THE EFFECTS OF ENVIRONMENTAL VARIATION ON INDIVIDUAL VARIATION

by

Morgan P. Kain

July, 2015

Director of Thesis: Michael W. McCoy

Major Department: Biology

Despite general consensus among ecologists that variation is ubiquitous and important in nature, ecological experiments have historically focused on changes in mean response assuming a constant environment and interchangeable individuals, potentially masking important mechanisms that underlie ecological dynamics. In this thesis I present statistical and methodological advances for studying variation and results from two experiments that investigate how environmental variation affects variation in individual phenotypes. In chapter one I present simulation-based power analyses for generalized linear mixed effects models that I designed, which detect how natural or manipulative treatments affect variation in responses, such as among- and within-individual variation. My results indicate power to detect differences in variance by treatment was low overall (in most cases >1,000 total observations per treatment needed to achieve 80% power) and heterogeneity in power across ratios of individuals to repeated measures with an optimal ratio that differed by target variance parameter. With these power analyses I hope to inspire novel experimental designs in ecology and evolution investigating the causes and implications of individual-level phenotypic variance. In chapter two I evaluate the effects of variation in predation risk by crayfish (*Procambarus sp*.) on among- and within-individual variationas well as average anti-predator behavior of freshwater snails (*Physa acuta*) using novel design that isolates changes in variance from changes in the mean. I found that both the type and magnitude of environmental variation cause changes in mean behavior and variation in individual behavior. Changes in the average response versus changes in how variable individuals are can have important implications for the evolution of behavioral reaction norms, and for the maintenance of variation in populations.

THE EFFECTS OF ENVIRONMENTAL VARIATION ON INDIVIDUAL VARIATION

A Thesis

Presented To the Faculty of the Department of Biology

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Biology

by

Morgan P. Kain

July, 2015

THE EFFECTS OF ENVIRONMENTAL VARIATION ON INDIVIDUAL VARIATION

by

Morgan P. Kain

APPROVED BY:

DIRECTOR OF
THESIS: _____
Michael W. McCoy, PhD

COMMITTEE MEMBER: _____
Jeffrey S. McKinnon, PhD

COMMITTEE MEMBER: _____
David G. Kimmel, PhD

COMMITTEE MEMBER: _____
Xiangming Fang, PhD

CHAIR OF THE DEPARTMENT
OF BIOLOGY: _____
Jeffrey S. McKinnon, PhD

DEAN OF THE
GRADUATE SCHOOL: _____
Paul J. Gemperline, PhD

ACKNOWLEDGEMENTS

None of this work could have been completed without the direction of my thesis advisor Dr. Michael W. McCoy. I would like to thank the members of my committee, Dr. Jeffrey McKinnon, Dr. David Kimmel, and Dr. Xiangming Fang for their intellectual contributions, patient questions and manuscript edits. I would also like to thank Dr. Ben Bolker for his collaboration and guidance. Finally, I would like to thank the 2013 and 2014 cohorts of the Michael McCoy and Krista McCoy labs, whose support has been invaluable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Variation is an intrinsic aspect of natural systems and has pervasive influence on our understanding in nearly all areas of ecology. For example, variation in climatic conditions, resource abundance, and disturbance can strongly affect the structure, dynamics and persistence of populations (Chesson, 1986; Menge & Sutherland, 1987; Stacey & Taper, 1992; Pigliucci, 2001; Ghalambor et al., 2010; Wolf & Weissing; 2012, Dingemanse & Wolf; 2013) and the function of ecosystems (Field et al., 1992; Swetnam & Betancourt; 1997, Levin; 1998, Hooper et al.; 2005). Despite general consensus among ecologists that variation is ubiquitous and important in nature, ecological experiments have historically focused on changes in mean response assuming a constant environment and interchangeable individuals, potentially masking important mechanisms that underlie ecological dynamics (Benedetti-Cecchi, 2003; Grimm & Railsback, 2005; DeAngelis & Mooij, 2005; Bolnick et al., 2011; Wolf & Weissig, 2012). Focus on the mean is primarily attributable to historical inertia in both ecology and statistics that considered variation to be uninteresting noise that masked our ability to detect a signal from the data; however, progress has also been inhibited due to difficulties in experimental design conflating changes in variance and changes in means (e.g. Navarrete, 1996; McCabe & Gotelli, 2000; Sih & McCarthy, 2002; Van Buskirk et al., 2002; Pecor & Hazlett, 2003), failure to appreciate how variation can affect ecological conditions (Ruel & Ayres, 1999) and complex and computationally intensive statistics for appropriate modeling of hierarchical variance (e.g. [G]LMMs: [generalized] linear mixed models and [D]HGLMs: [double] hierarchical generalized linear models).

Recently however, there has been an upsurge in the use of experimental design that isolates changes in variation from changes in the mean, which has led to a deeper understanding

of the importance of environmental variance (temporal variance in an environmental variable*sensu* Lawson et al., 2015) on, for example, community composition (Benedetti-Cecchi et al., 2006; Bertocci et al., 2007), organism morphology (Miner & Vonesh, 2004; Schoeppner & Relyea; 2009), and physiological tolerance to environmental stressors such as temperature (Estay et al., 2011; Foray et al.,2014; Paijmans et al., 2013; Vasseur et al., 2014). Additionally, increased access to [G]LMM and [D]HGLM (Gelman & Hill, 2006; Bolker et al., 2009; Zuur et al., 2009; Ronnegard et al., 2010; Martin et al., 2011; Dingemanse & Dochtermann, 2013; Schielzeth & Nakagawa, 2015) and expanded awareness of the presence of substantial phenotypic variation at the level of the individual (Nussey et al., 2007; Dingemanse et al., 2010) has led to substantial work quantifying organismal variation at multiple levels (e.g. among-individual variation in average behavior: e.g. Martin & Reale, 2008, Westneat et al., 2011; plasticity: e.g. Mathot et al., 2011, Dingemanse & Wolf, 2013; and within-individual "error" variation: Stamps, Briffa & Biro, 2012; Biro & Adriaenssens, 2013; Westneat et al., 2013). Yet, despite the well-recognized ecological importance of both environmental and organismal variation, few studies examine the interaction between them (but see Mathot et al., 2012). As a result we know little about how environmental variation affects organismal variation at multiple levels.

In this thesis I investigate how environmental variation affects variation in individual phenotypes. I present both statistical and methodological advances in studying variation as well as two experiments usinga design that isolates changes in variance from changes in the mean that allows for appropriate analyses of the effects of environmental variance on individual behavioral variation. This thesis is composed of two chapters, each of which is written as a unique publishable unit that should be read as independent pieces contributing to the overall thesis goal.

At the time of completion of this thesis, chapter one is in review in *PeerJ*and chapter two is in preparation forsubmission in *Ecology*.Due to the presence of co-authors on these submitted manuscripts, "we" is used in place of "I". Table and figure numbers have been updated to correspond to the list of tables and figures presented in the table of contents above.

In chapter one I present power analyses for generalized linear mixed effects models for detecting how natural or manipulative treatments affect variation in responses, with specific emphasis on between- and within-individual variation in phenotypic responses to variable environments. With growing interest in variance as the parameter of inquiry (Moore, Brodie & Wolf, 1997; Lynch & Walsh, 1998; Benedetti-Cecchi, 2003; Hill & Zhang, 2004; Nussey et al., 2007; Dingemanse et al., 2010; Tonsor et al.,2013; Westneat et al., 2014), these power analyses fill a need for flexible simulation-based power analyses that assess power to detect differences in random effects by treatment—the magnitude of variation present among repeated measures at a specific hierarchical level (Gelman & Hill, 2006; Zuur et al., 2009).

