1 Introduction

Temerpature and light cue numerous biological processes including....*list here*(). These cues often interact in complex ways and a major goal of biology is to quantify their both individual and interactive effects on biological activities (). This effort has only become more critical in recent decades as this such measurements have become essential for accurately predicting organisms' response to anthropogenic global change and informing numerous mitigation and adaptation stratigies ().

It is extremely difficult to partition the individual effects of temperature and light cues and their interactions based on *in situ* observational studies because temperature and light variables tend to covary in nature (). However, well designed experiments in controlled environments (e.g. growth chambers, greenhouses, mesocosm) can be used to manipulate temperature and and light variable independently allowing for robust comparisions of their individual and interactive contributions to biological processes. Indeed this approach has provided many fruitful insights regarding temperature and light signaling over the last century () and has great potentially to continue to advance fundamental and applied biological inquiries in the decades to come ().

However, controlled environment studies have their own challenges. Experimentalists must balance biological realism with statistical inference (), experimental effort with statistical power () and account for the effects of unmanipulated or unmeasured variables (). Below we highlight a particular challenge that can arise when experiments seeking to estimate the interactive effects of temperature and daylength on biological outcomes. Our example deals with the phenology (timing of recurring life cycle events, e.g. leaf budburst, flowering) in temperatue woody plants, but the issues and solutions we present should are broadly applicable to many other orgnaisms and biological processes that utalize temperature and daylength signals () (For a brief background overview of how temperature and light interact to influence the spring phenology of woody plants see box).

We begin below by reviewing the minimim experimental elements required to robustly test interactiions between two or more variables. We then detail the problem of inference that can arise manipulating the both temperature and light in experiments and demonstrate the extent of this issue a mathmatical and experimental examples. Finally, we counclude by outlining several possible solutions for overcomming these issues.

2 Testing interactions in contolled environmental

In order to have the stastical power to partition the individual and interactive effect of two or more variables in an experiment, one must:

1. Have at minimum two treatment levels of at least two variables of interest ().

2. Treatment levels must be full factorial (Fig. **??**a.). Full factoiral design are balanced (Fig. **??**b.) and orthoginal (Fig. **??**c.); which is to say that all possible treatment combinations are applie the treatments and the treatments are independent of each other (). **??**.

These two critical elements may seem obvious but can be conspiquously absent from many published studes. In the case of woody plant phenology, despite a well established understand that both temperature and light cues interatively influence phenological progress (), we found that in recently published database of Z controlled environment studies (), only X of them manipulated both light and temperature cues in the same experimentm and only Y of those X did so with a design that was both balanced and orthoginal (see Suppliment for details). This notable dearth of robust tests of light and temperature interactions may be the related to the common limitations of time, space, and resources that expirimentalist face (), but it may equally relate to a fundemental issues that arises from these variables themselves comprising of multiple axes of variabily.

3 Axes of environmental variation in experiments and their implications

For most environmental variables, including light and temperature, there are two axes of variation that can be manipulated in an experiment.

1. Intensity: The amount or quality of of a variable. We define temperature intensity as the amount of heat present in the system (measure in degrees). In the phenology literature this measurement is generally refered to as forcing. We define light intesity as the luninosity or irradience present in the system (measured in lumens or watts).

2. Periodicity: The interval at which the intesity of the variable is applied. Hereafter, we refer to the periodcity of light as photoperiod (often used synonomously with 'daylength") and the periodicity of temperature as theromoperiod.

For spring of woody plants phenology, it is generally accepted that photoperiodicity is the primarly light cue to which plants respond (() though see () regarding light intensity and phenology). For temperature, conventionally both intesity and periodicty drive phenological activity () and several metrics, (e.g. growing dgree hours, thermal sums, growing degree days) that combine these two axes have been developed (). This assumption is well supported; under natural conditions diurnal temperature fluctuations temperate regions can be quite large in the field, and several studies have found that diurnal temperature varition tstrongly influences plant phenology (). In fact, several studies suggest that the even if theromoperiodicity is not an explicit treatment vairiable (mainipulated systematically), incorperating it in experiments is essential for translating experimental result into real world predictions ().

