of flowering, or start of migration for a species, termed 'earlyness' by some authors), with earlier species tending to track more (studies included both birds and Lepidoptera, Diamond *et al.*, 2011; Ishioka *et al.*, 2013; Kharouba *et al.*, 2014; Jing *et al.*, 2016; Du *et al.*, 2017). While this is an important link, it is vulnerable to statistical challenges (see Box 'Challenges & opportunities in measuring tracking'). Few studies examined whether tracking correlates with resource acquisition traits; those that did generally found species with higher tracking also had traits associated with lower competitive abilities under low resources (e.g., being shallower rooted or lacking a taproot, Dorji *et al.*, 2013; Lasky *et al.*, 2016; Zhu *et al.*, 2016). These species were often also early (e.g., Dorji *et al.*, 2013; Zhu *et al.*, 2016), suggesting tracking may relate to a syndrome of traits that allow species to be rapid colonizers each season, but poor competitors for resources. Indeed, previous work has documented that species with earlier phenophases tend to have resource acquisition traits associated with lower competitive abilities (e.g., they tend to be of lower height, have shallower roots, narrower diameter vessels, thinner leaves, and grow faster, reviewed in Wolkovich & Cleland, 2014).

3.4 Box: Adding tracking and non-stationarity to a common coexistence model

To understand the role of environmental tracking by species in variable environments we use a simple model that allows within- and between-year dynamics to contribute to coexistence. As the model is akin to many commonly used seed germination models (Chesson *et al.*, 2004), we follow a similar terminology for ease; however the basic structure of our model could apply to other systems with one dominant (non-renewing) pulse of a limiting resource each season (e.g., water from rain or snowpack), or a multivariate resource pulse that acts effectively as one resource (e.g., nitrogen and light drawn down together over the season). In this model the environment is included between-years via variable germination, and within-years the environment is explicitly included as a resource pulse at the start of the season. We adjust the biologi-

cal start time of species (τ_i for species i) to also allow species to respond to the environment dynamically through what we refer to as tracking. Here, tracking effectively moves a species intrinsic start time closer to the environmental start time in that year, resulting in a higher germination fraction (making it effectively a superior colonizer, see SI for complete description and equations).

Species can co-occur via equalizing mechanisms, but they require stabilizing mechanisms to coexist. Species that is, cannot coexist given only variation in tracking—coexistence requires variation in another trait axis. As theory and empirical work suggest this trade-off may involve traits related closely to resource competition, we varied species' R^* . With variation in tracking and in R^* species can persist together as long as those species with a temporal niche advantage are also the inferior competitors (Fig. 3). That is, species that can draw resources down to a lower level and are thus the superior within-season resource competitors (lower R^*) can persist with species with that are inferior competitors but have realized biological start times closer to the environmental start time—a finding inline with currently observed empirical trade-offs (see Box ‘Trait trade-offs with tracking’). These trade-offs, however, are all environmentally dependent. They hold only so long as the environment is stationary.

We examined how trade-offs may be transformed by a non-stationary environment, by transitioning a stationary environment—in which two-species communities had persisted for 500 years—to non-stationary, via an earlier start of season (earlier timing of the resource pulse, τ_p , Fig. 3; see SI for more details). By changing a fundamental niche axis (the distribution of the environment, an axis along which these communities were structured), we shifted one major part of the trade-off: the new non-stationary environment favored an earlier start time than the previous stationary environment. This, in turn, reshaped our two-species communities, which depended on this trade-off for persistence.

While the non-stationary environment favored higher trackers (who in turn drove the extinction of species with lower tracking values from many two-species communities) some two-species communities persisted (257 out of 1698 two-species communities persisting after end of stationary, or 15.1%, Fig. 3). These two-species communities persisted because the same fundamental trade-off between biological start time and within-season competitive ability, while narrowed, was not fully lost. Taken together, these simulations show how non-stationarity can drive local species extinction and reshape the underlying assembly mechanisms of communities.

Our simulations support growing work that tracking should not be considered alone (Diamond *et al.*, 2011; Dorji *et al.*, 2013; Ishioka *et al.*, 2013; Kharouba *et al.*, 2014; Du *et al.*, 2017), but may be part of a larger trait syndrome. Indeed, this model trivially shows that multi-species communities cannot form given only variation in the temporal niche—a trade-off is required. Our results thus support empirical work showing a trade-off where trackers are also inferior resource competitors (Lasky *et al.*, 2016; Zhu *et al.*, 2016)—we show this must be the case for multi-species persistence; otherwise, the species best matched to the environment would drive the other extinct.

Finally, our results highlight that non-stationarity may reshape the balance of equalizing versus stabilizing mechanisms. As environments shifted from stationarity to non-stationarity, species that co-occurred via equalizing mechanisms persisted longer. While the outcome that equalized

species will be more similarly affected by environmental shifts is rather obvious, it has several important implications. First, it may make identifying which traits climate change promotes through stabilizing mechanisms more difficult. Second, it suggests climate change—or other factors that cause an environment to shift from stationary to non-stationary—may cause a fundamental shift away from assembly via stabilizing mechanisms.

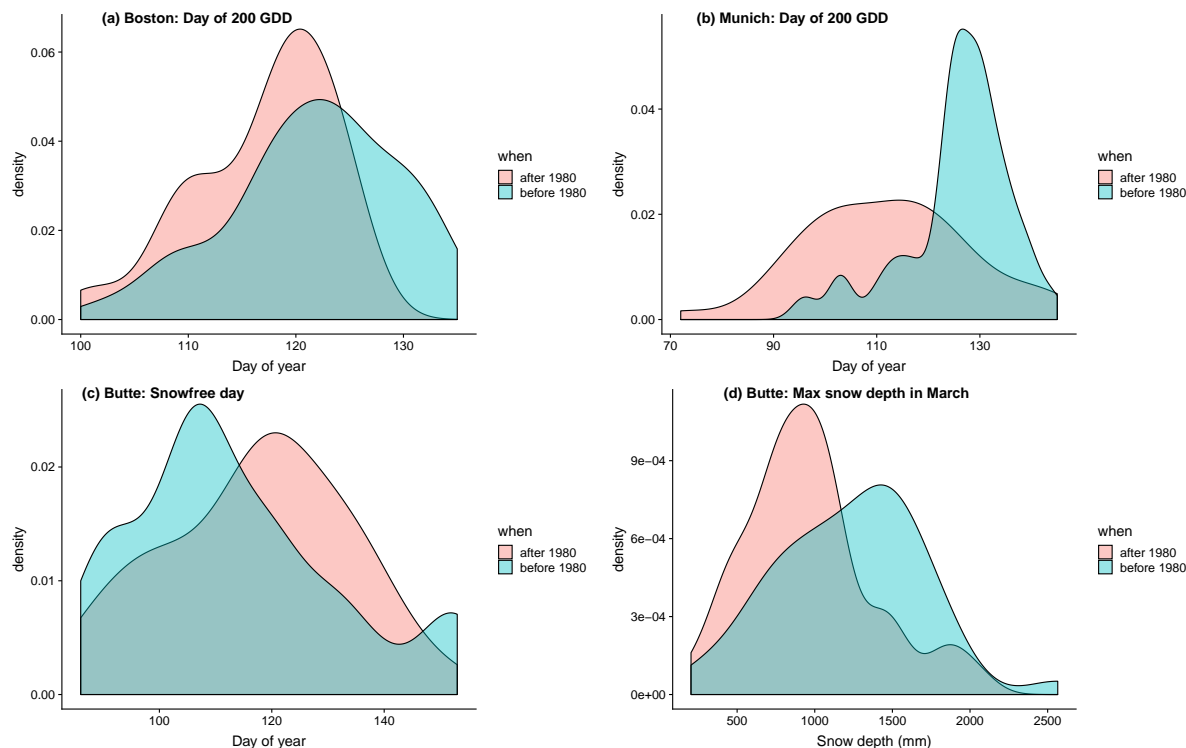


Figure 1: Examples of non-stationarity in climate variables linked to environmental tracking: shifts before and after 1980 (a major change-point in climate for many regions) in several metrics related to the start of growing seasons (a-c) or resource pulse connected to growing season length (d). Density plots of day of 200 growing degree day units (a metric of thermal sum, here based on 0 degree base temperature using daily minima in $^{\circ}\text{C}$) in Boston, MA, USA (a), and Munich, Germany (b), first snowfree day (followed by at least 9 snowfree days) in Crested Butte, CO, USA (c) and maximum snowdepth (mm) in March (often the month before the first snowfree day) in Crested Butte, CO, USA (d). Note that (c) and (d) are likely related, with lower snowpacks leading to an earlier first snowfree day. We selected sites that have been studied for plant phenological data and included at least 80 years of daily climate data from a Global Historical Climatology Network site; we subsetting data so that there were 40 years before and after 1980 for all sites.