In chapter two I describe two experiments that evaluate the effects of environmental variation on individual variation in behavior. Each experiment uses novel experimental design combining elements from Miner & Vonesh (2004), Benedetti-Cecchi (2006), and Schoeppner & Relyea (2009), and cutting-edge statistical approaches (DHGLM) that assess the importance of fixed and random effects in predicting both mean and variance structure (Lee & Nelder, 2006; Cleasby, Nakagawa & Schielzeth, 2015). Specifically, I quantified mean anti-predator behavior and both among- and within-individual variation in anti-predator responses of freshwater snails (*Physa acuta*)in response to varying magnitudes of coarse-grained and fine-grained variation in predation risk by crayfish (*Procambarus sp*.). In addition to my behavioral assays, I determined

the fitness consequences of environmental variation by quantifying *Physa* survival and reproduction.

Together this work helps to strengthen the foundation upon which novel questions about the effects and consequences of variation can be addressed. Additionally, my research indicates that increased research effort should be spent on ecological responses to environmental variation, especially in light of substantial theoretical results demonstrating increased prey suppression under pulsed predation regimes (Sih & McCarthy, 2002; Liu and Chen, 2003; Lie et al., 2006; Qian et al., 2009) and projections of changes in environmental variation due to climate change (Highes, 2000; Muller & Sotne, 2001; Luterbacher, 2004).Finally, these studies supply critical empirical support for the need to expand the focus in the field of behavioral ecology to include variation *per se* as a predictor of individual variation in behavior and towards quantifying and contrasting the magnitude of among- and within-individual variation among multiple groups of individuals.

CHAPTER 1: A PRACTICAL GUIDE AND POWER ANALYSES FOR GLMMs: DETECTING AMONG TREATMENT VARIATION IN RANDOM EFFECTS

Title: A practical guide and power analysis for GLMMs: Detecting among treatment variation in random effects

Authors: Morgan P. Kain[1*], Ben M. Bolker[2], Michael W. McCoy[1]

[1] Department of Biology East Carolina University, N108 Howell Science Complex, Greenville, NC 27858

[2] Department of Mathematics & Statistics, McMaster University, 314 Hamilton Hall, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

[*] Correspondence author. N108 Howell Science Complex East Carolina University Mail Stop 551 Greenville, NC 27858 E-mail: morganpkain@gmail.com

Abstract

In ecology and evolution GLMMs are increasingly used to test for differences in variation by treatment at multiple hierarchical levels. Yet, the specific sampling schemes that optimize the power of an experiment to detect differences in random effects by treatment/group remain unknown. In this paper we develop a blueprint for conducting power analyses for GLMMs focusing on detecting differences in variance by treatment. We present parameterization and power analyses for random-intercepts and random- slopes GLMMs because of their generality as focal parameters for most applications and because of their immediate applicability to emerging questions in the field of behavioral ecology. We focus on the extreme case of hierarchically structured binomial data, though the framework presented here generalizes easily to any error distribution model. First, we determine the optimal ratio of individuals to repeated measures within individuals that maximizes power to detect differences by treatment in among-individual variation in intercept, among-individual variation in slope, and within-individual variation in intercept. Second, we explore how power to detect differences in target variance parameters is affected by total variation. Our results indicate heterogeneity in power across ratios of individuals to repeated measures with an optimal ratio determined by both the target variance parameter and total sample size. Additionally, power to detect each variance parameter was low overall (in most cases >1,000 total observations per treatment needed to achieve 80% power) and decreased with increasing variance in non-target random effects. With growing interest in variance as the parameter of inquiry, these power analyses provide a crucial component for designing experiments focused on detecting differences in variance. We hope to inspire novel experimental designs in ecology and evolution investigating the causes and implications of individual-level phenotypic variance, such as the adaptive significance of within-individual variation.

Key-words: individual variation, behavioral ecology, reaction norm, plasticity, binomial distribution, hierarchical, sampling scheme

Introduction

Recent advances in computing power and access to increasingly sophisticated statistical tools such as generalized linear mixed effects models are changing research in ecology, evolution and behavior. Research questions and data analyses are no longer confined to the assumptions of clean experimental designs based on agricultural plots and Normal error distributions. Researchers now commonly incorporate multiple levels of hierarchical nesting (e.g. repeated measures) and can analyze data using a wide array of non-Gaussian error distribution models. This change is epitomized by the recent increase in use of linear and generalized linear mixed models ([G]LMMs: Touchon, J. & McCoy, W.M. unpublished data). These powerful tools permit appropriate modeling of variation among groups and across space and time, allowing for more accurate extrapolation of statistical results to unobserved data, as well as statistical tests of variance components (Gelman & Hill, 2006; Bolker et al., 2009; Zuur et al., 2009; Zuur, Hilbe & Leno, 2013).

The upsurge in the use of LMM and GLMM has been facilitated by several recent methods papers (Bolker et al., 2009; Martin et al., 2011; Dingemanse & Dochtermann, 2013; Schielzeth & Nakagawa, 2013) and textbooks (Gelman & Hill, 2006; Zuur et al., 2009; Zuur, Hilbe & Leno, 2013; Bolker, 2015) specifically aimed at non-statisticians. While these resources have accelerated the adoption of these tools, there are still too few resources guiding researchers through the choices that must be made *prior to* the initiation of a new experiment, such as the sampling scheme that will optimize the power of an experiment requiring analysis by linear (Moineddin, Matheson & Glazier, 2007; Scherbaum & Ferreter, 2009; Martin et al., 2011) and generalized linear (Johnson et al., 2014) mixed models. In this paper, we develop a blueprint for conducting power analyses for GLMMs using the lme4package (Bates et al., 2014) in the R

statistical programming environment (R Development Core Team, 2015). We focus on a specific application aimed at detecting differences in variance by treatment at multiple hierarchical levels.

Power analysis is fundamental to good experimental design, but is often overlooked (Jennions & Moller, 2003), or in the case of GLMMs, simply too difficult to implement for many practitioners. Power analyses can be especially daunting for GLMMs because they require large simulations with complex, non-Normal and non-independent data structures (Johnson et al., 2014). In this paper we take advantage of recent developments in the lme4 package in R that simplify the process of simulating appropriate data. Despite the increasing use of GLMMs in ecology and evolution and growing interest in variance, we are aware of no papers that present power analyses for statistical tests on variance using GLMMs, and only one paper presenting power analyses for fixed effects in GLMMs (Johnson et al., 2014). Indeed, Johnson et al.'s (2014) analysis illustrates that power analyses conducted for hierarchically structured experiments that do not incorporate random effects can generate biased estimates of fixed effects, highlighting the need for a better understanding of these approaches.