It follows that a common approach in phenology experiments that seems to balance prior knowlege, biological realism and experimental inferences is to vary photoperiodcity, and thermal intensity and

periodicty (). For example, a basic experiment might include a long (12 hours) and short (8 hours) photoperiod treatment and a high (30/20 C day/night) and low (20/10 C day/night)temperature treatment. Note that in this case the thermoperiodicty in not an explicity treatment (both high and low temperature treatment employ a diurnal fluction of 10 C), and is simply incorperated to enhance biological realism. At first glance, this design appears to meet the criterial of full factorial design, multiple treatment levels that are balanced and orthoginal, with mean high/low temperature treatments (25 and 15 C respectively) and long/short photoperiod treatment applied in all possible combinations.

Yet the orthogninality of this design is based on the assumption to a 12 hour thermoperiod. If, rather the thermoperiod is coupled with the photoperiod, orthogninality is violated, in that the daily mean temperature of the long/high treatment will be higher than that of the short/high treatment, and the long/low treatment slightly warmer than the short low (Fig **??**.a), because the warmer day time temperature are applied for different durations across the high temperature treatments. Ultimately, this non-orthoginality introducing covariation amoung the photoperiod and temperature treatments, making it stasticially impossible to differentiate their independent and interactive effects.

Of the X studies in the OSPREE database that manipuated both photoperiod and temperature, we found that at least Y may have this issue, suggesting that the true interactive effects of these cues on spring phenology is still quite poorly characterized. This may be in part why the relatively contribution of temperature and photoperiod cues to spring phenology remains a contentious debate in the phenology literature ().

Quantifying the uncertainty

To estimate the potential impact of this experimental artifact on estimations of cue effect sizes, we inteorgated the results of a large growth chamber phenology experiment by that employed this coupled photo-thermoperiod design. We used the simple geomotry to a plane to calculate the approximate maximum amount of the forcing (temperature) and photoperiod effect estimates that could potentially be mis-attrubuted due to the latent non-orthoginality of this design. While the original model estimates of the forcing and photoperiod effects (units are "phenological sensitivy" or Δ day of leaf out / Δ cue level) were estimate -9.5 and -4.5 caluted that as much about 3.0 units of of these effects could be miss attributed (for the full calculations see the Summliment.) Because forcing is expected to be the more dominant cue, we would expect the covariation of the variables may result in an over-estimation of the photoperiod effect and a weaker interaction estimate.

It is important to stress that this math is not meant to be rigid model correction. It is certainly possible that original model estimate approach the "true" values, but, due to the covariation of thermo and photoperiod there is no way of knowing for certain. (*Probably should think about how*

to phase all this without making it seem like the Flynn paper is bad or wrong).

Our estimate of "how much of each cue estimate could be misattributeddue to the covariation of thermo- and photoperiod" can be rephrased as a maximum prediction of the expected difference in effect size estimate from a coupled photo- and thermo-period experiment and an uncouple one in which diurnal photoperiod and thermoperiod are varied independently.

While we are aware of no experiments that explicitly compare these different designs, another later study by **?** utalized many overlapping treatment levels and species from the same sampling site and several treatment levels with the **?** study, but authors decoupled photo- and thermo-period, allowing for a reasonable comparision. We subset each dataset (publically available at HF and KNB) to include only the species shared amoung the two studies, and re-analyzed the data, to compare difference in the photoperiod and forcing effect estimates (see supplimental method). As we predicted the un-coupled design estimated a weaker (less negative) photoperiod effect, and stronger forcing and interaction effects than the coupled experimental design (**??**). It is worth saything that there may be other factors driving the differences between these experiments as the we conducted in different years, used different methods for applying an addition temperature pre-treatment ,chilling (see Suppliment), but this comparision is well matched to our mathmatical predictions and prior knowledge about how temperature and photoperiod are expect to interacting in phenology.