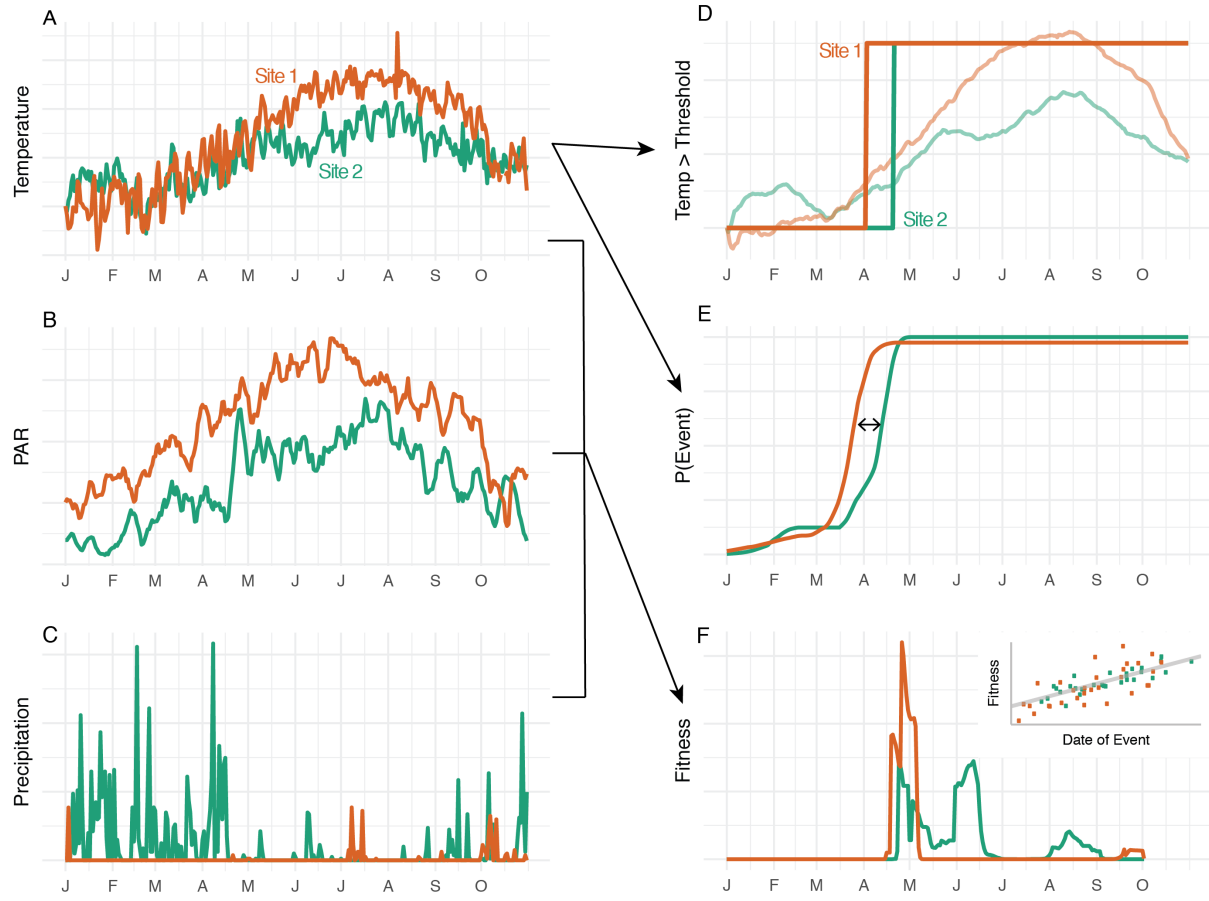


Figure 2: Tracking: Different components of a multivariate environment (A: air temperature, B: cumulative photosynthetically active radiation, C: precipitation) are included in measurements of the environment to understand the timing of an event (D), in an organism’s cueing from the environment which triggers the event (E), and in an organism’s fitness outcome that results from the event (F), conceptualized here for two sites. While we may often use simple measurements to estimate tracking (D: measured environmental tracking modeled as a temperature threshold), an organism’s cue system may be more complex (E: environmental tracking, here shown as a multivariate cue system where the changing probability of the event is a function of accumulated degree chilling hours and degree heating hours, based on the temperature environment in A). Ultimately, fitness is determined by the joint distribution of many environmental variables through time (F: fundamental tracking, shown here as the end-of-season fitness as a function of the timing a start-of-season life history event, where we model fitness dependent on (1) growth as a unimodal function of temperature and a saturating function of PAR and soil moisture, and (2) survival as a function of early season frost and summer heat stress). Of ultimate value to the organism—especially under a changing climate—is cue reliability (inset in F): the correlation between the timing of the event (E) and the fitness of the organism (F) across years. See SI ‘Fig. 2 methods’ for further methods and details.

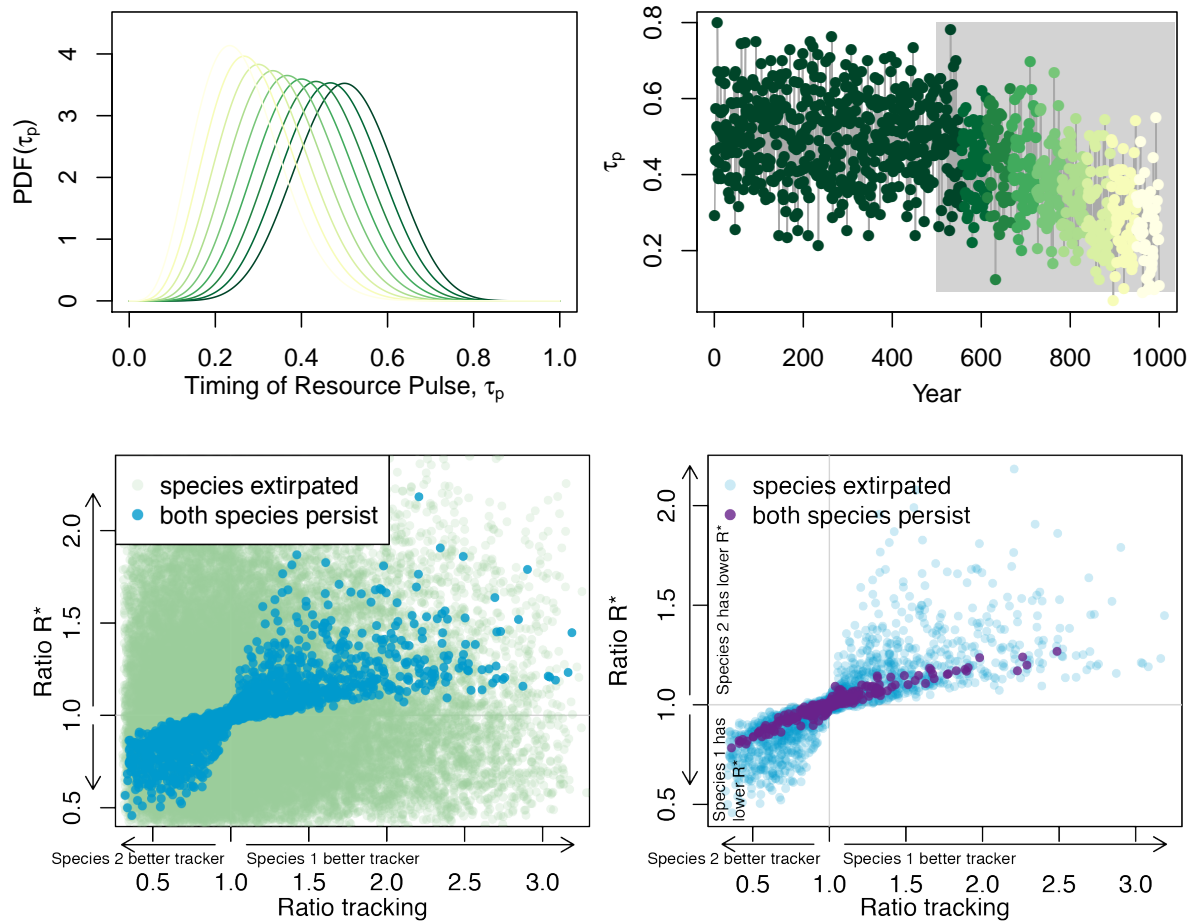


Figure 3: Example of how non-stationarity can reshape communities in a simple model. We shifted the environment (top panels) by changing the timing of the resource pulse from a stationary period ($\tau_p \sim \beta(10, 10)$ for the 500 years) to a nonstationary period ($\tau_p \sim \beta(5, 15)$ over the 500 years), then examined outcomes for two-species communities (bottom panels) where tracking (X axis: species 1/species 2) trades off with R^* (Y axis: species 1/species 2): each point represents one two-species community color-coded by whether both species persisted or one or more species was extirpated through 500 years of a stationary environment (bottom-left), followed by an additional 500 years of non-stationary environment (bottom-right, only two-species communities that persisted through the stationary period are shown).

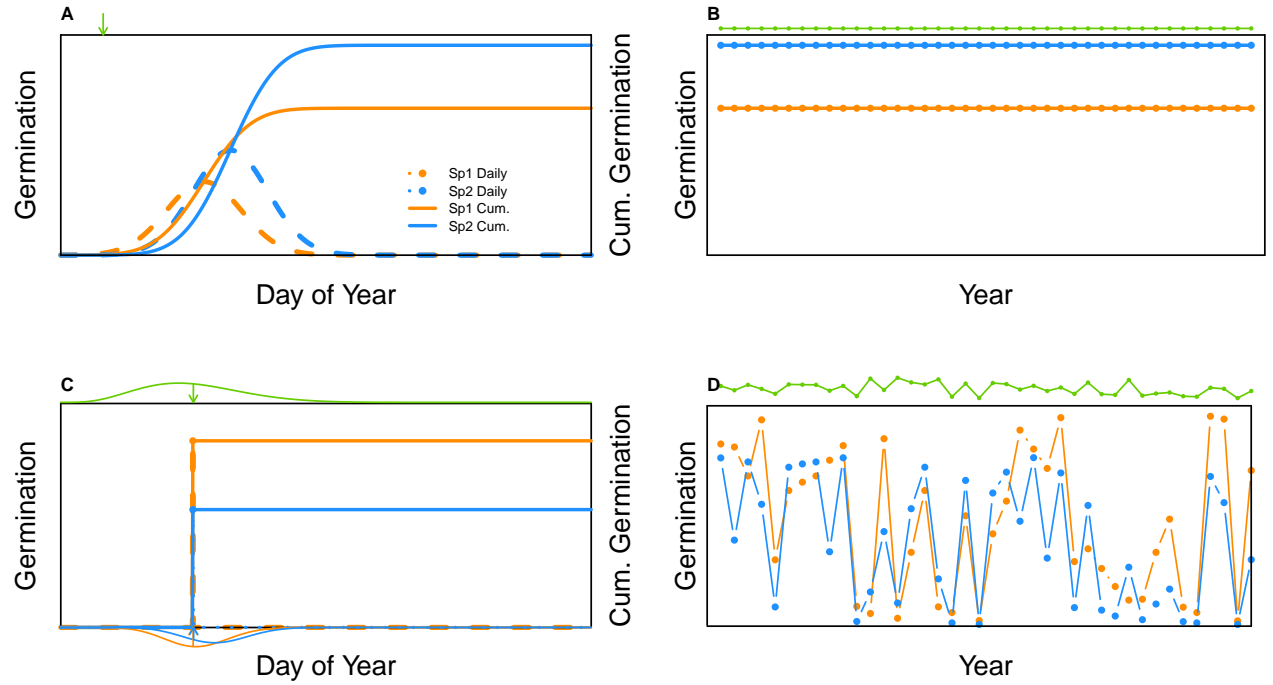


Figure 4: Tracking can be conceptualized as changes in priority effects or changes in storage effects. In a priority effect model (A,B), the coexistence mechanism is a within-year tradeoff between, an early-germinating species that pre-empts resources (sp 1) and late-germinating species that is a superior resource competitor (sp 2) (A, where green arrow indicates the start of season); no between-year variation is required to maintain coexistence. In a storage effect model (C,D), variation in the timing of the start of season (indicated by the distribution in green, top of C) results in differential species-response to the environment (illustrated by species-specific germination curves, bottom of C); this interannual variation in species-response to the environment (D)—along with a seedbank or other interannual storage mechanism—can maintain coexistence through reduced interspecific competition.

References

- Ackerly, D. (2009). Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19699–19706.
- Amano, T., Freckleton, R.P., Queenborough, S.A., Doxford, S.W., Smithers, R.J., Sparks, T.H. & Sutherland, W.J. (2014). Links between plant species' spatial and temporal responses to a warming climate. *Proceedings of the Royal Society B-Biological Sciences*, 281.
- Amarasekare, P. (2003). Competitive coexistence in spatially structured environments: a synthesis. *Ecology Letters*, 6.
- Angert, A.L., Huxman, T.E., Chesson, P. & Venable, D.L. (2009). Functional tradeoffs determine species coexistence via the storage effect. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 11641–11645.
- van Asch, M., Tienderen, P.H., Holleman, L.J.M. & Visser, M.E. (2007). Predicting adaptation of phenology in response to climate change, an insect herbivore example. *Global Change Biology*, 13, 1596–1604.
- Auld, J.R., Agrawal, A.A. & Relyea, R.A. (2010). Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society B-Biological Sciences*, 277, 503–511.
- Bagnall, D.J. (1993). Light quality and vernalization interact in controlling late flowering *Arabidopsis* ecotypes and mutants. *Annals of Botany*, 71, 75–83.
- Barabas, G., D'Andrea, R. & Stump, S.M. (2018). Chesson's coexistence theory. *Ecological Monographs*, 88, 277–303.
- Block, S., Alexander, J. & Levine, J.M. (2019). Phenological plasticity is a poor predictor of subalpine plant population performance following experimental climate change. *Oikos*, p. 10.1111/oik.06667.
- Bolmgren, K., Vanhoenacker, D. & Miller-Rushing, A.J. (2013). One man, 73 years, and 25 species. evaluating phenological responses using a lifelong study of first flowering dates. *International Journal of Biometeorology*, 57, 367–375.
- Bonamour, S., Chevin, L.M., Charmantier, A. & Teplitsky, C. (2019). Phenotypic plasticity in response to climate change: the importance of cue variation. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 374.
- Brown, C.J., O'Connor, M.I., Poloczanska, E.S., Schoeman, D.S., Buckley, L.B., Burrows, M.T., Duarte, C.M., Halpern, B.S., Pandolfi, J.M., Parmesan, C. & Richardson, A.J. (2016). Ecological and methodological drivers of species' distribution and phenology responses to climate change. *Global Change Biology*, 22, 1548–1560.
- Caffarra, A., Donnelly, A. & Chuine, I. (2011). Modelling the timing of *Betula pubescens* budburst. II. Integrating complex effects of photoperiod into process-based models. *Climate Research*, 46, 159–170.