While most applications of GLMMs to date have focused on detecting differences in fixed effects while appropriately accounting for random effects (e.g. Johnson et al., 2014), GLMMs are under rapid development and many new applications are now possible (e.g. modeling heterogeneous error variance: Kizilkaya & Tempelman 2005, Cernicchiaro et al., 2013). With growing interest in variance as the parameter of inquiry (Moore, Brodie & Wolf, 1997; Lynch & Walsh, 1998; Benedetti-Cecchi, 2003; Hill & Zhang, 2004; Nussey, Wilson & Brommer, 2007; Dingemanse et al., 2010; Tonsor, Elnaccash & Scheiner, 2013; Westneat, Wright & Dingemanse, 2014), there is an increased need for accessible, flexible simulation-based power analyses that

assess power to detect differences in random effects by treatment—the magnitude of variation present among repeated measures at a specific hierarchical level (Gelman & Hill, 2006; Zuur et al., 2009).

Here we present parameterization and power analyses for random-intercepts and random-slopes GLMMs that test for differences in variation by treatment in three key parameters: 1) Among-group variation in intercept; 2) Within-group variation in intercept; 3) Among-group variation in slope. We examine each of these comparisons in two contexts. First, we describe the optimal ratio of groups to observations within groups that maximizes power to detect differences in each variance parameter. In experiments with binomially distributed response variables, observations within groups are organized into *j* sampling occasions, eachcontaining *n* Bernoulli observations. Here we discuss the ratio of groups to total observations within groups (*n***j*), and consider different partitions of *n* and *j*. Second, we explore how power to detect differences in specific variance parameters is affected by increasing variation in non-target parameters (e.g., how power to detect differences in among-group variation decreases as within-group variance increases). We consider both random-intercepts and random-slopes models because of their generality as focal parameters for most applications, and choose to focus on the extreme case of hierarchically structured binomial data because binary response data (e.g. the presence or absence of a behavior) contains the least possible amount of information per observation and yet is a common data format for a variety of endpoints measured in ecology.

We use vocabulary and examples from behavioral ecology to illustrate our models because of their immediate applicability to emerging questions in this field. Specifically, we evaluate power to detect significant differences in among-individual variation in reaction norm intercept and slope, and within-individual variation in intercept between groups of individuals (Nussey,

Wilson & Brommer, 2007; Dingemanse et al., 2010). Our methods extend current approaches used in behavioral ecology for quantifying among-individual variation away from simply testing whether there is significant deviation from a null model of no variation (Martin et al., 2011; Van de Pol et al., 2012; Dingemanse & Dochtermann, 2013) toward quantifying and contrasting the magnitude of among- and within-individual variation among multiple groups of individuals.

In an effort to present a framework that is customizable for a diversity of research problems, we focus on a general sampling scheme in which several Bernoulli observations ($n > 1$) within multiple sampling occasions ($j > 1$) are available for each individual. Under this sampling scheme multiple probabilities of "success" (e.g. the probability of displaying a behavior) are available for each individual, which is necessary for quantifying within-individual variation (variation among sampling occasions in the probability an individual displays a behavior). However, we note that often in behavioral ecology only a single Bernoulli observation ($n = 1$) is available for each sampling occasion j. We include a description on how to modify this general case to accommodate single observations per sampling occasion in Supplement 1. Finally, while we focus on the binomial GLMM, the framework presented here generalizes easily to other error distribution models such as Normal, log-Normal, or Gamma (for continuous responses) or Poisson or negative binomial (for count responses).

Methods

Linear Mixed Model

We begin by introducing a general linear mixed model (LMM) to illustrate the variance components we are interested in (Figure 1) and their applications in behavioral ecology. We provide only a brief introduction to LMMs here because they have been extensively discussed in several recent reviews and textbooks (Gelman & Hill, 2006; Zuur et al., 2009; Stroup, 2012; Zuur, Hilbe & Leno, 2013; Dingemanse & Dochtermann, 2013; Bates et al., 2014; Bolker, 2015). We use the notation of Stroup (2012) to facilitate a transition to the binomial GLMM model, which is the focus of our power analyses.

A two treatment linear mixed model can be written as:

[1] $y_{ijk} \mid b_{0ik}, b_{1ik} \sim \text{Normal}(\mu_{ijk}, \sigma^2_{\varepsilon k})$

[2] $\eta_{ijk} = \beta_{0k} + b_{0ik} + (\beta_{1k} + b_{ik})X_{ij}$

[3] Identity link: $\eta_{ijk} = \mu_{ijk}$

[4] $\begin{bmatrix} b_{0ik} \\ b_{1ik} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{0k} & \sigma_{01k} \\ \sigma_{01k} & \sigma^2_{1k} \end{bmatrix}\right)$

Here, a single phenotypic measurement y_{ijk} of individual *i*, in environment *j* and treatment *k* is composed of three components: the treatment mean in environment *j* ($\beta_{0k} + \beta_{1k} X_{ij}$), the unique average response of individual *i* across the environmental gradient ($b_{0k} + b_{1k} X_{ij}$), and a residual error due to the variation around the mean of individual *i* ($\sigma^2_{\varepsilon k}$), which is assumed to be homogenous across *X* and among all individuals in treatment *k*, but is allowed to vary by treatment. Individuals vary from the treatment mean reaction norm in both their intercept (b_{0ik}) and slope (b_{1ik}), which together compose the total phenotypic variance attributable to among-individual variation. This individual contribution is quantified using a random intercepts and slopes model with a multivariate Normal (MVN) distribution [4]. Variation among individuals in

intercept and slope are σ^2_{0k} and σ^2_{1k} respectively; covariance between intercept and slope is given by σ_{01k}.In a LMM, the linear predictor directly predicts the mean, as shown by the identity link function in equation [3]. In a GLMM, the linear predictor predicts a function of the mean $g(x)$, which must be linearized through the use of non-identity link functions; for example, we use the standard logit (log-odds) link for Binomial GLMM.