4 Paths Forward

In the sections above we have systematically demonstrated that experiments which covary thermoperiod and photoperiod cannot robust differentiate the individual effect of temperature and photoperiod on a spring phenology (or any other biological process) or quantify their interactive influence. Given the paucity of interactive studies in the literature, it is clear that more well designed studies will be needed to better characterize the effect of these cues. Below we offer several generalized experiment designs that improve statistical orthoginality of controlled environmentl experiment that could be further developed and adjusted to fit the needs of expeimentalist across many subfield of biology.

1. **Manipulate photoperiod and temperature intesity with no thermoperiodicity**. This approach allows for the maintainence of statistical orthoginality across treatment combinations (**??**b.). The main drawback is that this design sacrifices the biological realism of diurnal temperature variation, which may make it more difficult to tranlate estimates from experiments to real world applications.

2. **Compensitory diurnal temperature fluctions**. There are almost unlimited pairs of integers that can reduce to the same mean (e.g. 24+26/2 = 30+20 = 25) and the non-orthoginality of the mean daily temperature that arises in a coupled photo-thermoperiod design could be

corrected for by proportunatly increasing the dirunal temperature fluction of the short photoperiod treatment realtively to the long treatments factor (**??**c.). However, if the differences between day and night temperature has a meanful biological effect (), this introduces another confonding, non-orthoginal factor for interpreting temperature and photoperiod effects.

3. **Uncouple thermoperiod and photoperiod.** By varying thermoperiod and photoperiod independently (**??**d.), statistical orthoginality can be maintained accross treatment. However, this approach it may introduce new artifacts that occur from the biological rather than statistical interactions that occur between light and temperature. For example, there is evidence that increasing temperatures in the first two hours of the photoperiod can be almost as effective for stimulating shoot elongation as similar temperature increases for the whole photoperiod (**?**), and that in phenology daytime warming can be as much as three times for effect that night time warming in cueing phenology (). With this design, treatments must inherantly differ in the amount of time the warmer daytime temperature extend into the dark nighttime light regime, introducing a new axes of non-orhtoginality.

In correcting one problem, each on of these designs introduces another, which may in fact be an intrinsic property of any experimental manipulation. This fact should caution experimentalists to continue to think carefully about our designs and perhaps most importanlty, remind us to be humble in our inference, and think critically about what is, and isn't accounted for in our work.