- CaraDonna, P.J., Iler, A.M. & Inouye, D.W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences*, 111, 4916–4921.
- Carter, S.K., Saenz, D. & Rudolf, V.H.W. (2018). Shifts in phenological distributions reshape interaction potential in natural communities. *Ecology Letters*, 21, 1143–1151.
- de Casas, R.R., Donohue, K., Venable, D.L. & Cheptou, P.O. (2015). Gene-flow through space and time: dispersal, dormancy and adaptation to changing environments. *Evolutionary Ecology*, 29, 813–831.
- Charmantier, A., McCleery, R.H., Cole, L.R., Perrins, C., Kruuk, L.E.B. & Sheldon, B.C. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, 320, 800–803.
- Chen, I.C., Hill, J.K., Ohlemuller, R., Roy, D.B. & Thomas, C.D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, 333, 1024–1026.
- Chesson, P. (2000). Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, 31, 343–366.
- Chesson, P. (2017). Aedt: A new concept for ecological dynamics in the ever-changing world. *Plos Biology*, 15.
- Chesson, P., Gebauer, R.L.E., Schwinning, S., Huntly, N., Wiegand, K., Ernest, M.S.K., Sher, A., Novoplansky, A. & Weltzin, J.F. (2004). Resource pulses, species interactions, and diversity maintenance in arid and semi-arid environments. *Oecologia*, 141, 236–253.
- Chesson, P. & Huntly, N. (1997). The roles of harsh and fluctuating conditions in the dynamics of ecological communities. *American Naturalist*, 150, 519–553.
- Chesson, P.L. & Warner, R.R. (1981). Environmental variability promotes coexistence in lottery competitive-systems. *American Naturalist*, 117, 923–943.
- Chevin, L.M. & Lande, R. (2015). Evolution of environmental cues for phenotypic plasticity. *Evolution*, 69, 2767–2775.
- Chevin, L.M., Lande, R. & Mace, G.M. (2010). Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *Plos Biology*, 8.
- Chmura, H.E., Kharouba, H.M., Ashander, J., Ehlman, S.M., Rivest, E.B. & Yang, L.H. (2019). The mechanisms of phenology: the patterns and processes of phenological shifts. *Ecological Monographs*, 89.
- Chuine, I., Bonhomme, M., Legave, J.M., de Cortazar-Atauri, I.G., Charrier, G., Lacoite, A. & Ameglio, T. (2016). Can phenological models predict tree phenology accurately in the future? the unrevealed hurdle of endodormancy break. *Global Change Biology*, 22, 3444–3460.
- Chuine, I. & Regniere, J. (2017). Process-based models of phenology for plants and animals. *Annual Review of Ecology, Evolution, and Systematics*, 48, 159–182.

- Cleland, E.E., Allen, J.M., Crimmins, T.M., Dunne, J.A., Pau, S., Travers, S.E., Zavaleta, E.S. & Wolkovich, E.M. (2012). Phenological tracking enables positive species responses to climate change. *Ecology*, 93, 1765–1771.
- Cohen, J.M., Lajeunesse, M.J. & Rohr, J.R. (2018). A global synthesis of animal phenological responses to climate change. *Nature Climate Change*, 8, 224–+.
- Colautti, R.I., Arington, J. & Anderson, J.T. (2017). Phenological shifts of native and invasive species under climate change: insights from the boechera - lythrum model. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 372.
- Collins, M., Knutti, R., Arblaster, J.L., Dufresne, T., Fichet, P., Friedlingstein, X., Gao, W., Gutowski, T., Johns, G., Krinner, M., Shongwe, C., Tebaldi, A., Weaver & Wehner, M. (2013). *Climate Change 2013: The Physical Science Basis – Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, chap. Long-term Climate Change: Projections, Commitments and Irreversibility, pp. 1029–1119.
- Cook, B.I. & Wolkovich, E.M. (2016). Climate change decouples drought from early wine grape harvests in France. *Nature Climate Change*, 6, 715–719.
- Cook, B.I., Wolkovich, E.M. & Parmesan, C. (2012). Divergent responses to spring and winter warming drive community level flowering trends. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 9000–9005.
- Craine, J.M., Wolkovich, E.M., Towne, E.G. & Kembel, S.W. (2012). Flowering phenology as a functional trait in a tallgrass prairie. *New Phytologist*, 193, 673–682.
- Deacy, W.W., Erlenbach, J.A., Leacock, W.B., Stanford, J.A., Robbins, C.T. & Armstrong, J.B. (2018). Phenological tracking associated with increased salmon consumption by brown bears. *Scientific Reports*, 8.
- Diamond, S.E., Frame, A.M., Martin, R.A. & Buckley, L.B. (2011). Species’ traits predict phenological responses to climate change in butterflies. *Ecology*, 92, 1005–1012.
- Donahue, M.J., Karnauskas, M., Toews, C. & Paris, C.B. (2015). Location isn’t everything: Timing of spawning aggregations optimizes larval replenishment. *Plos One*, 10.
- Donaldson-Matasci, M.C., Bergstrom, C.T. & Lachmann, M. (2012). When unreliable cues are good enough. *American Naturalist*, 182, 313–327.
- Dore, A.A., McDowall, L., Rouse, J., Bretman, A., Gage, M.J.G. & Chapman, T. (2018). The role of complex cues in social and reproductive plasticity. *Behavioral Ecology and Sociobiology*, 72.
- Dorji, T., Totland, O., Moe, S.R., Hopping, K.A., Pan, J. & Klein, J.A. (2013). Plant functional traits mediate reproductive phenology and success in response to experimental warming and snow addition in Tibet. *Global Change Biology*, 19, 459–472.

- Du, Y.J., Chen, J.R., Willis, C.G., Zhou, Z.Q., Liu, T., Dai, W.J., Zhao, Y. & Ma, K.P. (2017). Phylogenetic conservatism and trait correlates of spring phenological responses to climate change in northeast china. *Ecology and Evolution*, 7, 6747–6757.
- Ellner, S.P., Snyder, R.E., Adler, P.B. & Hooker, G. (2019). An expanded modern coexistence theory for empirical applications. *Ecology Letters*, 22, 3–18.
- Exposito-Alonso, M., Burbano, H.A., Bossdorf, O., Nielsen, R., Weigel, D., Exposito-Alonso, M., Rodriguez, R.G., Barragan, C., Capovilla, G., Chae, E., Devos, J., Dogan, E.S., Friedemann, C., Gross, C., Lang, P., Lundberg, D., Mitterdorff, V., Kageyama, J., Karasov, T., Kersten, S., Petersen, S., Rabbani, L., Regalado, J., Reinelt, L., Rowan, B., Seymour, D.K., Symeonidi, E., Schwab, R., Tran, D.T.N., Venkataramani, K., Van de Weyer, A.L., Vasseur, F., Wang, G., Wedegartner, R., Weiss, F., Wu, R., Xi, W.Y., Zaidem, M., Zhu, W.S., Garcia-Arenal, F., Burbano, H.A., Bossdorf, O., Weigel, D. & Genomes Field Expt, T. (2019). Natural selection on the arabidopsis thaliana genome in present and future climates. *Nature*, 573, 126–+.
- Fournier-Level, A., Perry, E.O., Wang, J.A., Braun, P.T., Migneault, A., Cooper, M.D., Metcalf, C.J.E. & Schmitt, J. (2016). Predicting the evolutionary dynamics of seasonal adaptation to novel climates in arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E2812–E2821.
- Fox, R.J., Donelson, J.M., Schunter, C., Ravasi, T. & Gaitan-Espitia, J.D. (2019). Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 374.
- Franks, S.J., Sim, S. & Weis, A.E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 1278–1282.
- Fu, Y.S.H., Zhao, H.F., Piao, S.L., Peaucelle, M., Peng, S.S., Zhou, G.Y., Ciais, P., Huang, M.T., Menzel, A., Uelas, J.P., Song, Y., Vitasse, Y., Zeng, Z.Z. & Janssens, I.A. (2015). Declining global warming effects on the phenology of spring leaf unfolding. *Nature*, 526, 104–107.
- Fukami, T. (2015). *Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects*, vol. 46 of *Annual Review of Ecology Evolution and Systematics*, pp. 1–23.
- Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N. & Hughes, K.A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, 525, 372–+.
- Ghalambor, C.K., McKay, J.K., Carroll, S.P. & Reznick, D.N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21, 394–407.

- Gienapp, P., Hemerik, L. & Visser, M.E. (2005). A new statistical tool to predict phenology under climate change scenarios. *Global Change Biology*, 11, 600–606.
- Gotelli, N.J. & Graves, G.R. (1996). The temporal niche. In: *Null Models in Ecology* (eds. Gotelli, N.J. & Graves, G.R.). Smithsonian Institution, Washington, DC, pp. 95–111.
- Gourbiere, S. & Menu, F. (2009). Adaptive dynamics of dormancy duration variability: evolutionary trade-off and priority effect lead to suboptimal adaptation. *Evolution*, 63, 1879–1892.
- Heberling, J.M., Cassidy, S.T., Fridley, J.D. & Kalisz, S. (2019). Carbon gain phenologies of spring-flowering perennials in a deciduous forest indicate a novel niche for a widespread invader. *New Phytologist*, 221, 778–788.
- Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K., Engelbrecht, F., Guiot, J., Hijioka, Y., Mehrotra, S., Payne, A., Seneviratne, S., Thomas, A., R. Warren, R. & Zhou, G. (2018). *Impacts of 1.5C Global Warming on Natural and Human Systems*.
- Inouye, B.D., Ehrlen, J. & Underwood, N. (2019). Phenology as a process rather than an event: from individual reaction norms to community metrics. *Ecological Monographs*, 89.
- Inouye, D.W., Morales, M.A. & Dodge, G.J. (2002). Variation in timing and abundance of flowering by *Delphinium barbeyi* Huth (Ranunculaceae): the roles of snowpack, frost, and La Niña, in the context of climate change. *Oecologia*, 130, 543–550.
- IPCC (2014). *Climate Change 2014: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Ishioka, R., Muller, O., Hiura, T. & Kudo, G. (2013). Responses of leafing phenology and photosynthesis to soil warming in forest-floor plants. *Acta Oecologica-International Journal of Ecology*, 51, 34–41.
- Jansson, R. & Dynesius, M. (2002). The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics*, 33, 741–777.
- Jing, J., Li, K. & Liu, Z.G. (2016). Effects of varying temperature on leaf phenology and herbivory of dominant tree species in subtropical evergreen broad-leaves forest in eastern china. *Polish Journal of Ecology*, 64, 53–69.
- Johansson, J. & Bolmgren, K. (2019). Is timing of reproduction according to temperature sums an optimal strategy? *Ecology and Evolution*, 9, 11598–11605.
- Johansson, J. & Jonzen, N. (2012). Game theory sheds new light on ecological responses to current climate change when phenology is historically mismatched. *Ecol Lett*, 15, 881–8.
- Kharouba, H.M., Ehrlen, J., Gelman, A., Bolmgren, K., Allen, J.M., Travers, S.E. & Wolkovich, E.M. (2018). Global shifts in the phenological synchrony of species interactions over recent decades. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 5211–5216.