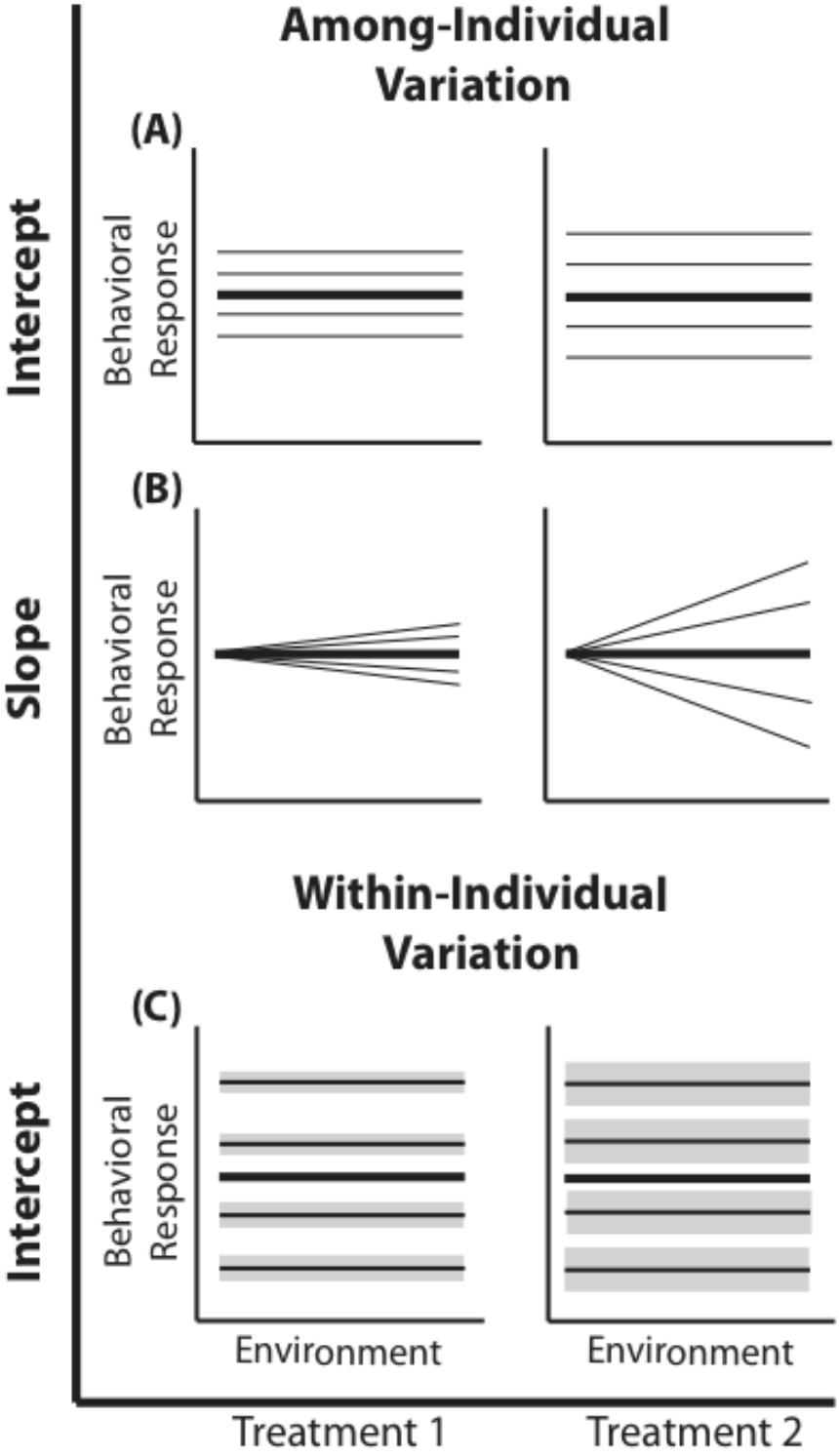

Figure 1: Reaction norm plots for a two treatment LMM. In all graphs bolded black lines depict treatment mean reaction norms and thin lines depict reaction norms of individuals. Grey envelopes in (C) illustrate the magnitude of within-individual variation. Here among-individual variation in intercept (A), slope (B) and within-individual variation in intercept (C) is larger in treatment 2.

Among-individual variation in intercept

In behavioral ecology among-individual variation in intercept σ^2_{0k} describes the amount of variation around average behavior that occurs among individuals (Figure 1). In field studies, σ^2_{0k} describes variation in individuals' average behavior in the mean-centered environment (Nussey, Wilson & Brommer, 2007; Westneat et al., 2011). Previous work has demonstrated that individuals from a diversity of taxa vary in their average behavior across different environments (Bell, Hankison & Laskowski, 2009). Yet, comparisons of among- and within-individual variation in average behavior (or other forms of plasticity) among groups, populations, or treatments remain underrepresented (e.g. Westneat et al., 2011; Dingemanse et al., 2012). For example, Westneat et al., (2011) found that female house sparrows vary less from one another in their average provisioning behavior than male sparrows. In the model presented here, the random intercept (b_{0ik}) for each individual (e.g. male and female nest provisioning rates are drawn from Normal distributions with different variances) is drawn from a treatment-specific Normal distribution.

Within-Individual Variation in Intercept

Within-individual variation in intercept ($\sigma^2_{\varepsilon k}$) is defined as the amount individuals vary around their own average behavior. Within-individual variation is routinely used for the calculation of repeatability in studies of animal personality (Bell, Hankison & Laskowski, 2009; Dingemanse et al., 2010) or more often is simply regarded as noise, despite the well established ecological and evolutionary implications of within-individual variation (Stamps, Briffa & Biro, 2012; Biro & Adriasenssens, 2013; Westneat, Wright & Dingemanse, 2014; Cleasby, Nakagawa & Schielzeth, 2015). For example, a variable predator environment may select for individual

prey that vary greatly around their mean behavior to remain unpredictable (Stamps, Briffa & Biro, 2012). LMMs can directly quantify patterns of within-individual variation when repeated measures within multiple individuals are available, facilitating comparisons of individual consistency between groups of individuals (Dingemanse et al., 2013). Here we are interested in determining if $\sigma^2_{\varepsilon k}$ differs by treatment. In other words, do individuals in one population or treatment exhibit more intra-individual behavioral variation than individuals from a second population or treatment?

Among-Individual variation in slope

Substantial empirical work has shown that individual animals in a variety of taxa display variation in phenotypic plasticity (Martin & Réale, 2008; Mathot et al., 2011; Dingemanse et al., 2012); using mixed models to quantify this variation has been the primary focus of several recent papers (Martin et al., 2011; Van de Pol, 2012; Dingemanse and Dochtermann, 2013). Among-individual variation in phenotypic plasticity has implications for the rate of evolutionary change, population stability and population persistence (Wolf & Weissing, 2012; Dingemanse & Wolf, 2013); thus defining those populations exhibiting greater individual variation in plasticity could help distinguish stable populations and populations with a high probability of micro-evolutionary change (Pigliucci, 2001; Ghalambor, Angeloni & Carroll, 2010). To quantify group differences in plasticity variation, multiple measurements within each individual across an environmental gradient are required. Here we are interested in determining if σ^2_{1k} differs by treatment.

Binomial GLMM

We assess power of a binomial GLMM for detecting differences in variation by treatment. This model can be written as:

[5] $y_{ijk} | b_{0ik}, b_{1ik,} v_{ijk} \sim \text{Binomial}(N_{ijk}, \pi_{ijk})$

[6] $\eta_{ijk} = \beta_0 + b_{0ik} + (\beta_1 + b_{1ik})X_{ij} + v_{ijk}$

[7] Inverse-logit: $\pi_{ijk} = 1/(1 + e^{-\eta ijk})$

[8] $\begin{bmatrix} b_{0ik} \\ b_{1ik} \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{0k} & \sigma_{01k} \\ \sigma_{01k} & \sigma^2_{1k} \end{bmatrix} \right)$

[9] $v_{ijk} \sim \text{Normal}(0, \sigma^2_{vk})$

Here, $y_{ijk}$is the number of "successes" in N_{ijk} observations of the *i*th individual in treatment *k* at the *j*th sampling occasion. When an environmental covariate (*X)* is present, one sampling occasion occurs at each level of the covariate *j*. In the absence of an environmental covariate, the linear predictor reduces to $\eta_{ijk} = \beta_0 + b_{0ik} + v_{ijk}$ and the *j*th occasion is simply a repeated sampling occasion in the same conditions. Note, when $N_{ijk} = 1$ there is only 1 observation per sampling occasion *j*, making y_{ijk} a Bernoulli response variable (see Supplement 1). When y_{ijk} is Bernoulli, overdispersion (v_{ijk})and thus within-individual variation is not identifiable.

In this model π_{ijk} describes the underlying probability of individual *i* in treatment *k* at occasion *j* exhibiting a behavior. Variation in $\pi$is determined by the linear combination of predictors on the logit (log-odds) scale: group intercept (β_0), group slope (β_1), individual unique intercept (b_{0ik}), slope (b_{1ik}),and observation level overdispersion that decreases predictive power at each observation (v_{ijk}). This linear predictor is transformed with the inverse logit link to produce π_{ijk}, which follows a logit-Normal-binomial mixed distribution.