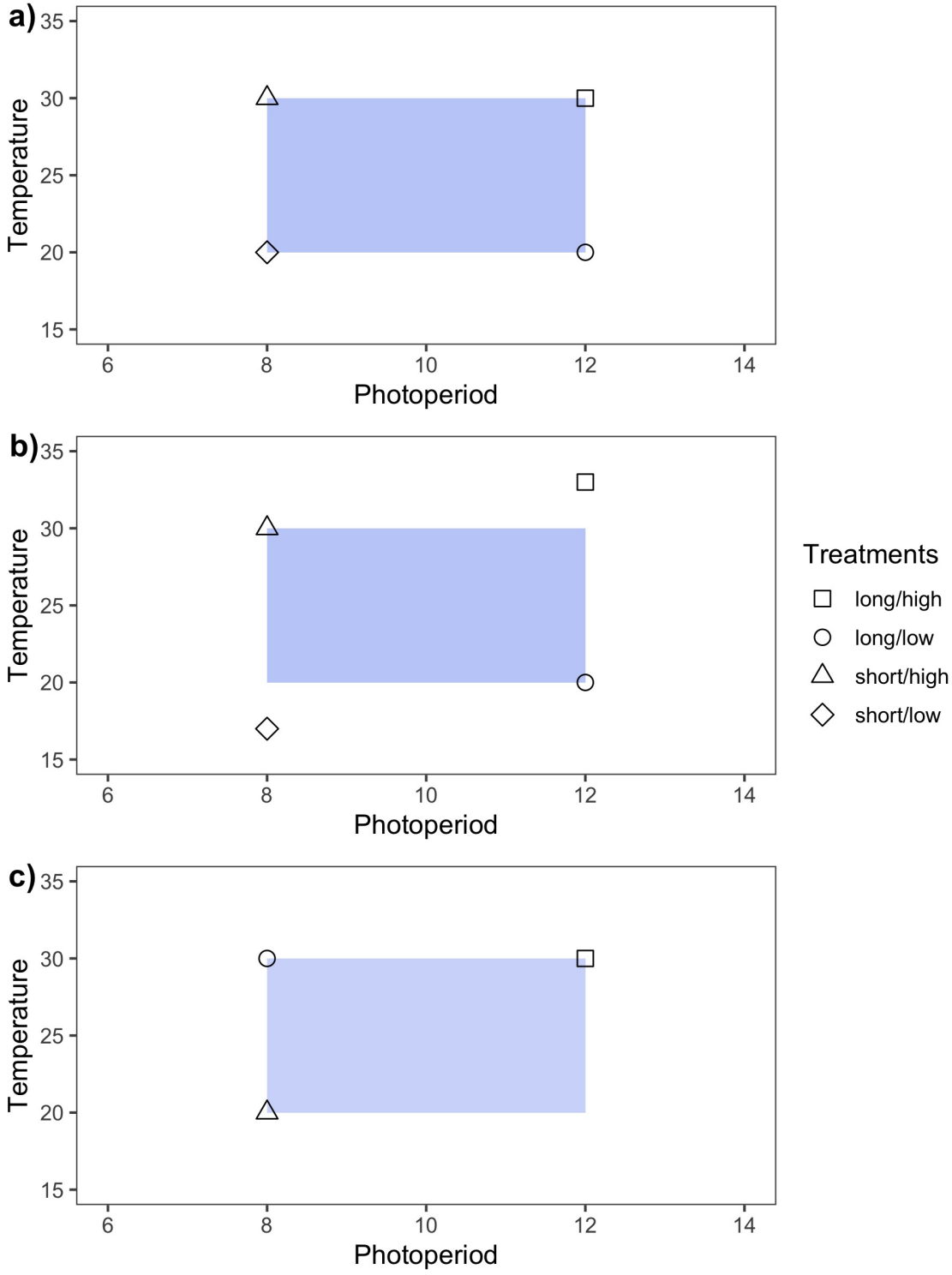

Figure 1: Idealized experimental designs demonstrate three approaches for varying temperature and light treatment level in controlled environment experiments. Design **a)** is fully factorial in that treatments levels are balanced and orthoginal. This design is appropriate for testing interactions between two or more variables. In **b)** the design is balanced both not orthoginal. Non-orthoginality in experiments often arises in experiments when there is covariation amoung the test variables is unaccounted for. In **c)**, the experimental design is orthoginal but unbalanced. Lack of balance in experiments often arises due to time, space or resource limitations.