- Kharouba, H.M., Paquette, S.R., Kerr, J.T. & Vellend, M. (2014). Predicting the sensitivity of butterfly phenology to temperature over the past century. *Global Change Biology*, 20, 504–514.
- Lasky, J.R., Uriarte, M. & Muscarella, R. (2016). Synchrony, compensatory dynamics, and the functional trait basis of phenological diversity in a tropical dry forest tree community: effects of rainfall seasonality. *Environmental Research Letters*, 11.
- Legault, G. & Melbourne, B.A. (2019). Accounting for environmental change in continuous-time stochastic population models. *Theoretical Ecology*, 12, 31–48.
- Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17, 183–189.
- Loken, E. & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, 355, 584–585.
- Mayfield, M.M. & Levine, J.M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, 13, 1085–1093.
- McCabe, G.J., Ault, T.R., Cook, B.I., Betancourt, J.L. & Schwartz, M.D. (2012). Influences of the el nino southern oscillation and the pacific decadal oscillation on the timing of the north american spring. *International Journal of Climatology*, 32, 2301–2310.
- Menzel, A., Von Vopelius, J., Estrella, N., Schleip, C. & Dose, V. (2006). Farmers’ annual activities are not tracking the speed of climate change. *Climate Research*, 32, 201–207.
- Metcalf, C.J.E., Burghardt, L.T. & Koons, D.N. (2015). Avoiding the crowds: the evolution of plastic responses to seasonal cues in a density-dependent world. *Journal of Ecology*, 103, 819–828.
- Morris, W.F., Pfister, C.A., Tuljapurkar, S., Haridas, C.V., Boggs, C.L., Boyce, M.S., Bruna, E.M., Church, D.R., Coulson, T., Doak, D.F., Forsyth, S., Gaillard, J.M., Horvitz, C.C., Kalisz, S., Kendall, B.E., Knight, T.M., Lee, C.T. & Menges, E.S. (2008). Longevity can buffer plant and animal populations against changing climatic variability. *Ecology*, 89, 19–25.
- Nakazawa, T. & Doi, H. (2012). A perspective on match/mismatch of phenology in community contexts. *Oikos*, 121, 489–495.
- Parnesan, C. (2006). *Ecological and evolutionary responses to recent climate change*, vol. 37 of *Annual Review of Ecology Evolution and Systematics*, pp. 637–669.
- Parnesan, C. (2007). Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Global Change Biology*, 13, 1860–1872.
- Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B. & Davies, T.J. (2017). A statistical estimator for determining the limits of contemporary and historic phenology. *Nature Ecology & Evolution*, 1, 1876–+.

809 Piersma, T. & Drent, J. (2003). Phenotypic flexibility and the evolution of organismal design.
810 *Trends in Ecology & Evolution*, 18, 228–233.

811 Provan, J. & Bennett, K.D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends*
812 *in Ecology & Evolution*, 23, 564–571.

813 Reed, T.E., Grotan, V., Jenouvrier, S., Saether, B.E. & Visser, M.E. (2013). Population growth
814 in a wild bird is buffered against phenological mismatch. *Science*, 340, 488–491.

815 Reed, T.E., Waples, R.S., Schindler, D.E., Hard, J.J. & Kinnison, M.T. (2010). Phenotypic
816 plasticity and population viability: the importance of environmental predictability. *Proceed-*
817 *ings of the Royal Society B-Biological Sciences*, 277, 3391–3400.

818 Revilla, T.A., Encinas-Viso, F. & Loreau, M. (2014). (a bit) earlier or later is always better:
819 Phenological shifts in consumer-resource interactions. *Theoretical Ecology*, 7, 149–162.

820 Richardson, A.D., Bailey, A.S., Denny, E.G., Martin, C.W. & O’Keefe, J. (2006). Phenology
821 of a northern hardwood forest canopy. *Global Change Biology*, 12, 1174–1188.

822 Rudolf, V.H.W. (2019). The role of seasonal timing and phenological shifts for species coexis-
823 tence. *Ecology Letters*.

824 Sagarin, R. (2001). Phenology - false estimates of the advance of spring. *Nature*, 414, 600–600.

825 Sale, P.F. (1977). Maintenance of high diversity in coral reef fish communities. *American*
826 *Naturalist*, 111, 337–359.

827 Screen, J.A. (2014). Arctic amplification decreases temperature variance in northern mid- to
828 high-latitudes. *Nature Climate Change*, 4, 577–582.

829 Simons, A.M. (2011). Modes of response to environmental change and the elusive empirical
830 evidence for bet hedging. *Proceedings of the Royal Society B-Biological Sciences*, 278, 1601–
831 1609.

832 Singer, M.C. & Parmesan, C. (2010). Phenological asynchrony between herbivorous insects
833 and their hosts: signal of climate change or pre-existing adaptive strategy? *Philosophical*
834 *Transactions of the Royal Society B: Biological Sciences*, 365, 3161–3176.

835 Smaldino, P.E. & McElreath, R. (2016). The natural selection of bad science. *Royal Society*
836 *Open Science*, 3.

837 Socolar, J.B., Epanchin, P.N., Beissinger, S.R. & Tingley, M.W. (2017). Phenological shifts
838 conserve thermal niches in north american birds and reshape expectations for climate-driven
839 range shifts. *Proceedings of the National Academy of Sciences of the United States of America*,
840 114, 12976–12981.

841 Steer, N.C., Ramsay, P.M. & Franco, M. (2019). nlstimedist: An r package for the biologi-
842 cally meaningful quantification of unimodal phenology distributions. *Methods in Ecology and*
843 *Evolution*, 10, 1934–1940.

- Stinchcombe, J.R., Weinig, C., Ungerer, M., Olsen, K.M., Mays, C., Halldorsdottir, S.S., Purugganan, M.D. & Schmitt, J. (2004). A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 4712–4717.
- Stocker, T., Qin, D. & Plattner, G. (2013). Climate change 2013: The physical science basis. *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers (IPCC, 2013)*.
- Tansey, C.J., Hadfield, J.D. & Phillimore, A.B. (2017). Estimating the ability of plants to plastically track temperature-mediated shifts in the spring phenological optimum. *Global Change Biology*, 23, 3321–3334.
- Thackeray, S.J., Henrys, P.A., Hemming, D., Bell, J.R., Botham, M.S., Burthe, S., Helaouet, P., Johns, D.G., Jones, I.D., Leech, D.I., Mackay, E.B., Massimino, D., Atkinson, S., Bacon, P.J., Brereton, T.M., Carvalho, L., Clutton-Brock, T.H., Duck, C., Edwards, M., Elliott, J.M., Hall, S.J.G., Harrington, R., Pearce-Higgins, J.W., Høye, T.T., Kruuk, L.E.B., Pemberton, J.M., Sparks, T.H., Thompson, P.M., White, I., Winfield, I.J. & Wanless, S. (2016). Phenological sensitivity to climate across taxa and trophic levels. *Nature*, 535, 241.
- Tufto, J. (2015). Genetic evolution, plasticity, and bet-hedging as adaptive responses to temporally autocorrelated fluctuating selection: A quantitative genetic model. *Evolution*, 69, 2034–2049.
- Tuljapurkar, S., Gaillard, J.M. & Coulson, T. (2009). From stochastic environments to life histories and back. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 364, 1499–1509.
- Venable, D.L. (2007). Bet hedging in a guild of desert annuals. *Ecology*, 88, 1086–1090.
- Visser, M.E. & Both, C. (2005). Shifts in phenology due to global climate change: the need for a yardstick. *Proceedings of the Royal Society B-Biological Sciences*, 272, 2561–2569.
- Visser, M.E. & Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. *Nature Ecology & Evolution*, 3, 879–885.
- Visser, M.E., Holleman, L.J.M. & Gienapp, P. (2006). Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia*, 147, 164–172.
- Vrba, E.S. (1980). Evolution, species and fossils - how does life evolve. *South African Journal of Science*, 76, 61–84.
- Vrba, E.S. (1985). Environment and evolution - alternative causes of the temporal distribution of evolutionary events. *South African Journal of Science*, 81, 229–236.
- Wadgymar, S.M., Ogilvie, J.E., Inouye, D.W., Weis, A.E. & Anderson, J.T. (2018). Phenological responses to multiple environmental drivers under climate change: insights from a

long-term observational study and a manipulative field experiment. *New Phytologist*, 218, 517–529.

Wainwright, C.E. & Cleland, E.E. (2013). Exotic species display greater germination plasticity and higher germination rates than native species across multiple cues. *Biological Invasions*, 15, 2253–2264.

Wilczek, A.M., Burghardt, L.T., Cobb, A.R., Cooper, M.D., Welch, S.M. & Schmitt, J. (2010). Genetic and physiological bases for phenological responses to current and predicted climates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 3129–3147.

Wilczek, A.M., Cooper, M.D., Korves, T.M. & Schmitt, J. (2014). Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 7906–7913.

Wilczek, A.M., Roe, J.L., Knapp, M.C., Cooper, M.D., Lopez-Gallego, C., Martin, L.J., Muir, C.D., Sim, S., Walker, A., Anderson, J., Egan, J.F., Moyers, B.T., Petipas, R., Giakountis, A., Charbit, E., Coupland, G., Welch, S.M. & Schmitt, J. (2009). Effects of genetic perturbation on seasonal life history plasticity. *Science*, 323, 930–934.

Willis, C.G., Ruhfel, B.R., Primack, R.B., Miller-Rushing, A.J., Losos, J.B. & Davis, C.C. (2010). Favorable climate change response explains non-native species' success in Thoreau's woods. *PLoS ONE*, 5, e8878.

Wolkovich, E.M. & Cleland, E.E. (2011). The phenology of plant invasions: A community ecology perspective. *Frontiers in Ecology and the Environment*, 9, 287–294.

Wolkovich, E.M. & Cleland, E.E. (2014). Phenological niches and the future of invaded ecosystems with climate change. *AoB Plants*, 6, plu013.

Wolkovich, E.M., Cook, B.I., Allen, J.M., Crimmins, T.M., Betancourt, J.L., Travers, S.E., Pau, S., Regetz, J., Davies, T.J., Kraft, N.J.B., Ault, T.R., Bolmgren, K., Mazer, S.J., McCabe, G.J., McGill, B.J., Parmesan, C., Salamin, N., Schwartz, M.D. & Cleland, E.E. (2012). Warming experiments underpredict plant phenological responses to climate change. *Nature*, 485, 494–497.

Wolkovich, E.M., Davies, T.J., Schaefer, H., Cleland, E.E., Cook, B.I., Travers, S.E., Willis, C.G. & Davis, C. (2013). Temperature-dependent shifts in phenology contribute to the success of exotic species with climate change. *American Journal of Botany*, 100, 1407–1421.

Wong, T.G. & Ackerly, D.D. (2005). Optimal reproductive allocation in annuals and an informational constraint on plasticity. *New Phytologist*, 166, 159–172.

Yang, L.H. & Czenzer, M.L. (2020). Seasonal windows of opportunity in milkweed-monarch interactions. *Ecology*, 101.

Yousey, A.M., Chowdhury, P.R., Biddinger, N., Shaw, J.H., Jeyasingh, P.D. & Weider, L.J. (2018). Resurrected 'ancient' daphnia genotypes show reduced thermal stress tolerance compared to modern descendants. *Royal Society Open Science*, 5.

- 917 Zettemoyer, M.A., Schultheis, E.H. & Lau, J.A. (2019). Phenology in a warming world: dif-
918 ferences between native and nonnative plant species. *Ecology Letters*.
- 919 Zhu, J.T., Zhang, Y.J. & Wang, W.F. (2016). Interactions between warming and soil moisture
920 increase overlap in reproductive phenology among species in an alpine meadow. *Biology*
921 *Letters*, 12, 4.

Note that reviewer comments are in *italics*, while our responses are in regular text, and all in-text citations generally cross-reference to the main text.

Editor's comments:

Reviewers were quite critical of a number of aspects of the article. In the end, I think the biggest issue is one of communication. The authors need to focus their arguments much more clearly and deliberately.