We use an observation-level random effect to model additive overdispersion (Browne et al., 2005), which models increased variance (following a Normal distribution with variance σ^2_{vk}) in the linear predictor on the link scale (Nakagawa & Schielzeth, 2010). Overdispersion is used to quantify within-individual variation because it models variation in π between each sampling occasion *j* for each individual. Here the magnitude of overdispersion is allowed to vary by treatment (for an example of multiple data sets where this occurs see Hinde & Demetrio, 2007), which is a focus of our power analysis.

The transformation through the inverse-logit function makes each of the three target variance components difficult to visualize with a concise figure. However, because the binomial GLMM model follows similar patterns as the LMM, we present power analyses for the binomial GLMM using the visual aid presented for the LMM (Figure 1). Finally, we simulate data for a fully balanced design without losing generality. See Martin et al., 2011 and Van de Pol, 2012 for a discussion on experimental designs where individuals are assayed in partially overlapping environments and when only single measurements are obtained for some individuals.

Simulations

All data were simulated in the R statistical programming environment using newly developed simulation capabilities of the lme4 package (Bates et al., 2014). Guidelines for parameterizing the GLMMs and running data simulations and power analyses are provided in Supplement 1. For a given total sample size, we present simulations for determining the optimal ratio of total number of individuals versus the number of repeated measures within individuals needed to provide power to detect a difference among treatments 80% of the time. We conducted simulations for multiple ratios of individuals to total observations within individuals, varying both sampling occasions (*j*) and Bernoulli observations within sampling occasions (*n*). Next, we describe simulations that evaluate how increasing "noise" (variation in non-target random effects) affects power to detect differences in targeted variance comparisons.

For both scenarios we simulate data with biologically relevant parameter values that illustrate common trends in power. At extreme parameter values the trends presented here may not hold due to interactions between the variance components that arise at the boundaries of binomial space. We do not dwell on these exceptions since they are unrealistic for most empirical data sets, but suggest exploration of these exceptions with code provided in Supplement 1.

We ran 2800 simulations for each combination of parameter values. The significance of a given random effect was assessed using likelihood ratio tests (LRTs) between models with and without the focal random effect. To correct for the known conservatism of the LRT when testing for $\sigma^2 = 0$ (due to a null value on the boundary of parameter space), we adopted the standard correction of dividing all p-values by 2 (Pinheiro & Bates, 2000; Verbeke & Molenberghs, 2000; Fitzmaurice, Laird & Ware, 2004; Zuur et al., 2009). This correction was appropriate for all p-

values because each LRT compared models that differed in only a single degree of freedom. Power is estimated as the percentage of simulations that provide a corrected p-value smaller than 0.05. We insured the validity of a nominal p-value of 0.05 by confirming that 2800 simulations of a scenario where standard deviations did not differ at all did not result in rejecting the null hypothesis more than 5% of the time. Under extremely low numbers of individuals (~2-4) power to detect differences in the null case exceeded 5% (~10-15%), possibly inflating power in these cases. Regardless, random effects cannot be reliably estimated with such low sample sizes and therefore in most cases such experimental designs should be avoided.

Scenario 1: Determining the optimal sampling scheme

Most researchers face limitations imposed by time, money and access to samples, and are therefore confronted with the question of how resources should be divided between individuals and measures within individuals. To investigate the optimal allocation of sampling effort between the number of individuals and number of observations per individual, we simulated two data sets for each variance comparison (See Table 1 for a summary of all simulations).First, using three hypothetical total numbers of Bernoulli observations *per treatment* (total sample size per treatment, TSS_T), we manipulated either the ratio of individuals to sampling occasions (σ^2_{0k} and σ^2_{1k}), or the ratio of individuals to Bernoulli observations within sampling occasions (σ^2_{vk}). For comparisons of σ^2_{0k} and σ^2_{1k} we manipulated the ratio of individuals to sampling occasions, holding the number of Bernoulli observations constant at 5, because power follows a non-monotonic pattern across these ratios for σ^2_{0k} and σ^2_{1k} (Figures 2, 3). Conversely, for comparisons of σ^2_{vk} we manipulated the ratio of individuals to Bernoulli observations and held the number of sampling occasions constant at 5 because power follows a non-monotonic pattern

across ratios of individuals to Bernoulli observations for $\sigma^2_{\nu k}$ (Figure 4). For comparisons of σ^2_{0k}, and $\sigma^2_{\nu k}$ we simulated TSS_T of 600, 1200 and 2400, and for comparisons of σ^2_{1k} TSS_T were 300, 600, and 1200. For example, for b_{1ik} with a TSS_T of 300, the most extreme ratios were 30 individuals with 2 sampling occasions and 2 individuals with 30 sampling occasions. While using only 2 samples for a grouping variable (individuals) is never suggested for a random effect, we include this combination as an illustration of the low power that results from an ill-conceived sampling scheme. For each variance comparison we simulated three different effect sizes (2, 2.5, and 3 fold difference in standard deviation by treatment).

Table 1. Parameter values for all simulations. For example, Scenario 1: Figure 2C illustrates power to detect differences in σ^2_{0k} across ratios of individuals to sampling occasions with a TSS_T of 2,400 at effect sizes of 2x, 2.5x, and 3x difference in standard deviation by treatment.

Target Variance	σ^2_{0k}					σ^2_{1k}					$\sigma^2_{\nu k}$				
Scenario	**1**				**2**	**1**				**2**	**1**				**2**
Figure	**2A**	**2B**	**2C**	**5A**	**6A**	**3A**	**3B**	**3C**	**5B**	**6B**	**4A**	**4B**	**4C**	**5C**	**6C**
Parameter	Sampling Occasions			Bernoulli Obs	$\sigma^2_{\nu k}$	Sampling Occasions			Bernoulli Obs	$\sigma^2_{\nu k}$	Bernoulli Observations			Sampling Occasions	σ^2_{0k}
TSS_T	600	1,200	2,400	240 - 3,600	2,400	300	600	1,200	120- 1,800	1,200	600	1,200	2,400	240 - 3,600	2,400
# Individuals	2-60	2-120	2-240	120-2	2-240	2-30	2-60	2-120	60-2	2-120	2-60	2-120	2-240	2-120	2-240
# Sampling Occasions	60-2	120-2	240-2	2-120	240-2	30-2	60-2	120-2	2-60	120-2	5	5	5	1-15	5
# Bernoulli Observations	5	5	5	1- 15	5	5	5	5	1-15	5	60-2	120-2	240-2	120-2	240-2
Effect Sizes	2; 2.5; 3	2; 2.5; 3	2; 2.5; 3	2.5	2.5	2; 2.5; 3	2; 2.5; 3	2; 2.5; 3	2.5	2.5	2; 2.5; 3	2; 2.5; 3	2; 2.5; 3	2.5	2.5

Next, we simulated data sets with increasing numbers of Bernoulli observations for comparisons of σ^2_{0k} and σ^2_{1k} (Figure 5A, B) and with increasing numbers of sampling occasions for comparisons of σ^2_{vk} (Figure 5C). For these simulations we used 1, 3, 5, 10 and 15 Bernoulli observations or sampling occasions. Ratios of individuals to sampling occasions (σ^2_{0k} and σ^2_{1k}) or individuals to Bernoulli observations (σ^2_{vk}) followed the intermediate TSS_T from the simulations described above. For example, for comparisons of σ^2_{0k} we simulated 1, 3, 5, 10 and 15 Bernoulli observations for ratios of individuals to sampling occasions ranging from 120:2 to 2:120. For all comparisons we simulated data using an effect size of a 2.5 fold difference in standard deviation by treatment.