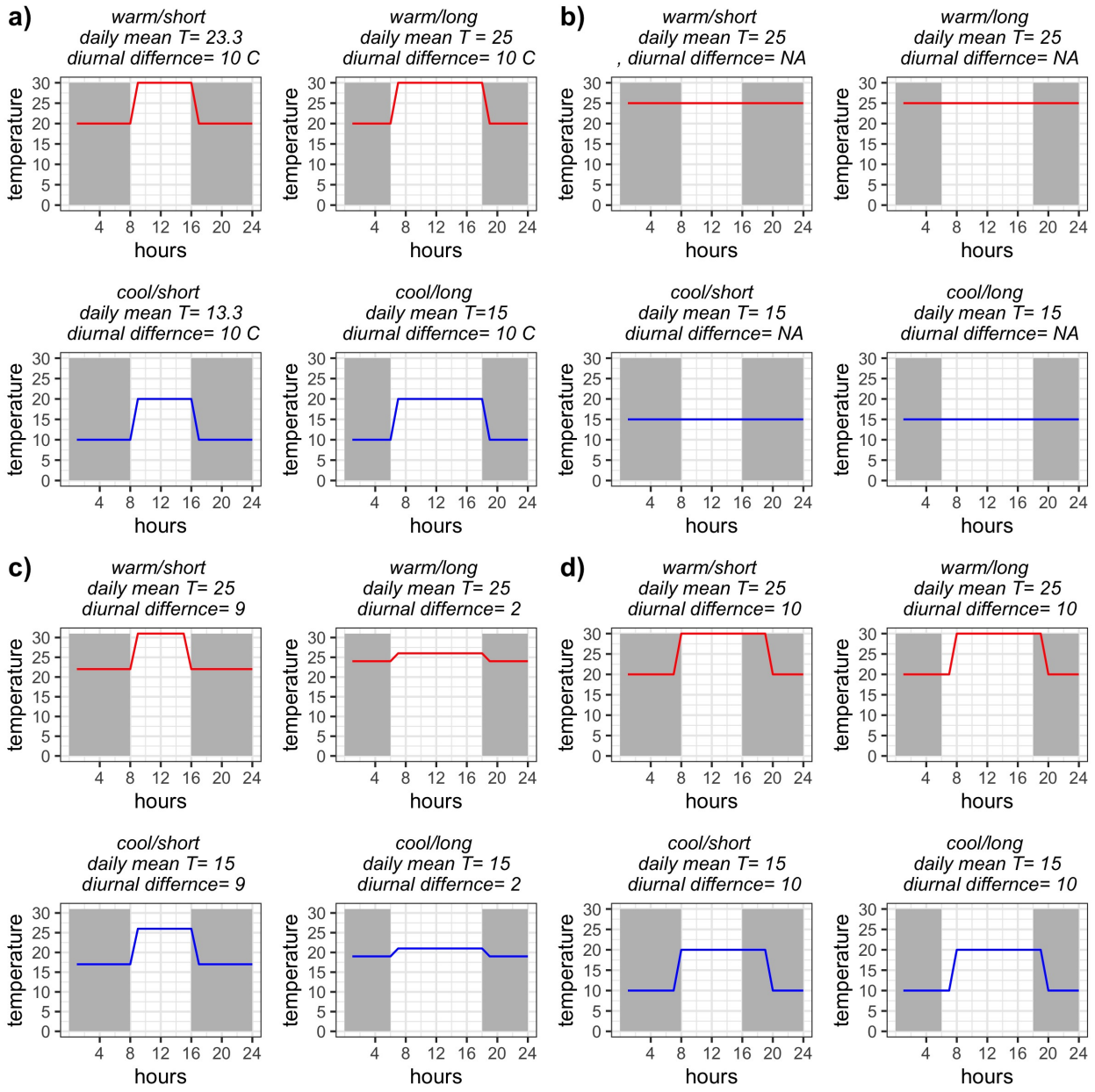

Figure 2: Conceptualized experimental designs to test temperature and daylength interactions on a biological response. In **a)** the design incorperates a standardize diurnal temperature fluctuation across all treatment. Because this thermoperiod is coupled with the photoperiod, while the same day and night temperatures are applied for the high and low temperature treatments respectivly, the mean daily temperatures differ across each photoperiod treatment generating non-orthoginality. Designs **b)**,**c)** and **d)** are all designs that can correct this non-orthoginality. Design **b)** manipulated temperature intesity only (no thermoperiodicity).In **c)** photo- and thermo- periods are still are coupled but the orthoginality of mean daily temperature is maintained by proportionately varying the diurnal temperature fluctations across treatments. In design **d)** standard diurnal temperature fluctuations are maintained but, thermoperiod and photoperiod are decoupled and varied independently, maintaining orthoginality daily mean temperatures.