We appreciate the editor's comments about clarity of message, and agree that more focused, clear arguments would do much to address the reviewers' concerns. Thus, we have overhauled the manuscript, especially sections 1.1-1.3 (now sections 1.1-1.2, 'Defining & measuring tracking' and 'Tracking in single-species environment') to be more precise and shorter (this material is now covered in sections 1.1-1.2, which present the relevant material, but in a space 35% shorter compared with our previous submission), while focusing our arguments around tracking in multi-species competitive environments (this is the one section that is now longer, by approximately 25%). We have overhauled our figure that defines tracking and been more careful in our definition of fundamental versus environmental tracking throughout the manuscript. Additionally, we have more clearly separated evolutionary and ecological theory, which we believe understandably led to some confusion. We believe the revised submission is much improved (and overall 20% shorter) and explain our changes in more detail in our point-by-point response to reviewers below.

Referee 1 comments:

The authors present a manuscript that attempts to summarize our current knowledge about ecological tracking, i.e. the ability of an organism to track the phenological niche. This is particularly interesting in the context of climate change and earlier onset of seasons in the northern hemisphere. The topic of phenological shifts is interesting, and I found the manuscript overall very well written.

We thank the reviewer for the positive comments on the manuscript's topic and writing style.

I have a few general concerns about the manuscript which I detail below, and some specific ones, which I will address later in a chronological order.

1) I am not familiar with the topic of ecological tracking, but I am very familiar with the literature regarding phenological shifts in response to climate change. In my opinion, ecological tracking appears to me as a rebranding of a phenomenon about much has been written. I am aware that the authors will disagree with this view, but their manuscript did not convince me that ecological tracking is fundamentally different from the widely observed phenological shifts. Maybe it is a subset of those, but it is nothing new. Nevertheless, the effect of phenological

changes on ecological communities is an interesting one.

We agree our manuscript’s topic is easily and readily applied to phenological shifts (as Reviewer 4 also noted), but we avoided this term given that phenology is generally defined as ‘the recurring timing of life history events’ (defined on line 20) and a number of events we review (and to which this manuscript applies) fall outside this definition. We understand that the definition of phenology may be evolving in the literature and have tried to be up front about the reasoning for our terminology; when we define tracking, we now state (line 73-line 75):

Both these definitions are readily applied to phenology—the timing of recurring life history events—though they can also apply to non-recurring life history events (e.g., seed germination), or events not normally defined as part of life history.

2) After carefully reading the manuscript, I did not understand what this manuscript actually is about and what the authors want to achieve with it. a) The authors claim it is a review, but many studies- and many reviews about them (e.g. by C. Parmesan or A. Menzel) - have described phenological shifts in response to climate change. Only very few of those are mentioned, and in the description of their narrow search criteria they end up with only a handful of studies, because it appears that the reviews and the studies therein were actively omitted.

This is a good point, as we were too broad in our previous draft of our aims (e.g., ‘we review current knowledge on tracking both in empirical data...’), and made it seem we were aiming to review the full literature on phenological shifts. This was not our aim, and we are more specific now (see line 54-line 61, see also changes to abstract):

Here, we review the concept of tracking as commonly used in the empirical climate change impacts literature and in related ecological theory. We provide definitions that distinguish between fundamental and environmental tracking, highlighting the distinction between measuring tracking and its fitness outcomes in empirical systems. Then, after briefly reviewing evolutionary theory that predicts variation in tracking across species and environments, we examine how well community assembly theory—especially priority effects and modern coexistence theory—can be extended to test the current paradigm that climate change should favor species that track.

Regarding the references—we did cite Menzel *et al.* (2006) and now also cite Parmesan (2006). As our aim is not a full review of all studies of phenological shifts we have attempted to balance older and newer references, as we wanted a mix of foundational studies with new work that gives up-to-date estimates based on to-date climate change (i.e., studies from the mid 2000s generally use data when global average warming was lower).

b) It is also not clear to me why they reviewed these papers and not the theoretical literature or the physiological literature. Both types of studies were discussed in detail in the manuscript but not reviewed at least I would doubt that the lack of studies identified by the authors regarding theory or the physiology of the cues can be based on a handful of studies. There must be

myriads of studies in animals and plants addressing the physiological basis of cueing for phenological events, e.g. flowering time in plants or breeding time or migration time in birds. I was particularly surprised that they also excluded theoretical studies in their search, while at the same time relying heavily on theoretical papers throughout the remaining manuscript to describe several aspects of ecological tracking and its consequences for populations and communities. If this was a review, why exclude theory?

We believe the reviewer is here (and below) referring to the part of the paper regarding a targeted systematic literature review for studies examining tracking and other traits together. This review is only mentioned in the Box ‘Trait trade-offs with tracking,’ and is not meant to be the focus of our paper. For this review we did not exclude physiological studies, though we did exclude modeling and theory studies because they did not have data (only 12 of 231 papers). We have tried to clarify this in our text within the Box and in the supplement (e.g., we have renamed this section ‘Literature review of studies examining tracking & traits,’ and we now open this section with “To examine current evidence of what traits may trade-off with tracking”).

We completely agree with the reviewer that theory and physiology are quite relevant to our topic and do review a number of relevant studies throughout the main text.

c) It was unclear to me whether they were searching for studies that explicitly talk about ‘ecological tracking’ (which are, I believe few), or any study that has ever observed a shift in phenology due to warming. The latter is not achieved, but it is also maybe not needed given the many reviews we already have. The former is probably not needed, too, because ecological tracking is, in my opinion, largely a rebranding of (adaptive) phenological shifts.

Again, we believe the reviewer is here referring to the part of the paper regarding a targeted systematic literature review for studies examining tracking and other traits together, which is only mentioned in the main text in the Box ‘Trait trade-offs with tracking.’ We have worked in the supplement to clarify that we are specifically looking for studies that examine tracking and traits at once; our search terms do not require the term tracking (or track*) but do require reference to a trait. Thus, many studies that only examine phenological shifts would be excluded, as finding those studies was not the aim of this systematic literature review.

d) If I accept it is not a review, then it is possibly an opinion paper or a perspective. I understood that the authors mention a whole suite of understudied aspects of ecological tracking and that they want to fuel a whole suite of new studies. However, for a perspective, the rationale for addressing some of the understudied aspects of ecological tracking is not always clear. For example, for studying mismatches between phenologies of coexisting species, it is not crucial to know the exact cue. Also, while the need for non-stationary models appears logical, I could not find anywhere clear predictions about why and how coexistence mechanisms would be changing differently in non-stationary systems compared to stationary (but fluctuating) ones. This is

regrettable because I assumed that the interaction between tracking and coexistence mechanisms was a main focus of this manuscript at least this would be an interesting topic.

We appreciate the reviewer's concern and it is in line with Reviewer 4's concerns as well. To address this we worked to focus more on the interaction between tracking and coexistence mechanisms. To do this we have merged two former sections and significantly streamlined the sections before 'Tracking in multi-species environments.' We have not completely removed these sections as we believe (as did previous reviewers in their comments) that some background is needed before the section on coexistence mechanisms. Additionally, we give an example of a model with a fluctuating environment where stationary and non-stationary outcomes are not the same in the Box 'Adding tracking and non-stationarity to a common coexistence model' and now mentioned on line 266-line 274, and we now provide a broader example on line 255-line 258. Finally, we have clarified why we believe the cues matter by updating the main text throughout (especially line 180-line 191) and Figure 2.

In fact, I would not expect large differences between a classical storage effect model and a model where the environment changes gradually and directionally over time, especially as storage effect models also look at environments with different statistical properties. Specifically, if say, we have a storage effect model (or a model addressing priority effects) where the environment does not fluctuate strongly, species would probably not be selected for being able to track, simply because tracking is not needed when the environment is stable. However, if we model (as in a classical storage effect scenario, or in a priority effect model) the environment as highly variable and unpredictable in time (and space), then species inhabiting such an environment must be able to track, because they cannot know what the ideal timing would be in any given year, unless there is a good cue (in which case the environment would not be unpredictable). Thus, I would expect a similar change from non-tracking to tracking when comparing stable with fluctuating (stationary) environments as when comparing a stationary with a non-stationary one. In other words, species inhabiting highly variable environments should be tracking, which may equip them with an advantage also in a gradually changing world. This idea has been voiced before in models (e.g. Bonebrake, T. C. & Mastrandrea, M. D. 2010. Proc. Natl. Acad. Sci. USA 107: 12581-12586) but also in experimental studies conducted in fluctuating habitats, where no effect of experimentally induced climate change was found.

So maybe the lack of a prediction about why we should look at non-stationary models and how their outcome would be different from what we know may be explained: the outcome would not be much different. It is also possible that the authors had attempted to exactly derive such a prediction in their model in the previous version of this manuscript, but I understood that they did in fact not produce any surprising results.

We can see that we did not make a strong enough case in our last submission for why outcomes from a fluctuating but stationary system would be different than in many non-stationary systems. We give an example of a model with a fluctuating environment where stationary and non-stationary outcomes are not the same in the Box 'Adding tracking and non-stationarity to a

common coexistence model’ and better highlight this in the main text of our revised manuscript (line 266-line 274):

As an example, we modeled a shift to earlier growing seasons using a common coexistence model where the environment is defined as a limiting resource that determines the start of growth each year. The timing of the resource relative to each species’ ideal timing determines the species-specific germination fraction each year, allowing us to include fundamental tracking. The shift to earlier seasons favored species that could track and narrowed the region of coexistence maintained by a trade-off between tracking and competitive ability (via R^* , see Fig. 3 and Box: ‘Adding tracking and non-stationarity to a common coexistence model’). Like all models, it rests on a number of assumptions, including that species’ cues remain as reliable in the non-stationary environment, but shows how non-stationarity could benefit trackers.

We also now provide a broader example on line 255-line 258, where we state, “For example, storage effect models predict shifts in communities when environmental change alters the long-term covariance between the environment and competition (i.e., decreasing $cov(E_i, C_i)$), leading to a decrease in the storage effect as a means of competitive coexistence.” Indeed, it is this changing covariance between environment and competition—on which the storage effect depends—that we expect to be different in stationary and non-stationary environments.

3) I was also not sure what exactly the topic of this manuscript is. From the previous reviews and the author’s replies I understood I that this manuscript aimed at coupling ecological tracking theory with coexistence theory, which would be an exciting topic. However, only approx. 10% of the manuscript is devoted to this topic. The remaining 90% are spread across several different and partly unrelated aspects of ecological tracking. These are, to name a few, the lack of physiological evidence for cueing, definitions of ecological tracking and measuring it, description of bet hedging as opposed to tracking, a brief note about the equivalence of phenotypic plasticity and ecological tracking, trade-offs between tracking ability and competitive ability (why this trade-off and no other one?), and some more. Interestingly, none of these various topics is actually reviewed in detail, which brings me back to my initial question of whether or not this is a review.

We appreciate the reviewer’s concern and have worked to streamline the manuscript so that more of the text is devoted to ‘Tracking in multi-species environments.’ Sections on physiological evidence for cueing, definitions of ecological tracking and measuring it, and review of plasticity versus bet-hedging are now 35% shorter, but we have not removed them because we believe they are critical background for examining ecological tracking and coexistence, and highlight areas where we need advances if we hope to better understand tracking and coexistence. We have tried to clarify this throughout (including edits to the abstract). Previous reviewer comments also stressed these connections and we think they are important, but we could have done better to present them more briefly and as background, which we now do.

In my opinion, the authors do themselves a disservice by evoking expectations about linking ecological tracking with coexistence theory, when in the end they spread sometimes thinly across several aspects of ecological tracking. The manuscript could thus really profit from being concise in the selection of aspects discussed and then discuss these aspects exhaustively.