In all Scenario 1 simulations, both β_0 and β_1 were constrained to a single value for all treatments. For comparisons of among-individual variation in intercept no environmental covariate was used, and $\sigma^2_{vk}$was held constant among treatments. For comparisons of among-individual variation in slope we held σ^2_{vk} constant. Finally, for comparisons of within-individual variation in intercept, no environmental covariate was included and σ^2_{0k} was held constant among treatments. All parameter values used in simulations for both Scenarios can be found in Table A1.1.

Our goal in Scenario 1 was to isolate changes in a single variance parameter, but exploration of the dependence among multiple variance components and the mean may be warranted if it is relevant for a specific problem. Incorporating concurrent changes in intercept, slope and overdispersion parameters can be easily implemented with slight modifications to the code presented in the online supplement. We show initial results of relaxing some of these assumptions in Scenario 2, but full exploration of these possibilities are beyond the scope of this paper.

Scenario 2: Measuring the ratio of overdispersion to effect size

Decreasing the ratio of the variance in the target random effect to total variance influences power to detect differences in the target variance among treatments. Therefore, we simulated four levels of "noise" (magnitude of non-target random effect variance) assuming a Normal distribution with increasing standard deviations (0.1, 0.5, 1.0, 2.0) (Figure 6). These correspond to ratios of target variance parameter effect size to non-target variance of 25:1, 5:1, 5:2, and 5:4. For comparisons of σ^2_{0k} and $\sigma^2_{1k,}$"noise" was simulated with increasing variation in within-individual variation (σ^2_{vk}), while for comparisons of σ^2_{vk} noise was simulated with among-individual variation in intercept (σ^2_{0k}). For each variance parameter ratios of individuals to repeated measures followed the largest TSS_T sampling scheme used in Scenario 1 and an ES of a 2.5x difference in standard deviation by treatment.

Results

Scenario 1: Determining the optimal sampling scheme

Power to detect differences between treatments for each variance component increases with total sample size (TSS_T) and effect size (ES) (Figures 2-5). For a given TSS_T power depends on the ratio of the number of individuals to the number of repeated measures per individual. However, the optimal ratio of individuals to repeated measures varies depending on TSS_T and target variance parameter. For example, power to detect both σ^2_{0k} and σ^2_{1k} is non-monotonic across ratios of individuals to sampling occasions (Figures 2, 3), but is an increasing function of the number of Bernoulli observations within sampling occasions (Figure 5A, B). Power to detect σ^2_{0k} is maximized at a ratio of individuals to repeated measures of approximately 6:5 under low sample sizes ($TSS_T = 600$) (Figure 2A) but a ratio of approximately 2:1 is optimal under larger sample sizes ($TSS_T = 2400$) (Figure 2C). At low sample

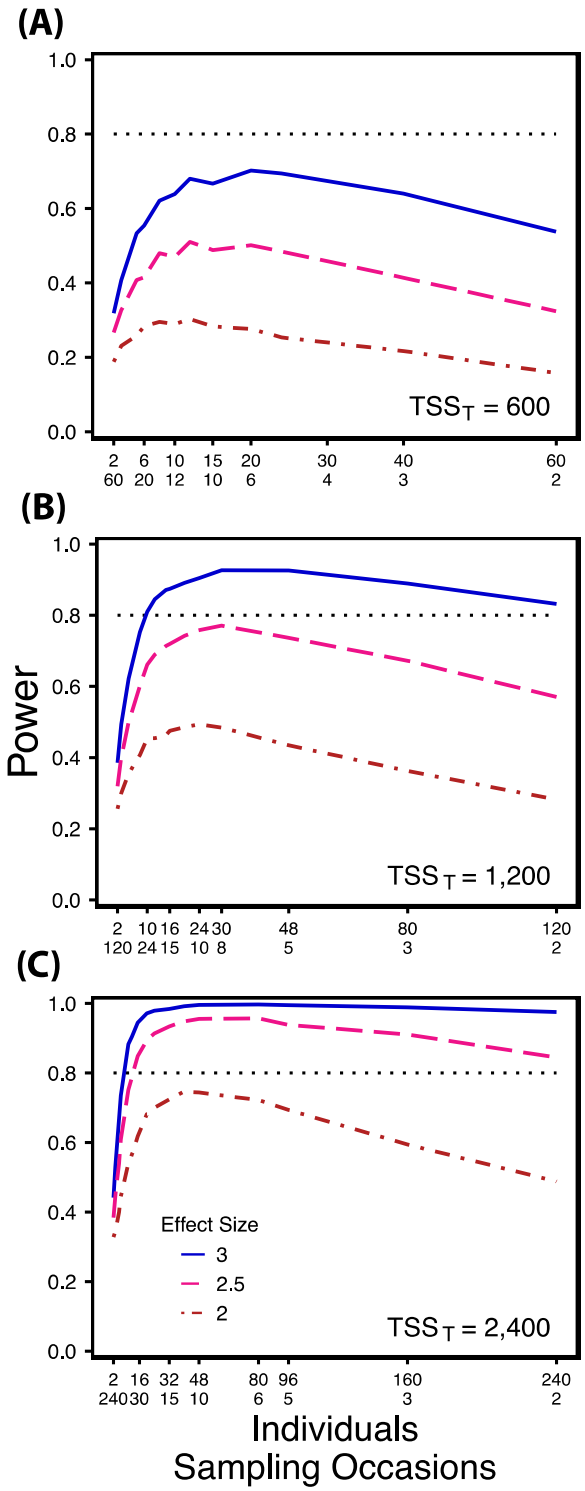

Figure 2: Power to detect differences by treatment in σ_0 for three effect sizes (ratio of σ_0 between treatments) and three TSS_T (total sample size per treatment) [(A) = 600, (b) = 1,200, (C) = 2,400]. Each scenario was simulated with 5 Bernoulli observations per sampling occasion.

sizes ($TSS_T = 300$), power to detect σ^2_{1k} is maximized at a ratio of approximately 12:5 (Figure 3A), while larger sample sizes ($TSS_T =$ 600, 1200) favor a ratio heavily weighted towards having more individuals (approximately 5:1) versus more repeated measures (Figure 3B, C). Power to detectσ^2_{1k} is higher overall and less sensitive to deviations from the optimum ratio than power to detect σ^2_{0k} (Figure 3).

Power to detect σ^2_{vk} follows a strikingly different pattern than σ^2_{0k} andσ^2_{1k}. Power to detect σ^2_{vk} is non-monotonic across ratios of individuals to the number of Bernoulli observations within sampling occasions (Figure 4), and is an increasing function of the number of sampling occasions (Figure 5C). At low sample sizes (e.g. $TSS_T = 600$) power to detect $\sigma^2_{vk}$is maximized by devoting nearly all of the available resources to repeated measures within individuals (Figure 4A); however, at larger sample sizes (e.g. $TSS_T =$ 2,400) power is maximized at a ratio of individuals to Bernoulli observations of approximately 1:2 (Figure 4).