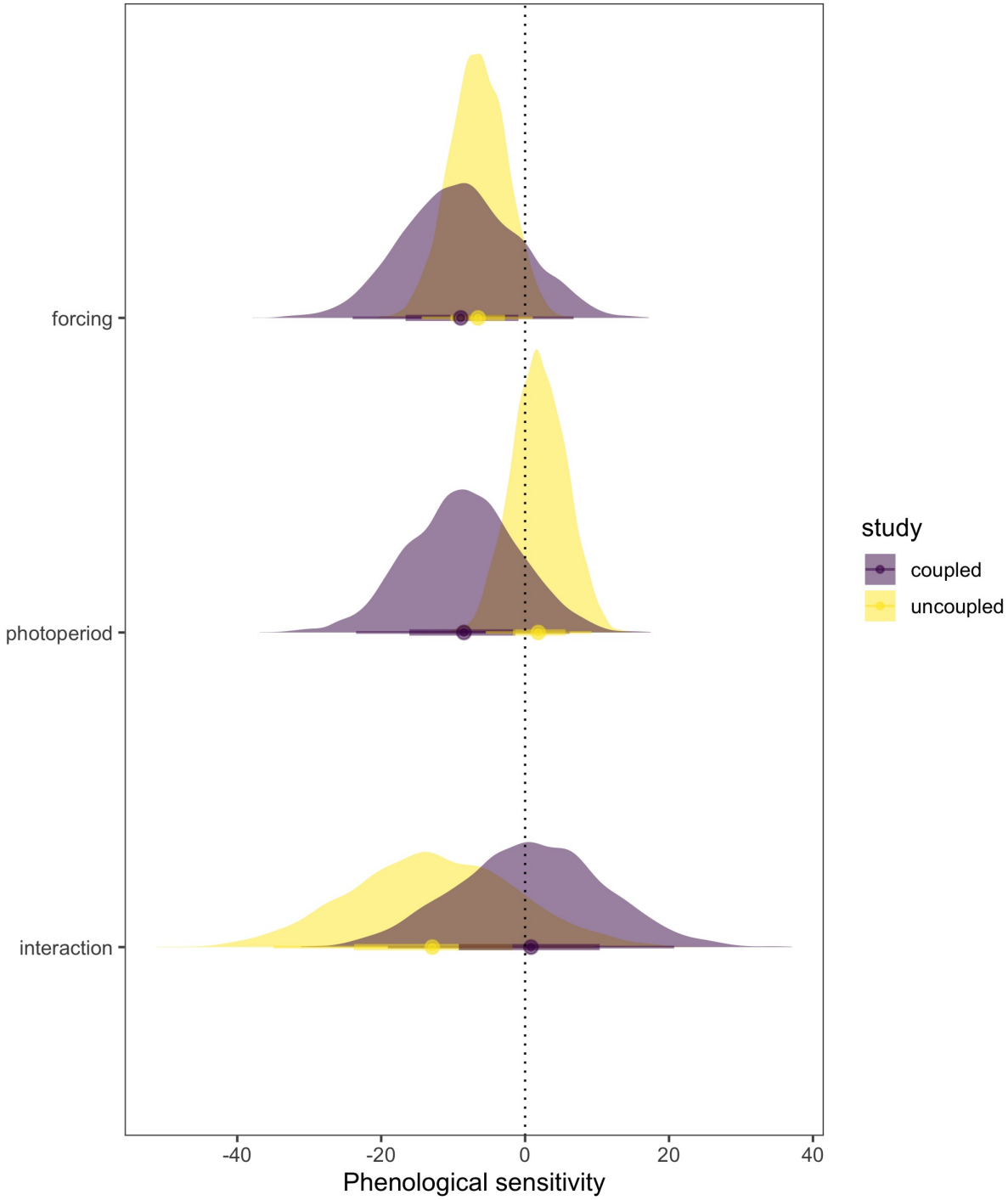

Figure 3: Need a caption but basically, estimates differs in expected ways