As outlined above, we have worked to streamline the sections outside of those focused on tracking in multi-species environments.

4) It is not clear to me why out of all possible biotic interactions, competition is dealt with so prominently. I understand that competition is the other side of the coexistence coin, but since coexistence theory is not the core of the manuscript, other biotic interactions should have been discussed, too. There could be positive interactions that are decoupled by climate change and (as mentioned by the authors) decoupling of interactions among trophic levels. The subsequent focus on trade-offs between tracking and competitive ability appears to me equally arbitrary. If we accept that plasticity comes at a cost, it can trade-off with any trait. For example, I would think that stress resistance (which in plants is assumed to trade-off with competitive effect ability) would trade-off with tracking, ability, too. Also, there could be trade-offs between phenological plasticity (i.e. tracking) and plasticity in other traits that enable fitness homeostasis even if no ecological tracking occurs. This relationship is not addressed. However, it could be fundamental if organisms are highly plastic in other traits, in which case they may not even need to track.

We focus on the trade-off between tracking and competitive traits as it is predicted by theory and the most supported by empirical evidence. Additionally, current coexistence theory outlines how physiological stress may change the timescales of species interactions (by slowing down growth, for example), but it should not fundamentally reshape the mechanisms of coexistence (Chesson & Huntly, 1997).

We realize, however, we did not make our focus very clear in our previous draft. Our current draft streamlines early sections to focus on tracking in multi-species environments, specifically with a focus on competitive environments. We now state in the abstract (line 8-line 15):

We argue that much current theory for tracking ignores the importance of a multi-species context beyond trophic interactions. Yet community assembly theory predicts competition should drive variation in tracking and trade-offs with other traits. Existing community assembly theory can help us understand tracking in stationary and non-stationary systems and, thus, predict the species- and community-level consequences of climate change.

We also have tried to highlight why this perspective is important throughout, including line 37-line 38, “Further, there has been comparatively little work connecting tracking to community assembly theory, which shows temporal sequencing and environmental variability can alter the relative fitness and niche differences between species that determine coexistence, suggesting important ecological constraints to tracking” and in our revised section ‘Tracking in multi-species environments.’

5) *Ecological tracking is regarded exclusively as a plastic response. However, the (very few) solid studies on evolutionary change in response to climate change indicate that phenological traits could be among the first under real selection. I was asking myself why plasticity should be the main mechanism by which species can track, and whether we need this assumption for defining ecological tracking, or whether the definition could also embrace rapid evolutionary change.*

We understand the reviewer's concern that adaptive tracking (*sensu* Simons, 2011) theoretically could equally explain tracking and we understand the concern that there not many rigorous studies on evolutionary change in response to climate change. However, most studies (of which we are aware) that have estimated plastic versus evolutionary change in phenology find it is mostly due to plasticity and many phenological traits are highly plastic (if the environment is defined in calendar time) thus we have retained our focus on plasticity but now have worked to be clear about this, line 95-line 102:

Environmental tracking at the individual-level is a purely plastic response to environmental variation (in line with current findings on most climate change responses, Bonamour *et al.*, 2019), with the plasticity itself an outcome of selection (Chevin *et al.*, 2010) through its connection to fundamental tracking (Fig. 2). At the population-level, tracking may also incorporate evolutionary change—change in genotype frequencies—depending on both the timescales of study and the species' generation time (this evolutionary response can be predicted as the difference between the environmental sensitivity of phenotypic selection and an organism's plasticity, $|B - b|$ in Chevin *et al.*, 2010).

Specific comments (chronological order, line numbers are references given): Line 1-12: reference to the many studies and reviews about 'escape in time' is missing (e.g. Parmesan, Menzel, and many more). This leaves the impression that we know nothing about ecological tracking, which is, in my opinion, not true.

We now cite Menzel *et al.* (2006); Parmesan (2006) on line 29.

37ff: *Do we need to show that tracking is related to fitness? Isn't that self-evident and if not, why?*

We have removed this line, but have worked to address this in the section 'Defining & measuring tracking.'

74ff: *Is it true that we know nothing about environmental cues? I did not take the time to dive deep into the literature but I would think that studies on birds and plants are plentiful. Maybe the mechanistic studies (i.e. experimental) are rarer than correlations (but they do exist, e.g. reciprocal transplant studies and not only Arabidopsis), but even evidence for correlations of*

e.g. flowering time with e.g. growing degree day units is abundant.

In streamlining the manuscript we have removed this paragraph.

84ff: The advancement in phenology by certain numbers of days has been demonstrated by C. Parmesan or A. Menzel (and others) much earlier than what is cited here. I am puzzled why their work is not cited.

We now cite Menzel *et al.* (2006); Parmesan (2006) on line 65.

93ff: Why is it so crucial to know the exact physiological mechanism of tracking and why the cue? For example, if we are mostly interested in the same trophic level and competitive interactions, we may, as a first approximation, assume that the organisms use a similar set of cues. Also, if it is true that we know nothing about the relationship between physiology and the cue, this seems a rather bleak perspective and may lead to the conclusion that we will never understand ecological tracking. So why is this important?

We have updated Figure 2 and the text throughout (e.g., line 180-line 191) to clarify why we believe understanding the cues is important, but have otherwise worked to shorten this section to address this reviewer's and reviewer 4's concerns.

192-194. Some variable environments do provide cues, e.g. in the Sonoran desert annual system (see Pake, C. E. and Venable, D. L. 1996. Ecology 77: 1427-1435), the amount of the first rainfall in a year seems to partly predict the rainfall of the season. Predictive germination has also been addressed from a theoretical perspective by Cohen (1967) and subsequent authors.

We agree and cite papers by Venable, which build on Cohen's work throughout the manuscript (e.g., line 21, line 162).

195ff: One important aspect of the cueing seems to me the reliability of the cue. Unfortunately, the authors do not mention this and only focus on benefits and costs. To me, this seems a key aspect which is tightly related to the costs (i.e. low reliability, high potential costs). The reliability is not touched upon in the cost-benefit discussion.

We agree cue reliability is important; we now define it on line 88, and explain its importance on line 178-line 191:

Critical to predictions is whether cue systems maintain their reliability with change; i.e., whether they continue to yield high fundamental tracking (Bonamour *et al.*, 2019). Consider a simple case in which an organism's cues evolved based on a correlation between peak prey abundance and daylength: in a stationary environment

the daylength cue may be fairly reliable (generally predicting peak prey abundance based on daylength, with some interannual variation), but would become unreliable, and lead to fitness declines, if warming continually advances peak prey abundance. Multivariate cues are often argued to be more reliable because they can capture multiple attributes of the environment (Dore *et al.*, 2018; Bonamour *et al.*, 2019), but they may be equally vulnerable to failure if non-stationarity decouples the cues from fundamental tracking (Bonamour *et al.*, 2019) and thus optimal fitness is no longer associated with the cue system. Predicting the outcome of non-stationarity from the stationary environment requires that researchers know: (1) the full cue system of an organism, (2) how it relates to fundamental tracking, and (3) how both that cue system and the underlying fundamental model shift with non-stationarity.

208ff: The discussion about bet hedging is too much black and white (i.e. between not germinating and germinating). There is also plasticity in germination rates and some of it is driven by cues (see literature about predictive germination). I would actually assume that in the ‘classical’ bet-hedging system (desert annuals), tracking ability would be selected for very strongly because in a fluctuating environment, plants need to respond very plastically to the ever-changing conditions. So the idea that there is either tracking or bet-hedging is not plausible for me.

Agreed, we have re-written the section on bet-hedging (line 158-line 170):

Tracking should generally not be favored in unpredictable environments (e.g. when early season climate cannot be used to predict later season climate), or environments where species otherwise face high uncertainty in the timing of investment decisions. Instead theory suggests the optimal strategy may often be to bet-hedge (Venable, 2007; Donaldson-Matasci *et al.*, 2012; de Casas *et al.*, 2015) via a high diversity of timings or a conservative timing. Because bet-hedging, by definition, maximizes geometric-mean fitness in the long-run, its short-term outcomes can appear maladaptive. How often observed ‘maladaptations,’ which may easily include species that do not track or appear to track poorly, are actually the outcome of bet-hedging is difficult to estimate, as robustly assessing bet-hedging requires studies of fitness over longer timescales than many current field experiments (Simons, 2011). Environmental variation often includes both predictable and less predictable aspects. In such cases theory predicts organisms may evolve tracking that is a mixed strategy between bet-hedging and plasticity (Wong & Ackerly, 2005).

217-229: This paragraph does not appear to contain much information, so it could be left out.

We have shortened this into one sentence that we include regarding constraints and plasticity, line 144-line 150.

243ff: I am missing an in-depth discussion about plasticity, i.e. the ability to maintain fitness (fitness homeostasis) even when the environment fluctuates strongly. Plasticity is expected to

evolve under unpredictably varying conditions, and tracking is only one aspect of that plasticity. There should be trade-offs among the different types of plasticity.

We have revamped the section on plasticity (line 136-line 157) and worked to shorten it. Given this reviewer and reviewer 4's request to focus the paper more we have kept this section short.

1.4: This paragraph is entirely devoted to tracking-competitive ability relationships. It seems logical that tracking ability should also trade-off with tolerance to stress (e.g. low temperatures if e.g. bud burst is early) which in turn may trade-off with competitive ability.

We believe the reviewer means that tracking could co-vary with stress tolerance, which we agree with, and now mention on line 214 and line 395.

336ff: Isn't the storage effect the same as tracking only that it is about inter-annual variation and not variability in intra-annual timing? So what would then be the fundamental difference between stationary and non-stationary models when, e.g. we start with a storage effect model in a randomly fluctuating environment where species must already be able to track? I feel it would be crucial to provide clear predictions about what non-stationary models may predict in contrast to 'classical' models. Without these, the call for more and different models' is not very well justified. Here, the main justification is that it has not been done', but not this is why stationary models are entirely misleading'. Unfortunately, the Box remains vague about this.

The storage effect model depends on inter-annual variation in species-responses that result in positive covariance between the environmental response (species fitness in response to the environment without competition) and competitive response (the decrease in the fitness due to both intra and interspecific competition). Certainly, tracking is one mechanism by which a species can increase its fitness by (for example) germinating more in a 'good' year; however, if a 'good' year for species A is also a 'good' year for species B, then there will be increased competition; i.e., positive covariance between the environment and competition. This increase in competition in 'good' years is fundamental to the storage effect coexistence mechanism. Under non-stationary environments, we expect that this covariance between environment and competition is likely to change, either because of differential responses to changing environmental cues or the direct effects of the environment on competing species. Please see our reply to second part of point d) above (comment starting with 'In fact, I would not expect large differences between a classical storage effect model...').

1.5 I found this section somewhat if not completely- redundant with the sections before and was not sure why it is needed. Much of the discussion here remains somewhat vague. The conclusions are that we need more interdisciplinarity, more understanding and measuring of tracking, more looking at trade-offs with selected traits, and more models that are different from the current ones. Overall, this is not the strongest section of the manuscript. It could be merged

with the previous sections and made much more concise.

We appreciate the reviewer's concerns. We have shortened the previous sections so that this section is less redundant, and this section is now shorter by roughly 50%.

Box

578-581: Could the finding of early species tracking more simply be due to the fact that response to environmental variables (e.g. higher temperatures) follow a logistic curve where the late species attain high fitness because they are always in their climatic comfort zone? Whereas the early species experience, during their life or evolutionary history a much larger range of temperatures, some of which are clearly decreasing fitness?

This is an interesting hypothesis and possible, but we are not aware of any formal studies of this.

600ff: Many models and data have been published about within-season timing of (germination) events. They could make a valuable contribution to this section (e.g. Simons, A. M. 2009. Proc. R. Soc. B 276: 1987 1992. Simons, A. M. 2011. Proc. R. Soc. B 278: 1601 1609).