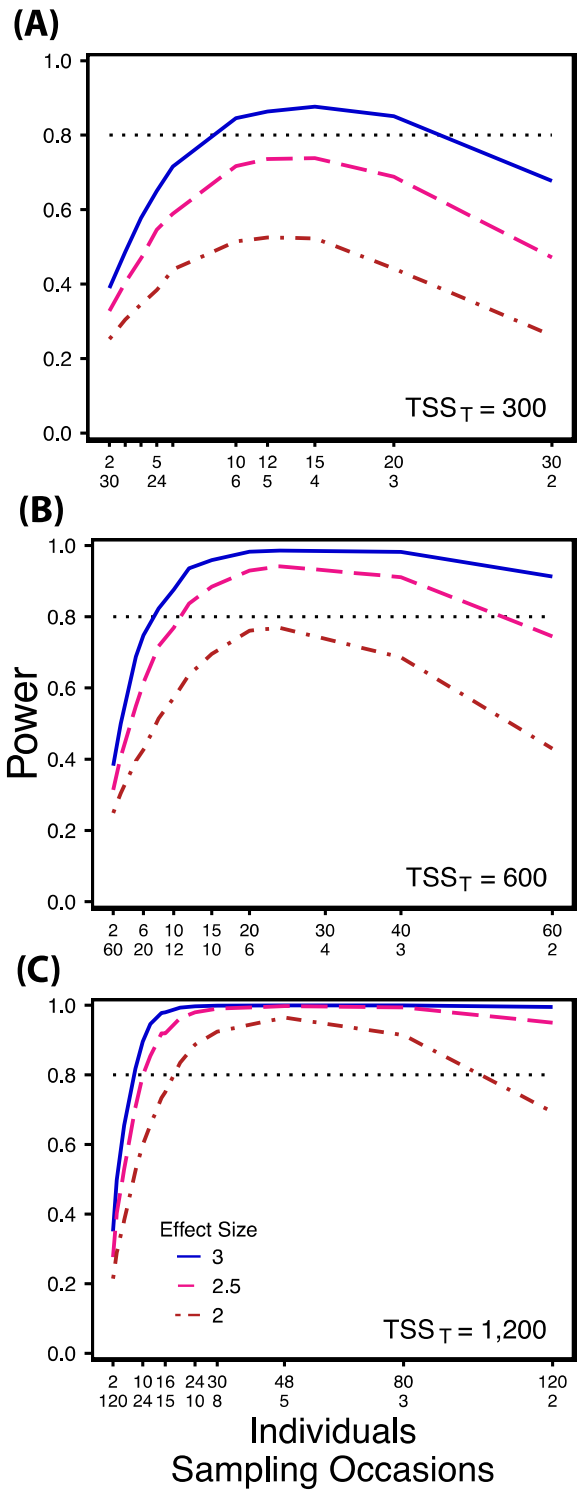

Figure 3: Power to detect differences by treatment in σ_1 for three effect sizes and three TSS_T [(A) = 300, (b) = 600, (C) = 1,200]. Each scenario was simulated with 5 Bernoulli observations per sampling occasion.

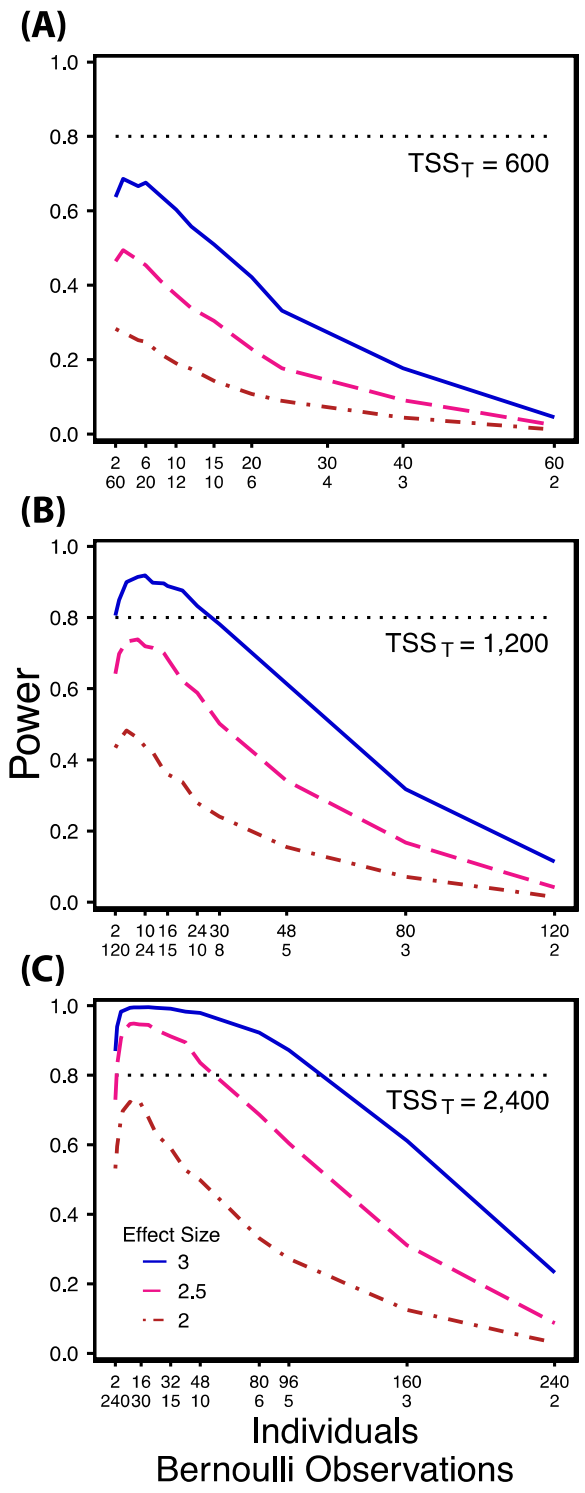

Figure 4: Power to detect differences by treatment in σ_v for three effect sizes and three TSS_T [(A) = 600, (b) = 1,200, (C) = 2,400]. Each scenario was simulated with 5 sampling occasions.