We agree and now cite this paper (line 167) in our section on evolutionary theory. This box is focused on one particular model (an ecological model with no evolution) and for clarity we mention only the relevant model in the Box. Throughout the manuscript we have also worked to clarify where we are speaking mainly about evolutionary versus ecological models.

607ff: I believe that a similar storyline could be created with stress tolerance instead of competitive ability.

Agreed, we focus here on competitive ability as that is what the literature has found evidence for.

Referee 2 comments:

The resubmitted paper 'How environmental tracking shapes species and communities in stationary and non-stationary systems' by Wolkovich and Donahue deals with environmental tracking, specifically how environmental tracking can be measured and analyzed, how it may influence species co-existence and species responses to climate change. I think the topic of the paper is novel and highly relevant, and overall the authors did a very good job in reviewing the literature on the topic. I specifically like the part about how tracking may trade-off with other traits (e.g. those related with competition) and thereby shape the co- existence among species in ecological communities.

We thank the reviewer for their comments and have worked to retain the better parts of the

manuscript while improving the rest of it based on feedback from this and the other reviewers.

I only have one point to criticize: although the authors highlight that researchers are increasingly recognizing the need to consider multiple climate variables (L 14) this review is mainly focused on environmental tracking in response to temperature changes. I am aware that there is much more known about phenological responses to temperature change compared to precipitation change, which is also supported by the result of the literature search in the Supplement. However, as this review deals with climate change and not only climate warming and we know that climate change is complex and multivariate, I would love to see more examples in the text about environmental tracking and precipitation change. Are there any studies about how temperature and precipitation change may interactively affect environmental tracking (e.g. via changes in snow cover)? If not, I think this could be highlighted in the future directions paragraph more explicitly. Just out of curiosity, would it be possible to include such interactive effects of multiple resources in the model?

We appreciate this comment and completely updated Figure 2 to address it, working to show that both temperature and precipitation are likely critical for many organisms. We have added citations to interactive changes in climate (line 355) and now state (line 448-line 450) “Additionally, climate change has decoupled historical relationships between precipitation and temperature in some systems (e.g., Cook & Wolkovich, 2016; Wadgymar *et al.*, 2018).” We also now mention megadroughts and pluvials (line 413) and have altered our final sentence (line 419). We mention snowpack in model Box (line 429 and line 540) and do believe it could be addressed in the model by developing a more complex environment and cue system. On evolutionary timescales this question is addressed somewhat in some models, for example in Chevin & Lande (2015), which we now cite (line 354).

This question could also addressed in our model by changing the size of the resource pulse—which could be considered to model the flush of water and soil nutrients at the start of many snowpack controlled systems—and its abiotic loss rate (ϵ), which would be higher with increased temperatures (and hence higher evaporative loss) in many systems. We included one version this simulation in a previous draft of the manuscript but removed it to focus the manuscript in response to previous reviewer concerns. It would be a great area of future study for a manuscript focused on such interactive effects (and where greater exploration of parameter space for this question would be possible).

L 502 Not only temperature is rising but we already and will face non-homogeneous but fundamental differences in the precipitation regime around the globe

Good point, we now write “in the altered climates of our future” line 419).

Referee 3 comments:

In a review piece, Wolkovich & Donahue comprehensively present the idea of environmental tracking by species in stationary and non-stationary environments. This review is loaded with information and touches on several fundamental ecological ideas in relation to environmental tracking by species. The effort therefore is commendable with a potential to motivate new research avenues for climate change ecology-particularly the phenology research. Having said that, I also struggled at various places to grasp the core idea authors were intending to communicate. I outline them below.

We appreciate the reviewer's time and comments to improve our manuscript. We agree that our previous draft was perhaps so loaded with information that the most important and salient points were lost, and we have worked to fix this as we outline below.

I definitely agree with phenology as a trait and tracking as a plasticity of this trait (lines 244-246). I also liked how authors relate the idea of subsequent trade-offs in traits owing to costs associated with plasticity. I, however, missed examples of which traits and plasticity of them are going to trade-off the most with tracking, and how these may differ in stationary and non-stationary environment. Can we also say something whether the strength of trade-offs may differ in these two environments?

This is a good point. We have addressed this in two ways. First we have re-written the section 'Predicting variation in environmental tracking in non-stationary systems.' This focuses mainly on the cues underlying tracking (and not traits) but lays out more clearly how to predict how well species will track non-stationarity—a first step to understanding trade-offs. Second, we now highlight our example model in the main text of (line 266-line 274), which shows that the trade-off space narrows and tracking is more favored in this non-stationary example. Beyond that, we are not sure there is enough empirical or theoretical evidence for stronger or more specific predictions, which we have tried to outline in our Future Directions section 'What major traits trade-off with tracking?' The question of plasticity in other traits is especially interesting—it might be possible to make some predictions if we understood tracking better (e.g., how much does it reduce the *experienced* environment for certain events across years or generations)—but we feel too little is currently known.

Difference in species' ability to track environmental changes as something similar to competition-colonization trade-off is further a stimulating idea (lines 273-280). I was, however, left guessing if authors modelled this at all in their theoretical frameworks. My initial impression was that figure 3 gets at this, but I am not really sure if two species scenarios in figure 3 relate one species as a competitor (lower cue) and the other as colonizer (higher cue). Can this be clarified or if possible implemented?

Good point, our example model does effectively trade-off superior colonizers (which tracking begets) with superior competitors, we now clarify this in the model box (line 547).

Line 5 (Abstract): species responses

We have changed this on line 2, which we hope is the requested change. We have also added line numbers to the abstract to help with identifying the exact change requested.

Line 12 (Abstract): through the lens of which ecological theory? Later, you mention community ecology theory. Perhaps, use the latter to be consistent.

Done, line 8.

Line 2: Perhaps, use more recent IPCC citation.

We believe this is the most recent citation from IPCC Working Group II ('Impacts, Adaptation and Vulnerability') that considers various warming levels and a full report on impacts. We now also cite the more recent report focused on 1.5 C of warming (line 18); if the reviewer is referring to another report, please let us know.

Lines 10-12: The "indirect effects of climate change" is not very clear. Why could it not be a direct effect of climate change? Please clarify.

Good point, we have changed to fitness consequences (line 28).

Line 21: Can you elaborate which foundational ecological theory is meant here?

This was unnecessarily vague; we have changed to 'community assembly theory' (line 38), and the full sentence now reads, "Further, there has been comparatively little work connecting tracking to community assembly theory, which shows temporal sequencing and environmental variability can alter the relative fitness and niche differences between species that determine coexistence, suggesting important ecological constraints to tracking."

Line 43: Which basic community ecology theory? Please be specific when mentioning a theory as you did in lines 23-26.

Done, we now write "community assembly theory—especially priority effects and modern coexistence theory" (line 60).

Lines 237-240: Would not this be a trophic mismatch case still predictable from the stationary environment? Or does this imply that trophic mismatch will not occur in the non-stationary environment? Please clarify.

Good point, we have tried to clarify this without adding too much text, the text now reads (line 180 to line 191):

Consider a simple case in which an organism's cues evolved based on a correlation between peak prey abundance and daylength: in a stationary environment the daylength cue may be fairly reliable (generally predicting peak prey abundance based on daylength, with some interannual variation), but would become unreliable, and lead to fitness declines, if warming continually advances peak prey abundance. Multivariate cues are often argued to be more reliable because they can capture multiple attributes of the environment (Dore *et al.*, 2018; Bonamour *et al.*, 2019), but they may be equally vulnerable to failure if non-stationarity decouples the cues from fundamental tracking (Bonamour *et al.*, 2019) and thus optimal fitness is no longer associated with the cue system. Predicting the outcome of non-stationarity from the stationary environment requires that researchers know: (1) the full cue system of an organism, (2) how it relates to fundamental tracking, and (3) how both that cue system and the underlying fundamental model shift with non-stationarity.

Lines 254-256: But what about the benefit side of the tracking? And which other traits those be where trade-off with tracking will be higher?

We have re-worked this entire section to address the concerns of reviewers 1 and 4, working to shorten and clarify it. While this sentence still remains (line 198-line 200) it is now presented more clearly as a contrast to the benefits of tracking. Which traits may trade-off is covered in the following section of the manuscript.

Line 309: two 'the's

Fixed (line 260).

Lines 386-388: Please use this example as a separate sentence.

Done (line 355).

Lines 408: Please provide more example studies when you suggest "many studies".

Done (line 375).

Line 416: Here one or two examples will help the readers.

In revising to address reviewer 1's request to significantly shorten this section and focus the

paper we have removed this section.

Referee 4 (Ally Phillimore) comments:

There are some very interesting ideas in this review on phenological responses to environmental change and I can see it making a stimulating contribution. However, there are a lot of aspects that require attention, including the structure. In general I found the ms rather imprecise in its use of terminology and quite verbose. I hope the comments below are useful in revising the ms. I have not really commented on the coexistence theory aspect as I am not sufficiently familiar with this literature.

We are glad the reviewer thinks this piece could make a stimulating contribution, and agree that there was room for streamlining and conciseness, and greater precision in our language. We have worked to address these issues and explain them in more detail below.

My biggest criticism of the ms is that the term ‘environmental tracking’, which is central to the ideas being developed is not clearly defined, despite having a section devoted to its definition. A clear definition is provided for ‘fundamental tracking’, but then the text switches to environmental tracking without providing a definition (except in fig 2). This term seems to be applied more loosely to any case of phenological change, but initially without any discussion of what the yardstick is (Visser and Both 2005), meaning that its unclear that ‘tracking’ is taking place, for instance the response could be maladaptive. The yardstick for tracking could (from hardest to quantify to easiest) be the rate at which (i) the optimum is changing (as in Chevin’s B or the author’s fundamental tracking); (ii) a resource is shifting or (iii) the environment is changing (Amano et al. 2014). Related ideas are introduced from line 100, but you might consider introducing them sooner. Overall I found sections 1.1 and 1.2 quite muddled. I think ‘environmental tracking’ as used in these sections is synonymous with how the existing literature would refer to ‘phenological responses’ to cues (line 109), and I don’t see that introducing new terminology brings something useful to the table unless there is also some discussion of how much the environment is shifting, i.e. we need to know something about what is being tracked. Another concern is that introducing new poorly defined terms will just generate greater confusion in the field.

This was also a concern of Reviewer 1 and something we have struggled with (and the literature clearly has as well). One thing we struggled with is how broad the definition of phenology needs to be to include the diversity of events we include in the paper and to which we believe the topic of the paper applies. We have tried to clarify this in several ways. We have changed the title to be more specific without (hopefully) being jargony and we now try to address this head-on when we define tracking—we now state (line 73-line 75):

Both these definitions are readily applied to phenology—the timing of recurring life history events—though they can also apply to non-recurring life history events (e.g.,

seed germination), or events not normally defined as part of life history.

We now provide a new Figure 2 to clarify our definitions and we have overhauled the text where we define environmental tracking (line 66-line 73):

Conceptual and theoretical treatments of tracking often relate how well an organism matches the timing of a life history event to the ideal timing for that event, what we refer to as ‘fundamental tracking’. In contrast, empirical studies of tracking often focus on estimating a change in the timing of an event relative to a measured environmental variable, with the aim of measuring what we refer to as ‘environmental tracking’ (Fig. 2)—the change in timing of a major biological event due to an organism’s cue system given change in the environment (though most studies lack the required knowledge of the underlying cue system, Chmura *et al.*, 2019).

We have then restructured this section to contrast fundamental tracking and ‘environmental tracking,’ which agree with yardsticks (i) and (iii) of the reviewer. We further clarify what we mean by environmental tracking (line 90-line 102):

Environmental tracking depends on the intersection of the environment’s variability—which aspects of the environment vary, how (e.g., temporally each year, spatially at x scale) and how much—and an organism’s response to the environment via its proximate cues. If the varying components of the environment are not in the organism’s set of cues, then the organism does not ‘track’ per this definition (although covariation with other environmental components might give the appearance of tracking). Environmental tracking at the individual-level is a purely plastic response to environmental variation (in line with current findings on most climate change responses, Bonamour *et al.*, 2019), with the plasticity itself an outcome of selection (Chevin *et al.*, 2010) through its connection to fundamental tracking (Fig. 2). At the population-level, tracking may also incorporate evolutionary change—change in genotype frequencies—depending on both the timescales of study and the species’ generation time (this evolutionary response can be predicted as the difference between the environmental sensitivity of phenotypic selection and an organism’s plasticity, $|B - b|$ in Chevin *et al.*, 2010).

We have avoided yardstick (ii) purposefully and attempt to address that in this section as well, when we write (line 78-line 84):

This is a foundational concept of the trophic mis-match literature (Visser & Gienapp, 2019), which often assumes the peak timing of a resource defines the ideal timing for phenological events dependent on that resource (e.g. egg laying dates dependent on caterpillar abundance, Visser & Both, 2005). For most events, however, fitness outcomes are likely dependent on a suite of interacting forces (e.g., Reed *et al.*, 2013)—for example, egg laying dates for migratory birds may depend both on the timing of peak prey abundance and the need to leave nesting grounds before winter.

This is a tricky topic and it's why we believe this paper would be useful to the field, but we appreciate we need to be exact and clear and we hope our updates to the text and Figure have addressed this problem.

The section 1.3 on 'understanding variation in environmental tracking' is rather long and doesn't offer up novel perspectives. I think it could be greatly reduced by briefly summarising some of the theoretical literature on the evolution of plasticity in response to environmental cues.

Agreed, we have overhauled this section and shortened it considerably, see line 136-line 170. We have especially shortened sections 1.1-1.3 to be more precise and shorter (this material is now covered in sections 1.1-1.2, which present the relevant material, but in a space 35% shorter compared with our previous submission).

I was surprised to see plasticity really only mentioned half way through the review (around line 245), given that along with any shifts in the environmental cues, this is the most important determinant of the phenological response at least in the short/medium term. I suggest that this could be mentioned earlier when you define 'environmental tracking'. For instance, you could briefly outline the processes that can allow tracking, which I think are plasticity at the individual level, adaptation at the population level and species sorting at the community level. In lines 65-66 the mechanism underpinning a plastic response is defined and you might draw attention to that.

Agreed, we now mention plasticity much earlier (line 95 to line 97) and here we discuss also the population level:

Environmental tracking at the individual-level is a purely plastic response to environmental variation (in line with current findings on most climate change responses, Bonamour *et al.*, 2019), with the plasticity itself an outcome of selection (Chevin *et al.*, 2010) through its connection to fundamental tracking (Fig. 2). At the population-level, tracking may also incorporate evolutionary change—change in genotype frequencies—depending on both the timescales of study and the species' generation time (this evolutionary response can be predicted as the difference between the environmental sensitivity of phenotypic selection and an organism's plasticity, $|B - b|$ in Chevin *et al.*, 2010).

We have also overhauled the entire section on this (mentioned just above) so that it opens with plasticity theory and focuses mainly on this.

I like the section on Tracking in Multi-Species Environments, as this is the first part of the ms that introduces some novel perspectives. I think the ms would be improved if the preceding components were edited down, so that you get to this point much sooner. In general I thought the second half of the ms was more stimulating and well-explained than the first.

Thanks, this was a good point. As mentioned above we have cut the earlier sections by 35% to get to this section sooner (and overall shortened the manuscript by 20%). We also have made edits to the abstract and introduction to clarify our focus on this topic, laying out the earlier sections as important background.

Minor Comments

Environmental tracking: Where this idea is introduced (line 45) I think it might help to begin at the population level with a clear evolutionary definition of environmental tracking, where —B-b— is small following the equation in Chevin et al. 2010.

We have worked to clarify our definition of environmental tracking and now specifically include this equation when defining the component of it due to evolution (see line 95 to line 102, quoted above in two places).

Line 6. What proportion? The cited paper by Cook et al. is just about phenology so doesn't support the general point. A paper by Amano et al. 2014 finds that UK plant species that shift less in terms of phenology have a greater tendency to range shift. I think this finding has been replicated in other systems but can't remember the reference.

We now cite Amano *et al.* (2014) on line 23 We also worked to find other studies that incorporate at once range and phenological change, including reaching out to colleagues when we struggled to find citations. Several colleagues mentioned they are working on projects related to this topic, but have not published them and generally did not recommend other citations. The most relevant paper we found was Socolar *et al.* (2017), which we now also cite, though this paper does not provide species-specific estimates. We would be happy to include other citations if suggested, but our research suggests this may be a broad area in need of further work.

Line 14. And evolutionary theory, particularly Chevin et al. 2010 PLOS Biol.

We now cite Chevin *et al.* (2010) on line 33.

Line 15. I think the terminology in this sentence is confusing. From an evolutionary biology perspective plasticity has a clear meaning (a change in genotype's phenotype in response to the environment), but here I think it is being used to more vaguely imply flexibility, and I think 'flexibility' would be a less loaded term. Also note that tracking can involve evolution.

Agreed, we now say “phenotypic flexibility” on line 27.

Line 29-36. I agree that climate change has greatly exacerbated the non-stationary aspect of climate, but looking at historical records it seems as though climate is often somewhat non-stationary.

Agreed, we discuss this in the Box on ‘Environmental variability & change.’

Line 56. I think a more precise/mathematical definition of cue quality could be helpful, e.g., something based on the sum of squares between optimum and actual event timing (RMSE?). Also note that the literature on the evolution of plasticity uses the term ‘cue reliability’ to refer to the correlation between the environment of development and the environment of selection.

We have worked to address this through edits to the section ‘Defining tracking’ and in a new Figure 2. We agree cue reliability is important; we now define it on line 88, and explain its importance on line 178-line 191:

Critical to predictions is whether cue systems maintain their reliability with change; i.e., whether they continue to yield high fundamental tracking (Bonamour *et al.*, 2019). Consider a simple case in which an organism’s cues evolved based on a correlation between peak prey abundance and daylength: in a stationary environment the daylength cue may be fairly reliable (generally predicting peak prey abundance based on daylength, with some interannual variation), but would become unreliable, and lead to fitness declines, if warming continually advances peak prey abundance. Multivariate cues are often argued to be more reliable because they can capture multiple attributes of the environment (Dore *et al.*, 2018; Bonamour *et al.*, 2019), but they may be equally vulnerable to failure if non-stationarity decouples the cues from fundamental tracking (Bonamour *et al.*, 2019) and thus optimal fitness is no longer associated with the cue system. Predicting the outcome of non-stationarity from the stationary environment requires that researchers know: (1) the full cue system of an organism, (2) how it relates to fundamental tracking, and (3) how both that cue system and the underlying fundamental model shift with non-stationarity.

Line 62. Do you simply mean that in different locations if the individuals have the same reaction norms but environment differs then the outcome will differ? This could be explained in clearer language. Also there is a large literature by the likes of Scheiner, Lande, Chevin, Tufto, Hadfield on the evolution of cues and plasticity that goes uncited here.

In streamlining the manuscript we have deleted this sentence, though throughout this section we do cite most of the mentioned authors.

Line 64-67. Here the definition of tracking seems to be at odds with the evolutionary literature. The mechanism described is a plastic response to a cue, whereas in evolutionary biology tracking is usually with respect to a fitness optimum. This also seems to be at odds with your definition

of ‘fundamental tracking’ (line 48-49).

We have worked on this, please see our related comments above, and updated section line 65-line 102

Line 67. The organism is only expected to track the optimum proportional to the correlation between the environment of development and environment of selection.

In streamlining the manuscript we have deleted this sentence.

Line 84. Here you outline a series of papers that present information on phenological responses to temperature. However there is an absence of information on what the ‘fundamental tracking’ or shifts in the optimum are doing. I think various methods exist for generating a yardstick (Visser and Both 2005) for fundamental tracking. One option is to use the response of resources. Alternatively, the estimation of the “environmental sensitivity of selection” (Chevin 2010) and use of this in prediction is an informative avenue (Vedder et al. 2013, Gienapp et al. 2013). We also use a space for time approach to estimate tracking of the optimum in plants (Tansey et al. 2017). In terms of environmental tracking another interesting perspective is that presented in the Amano et al. paper I mention above.

We agree this was not clear, and we have addressed this by being much clearer about where different types of papers fall given our definitions (line 66 to line 73), where we write:

Conceptual and theoretical treatments of tracking often relate how well an organism matches the timing of a life history event to the ideal timing for that event, what we refer to as ‘fundamental tracking’. In contrast, empirical studies of tracking often focus on estimating a change in the timing of an event relative to a measured environmental variable, with the aim of measuring what we refer to as ‘environmental tracking’ (Fig. 2)—the change in timing of a major biological event due to an organism’s cue system given change in the environment (though most studies lack the required knowledge of the underlying cue system, Chmura *et al.*, 2019).

In the following paragraphs (though line 102) we have worked to be more explicit in our definitions of fundamental and environmental tracking.

Line 90-92. With respect to consumers tracking prey is this just the phenological shift shown? Here I think there is an opportunity to quantify whether tracking is adaptive (based on Ghalambour et al’s 2007 definitions of adaptive plasticity).

We have worked to address this in the updated section on ‘Defining tracking.’ In streamlining the section the reviewer refers to have we have deleted this sentence.

Line 174. See also Reed, T. E., Jenouvrier, S., & Visser, M. E. (2013). Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. Journal of Animal Ecology, 82(1), 131-144.

This is a great citation; the text mentioned here has been deleted in streamlining the paper, but we now cite this paper when discussing the complexity of total fitness and defining a relevant ‘yardstick’ (line 82 and line 111).

Line 201. See Chevin et al. 2015.

This is an interesting paper and outlines the challenges of predicting cues, their fitness consequences as well as measuring them. Employing a multivariate selection environment they find hyper-adaptation that can appear maladaptive if viewed via only one axis. We have clarified on this line that we are referring to empirical studies (line 153), while Chevin & Lande (2015) is theoretical (we now cite it on line 354). This comment, however, highlighted that we were not clear enough about the assumptions we were making when referring to multivariate cues, we have now worked to clarify this (line 184-line 188), where we write, “Multivariate cues are often argued to be more reliable because they can capture multiple attributes of the environment (Dore *et al.*, 2018; Bonamour *et al.*, 2019), but they may be equally vulnerable to failure if non-stationarity decouples the cues from fundamental tracking (Bonamour *et al.*, 2019) and thus optimal fitness is no longer associated with the cue system.”

Line 250. Is there a theory reference for this? I would have thought that the plastic response to each multivariate cue would be lower than the response to a single reliable cue.

This is a good point, we were implicitly assuming multivariate cues yield greater reliability than a single cue; in re-organizing this section as the reviewer requested we have deleted this point, but we have tried to clarify our reasoning and provide references in the manuscript as noted above.

Line 253. Evidence that the most plastic species have fared best, Willis et al.

In re-organizing this section as requested we have deleted this sentence, but we cite Willis *et al.* (2010) several times in the current draft.

Line 420. This recommendation is a bit vague. Is there something quantitative that researchers should do?

Good point, we have clarified this (line 375-line 382) as much as possible while aiming to shorten this section at the request of reviewer 1.

Box. 2. An additional challenge for observational studies is teasing apart the influence of photoperiod. This may only be possible for spatiotemporal or experimental studies.

This a good point, which we now make on line 493.