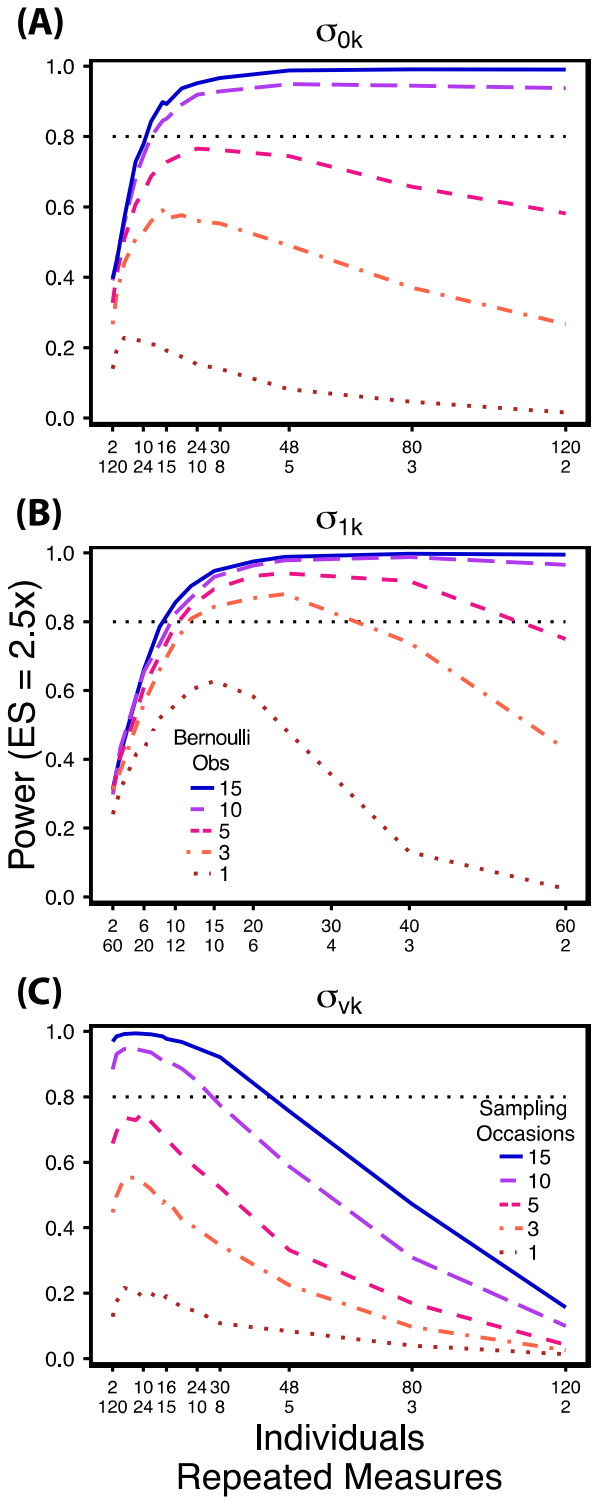

Figure 5: Power to detect differences by treatment in σ_0 (A) and σ_1 (B) under increasing Bernoulli observations per sampling occasion; σ_v (C) under increasing sampling occasions. In (A) and (B) ratios of individuals to sampling occasions follow figures 2B and 3B respectively. In (C) ratios of individuals to Bernoulli observations follows figure 4B.

Scenario 2: Power under increasing non-target random effect variance

Power to detect differences in variance components is strongly affected by the proportion of total variance that can be attributed to the target variance component (Figure 6). Increasing variance in non-target random effects decreases power to detect differences in the target variance parameter by treatment. However, the ratio of target to non-target variance does not alter the optimal ratio of individuals to repeated measures for the target variance comparison (Figure 6). Panel A demonstrates that power to detect σ^2_{0k} decreases substantially as the magnitude of within-individual variation increases. Detecting differences in σ^2_{1k} depends only on total random effect variation at extreme ratios of individuals to sampling occasions (e.g. 80:3) (Figure 6B). Finally, detection of σ^2_{vk} is largely independent of the magnitude of among-individual variation at large ratios of ES to non-target variance, as indicated by overlapping curves in Figure 6C. However, when among-individual

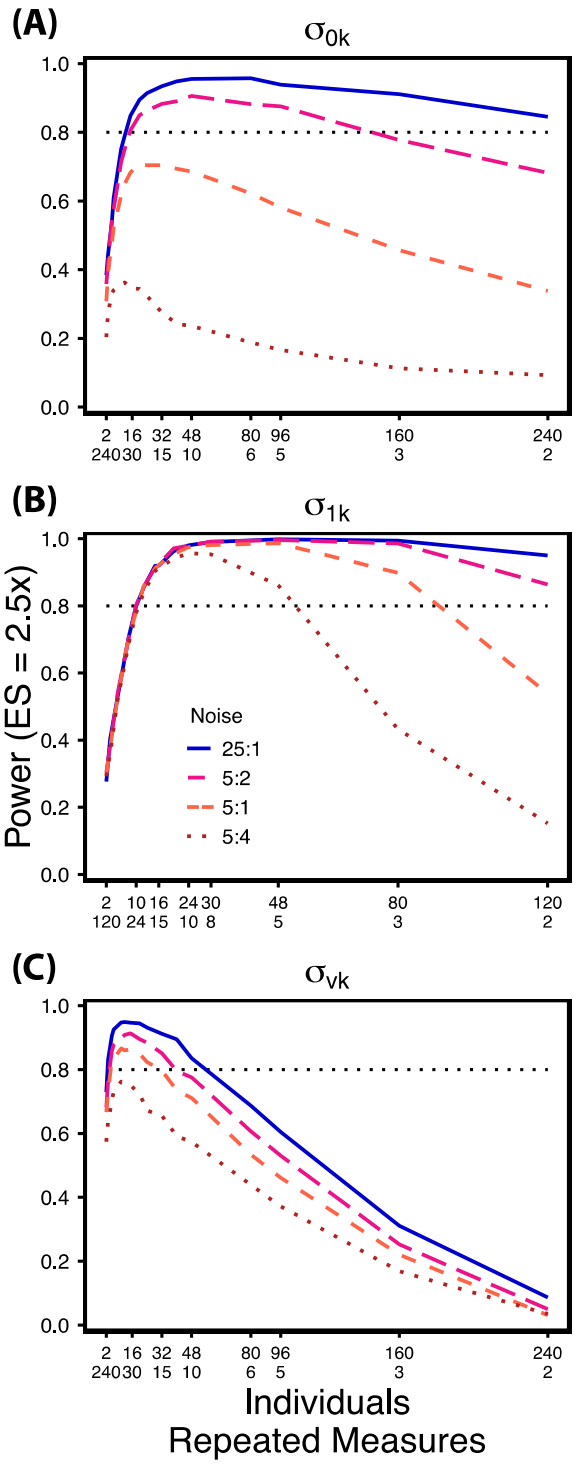

Figure 6: Power to detect differences by treatment in σ_0 (A) and σ_1 (B) under increasing variation in σ_v; σ_v (C) under increasing variation in σ_0. Noise is given as the ratio of effect size to variation in the non-target variance parameter. In (A) and (B) ratios of individuals to sampling occasions follow figures 2C and 3C respectively. In (C) ratios of individuals to Bernoulli observations follows figure 4C.

variation in intercept is very large (Figure 6C: Red curve), power to detect σ^2_{vk} decreases because individual mean responses approach 0 or 1, reducing the amount of detectable within-individual variation.

Discussion

The power analyses presented here establish a framework for designing experiments focused on detecting differences in variance components by treatment using GLMMs. These results should serve as a baseline upon which researchers can expand to address their own specific problems. Nevertheless, our findings reveal some important general trends that should be considered when designing experiments. Our results demonstrate heterogeneity in power across sampling schemes (ratio of individuals to repeated measures and partitioning of repeated measures into sampling occasions and Bernoulli observations), and differences in which sampling scheme maximizes power for different components of variance (Figures 2-5). As expected, power declines rapidly for low sample sizes and small effect sizes (Figures 2-4). However, for large TSS_T and relatively large effect sizes (3 SD difference between treatments), > 80% power is retained across many different combinations of individuals to repeated measures for each component of variance (Figures 2-5). Not surprisingly, power to detect differences in the target random effect declines with increasing variance in the non-target random effects (Figure 6).

Power to detect σ^2_{0k} is non-monotonic across ratios of individuals to sampling occasions, and is an increasing function of the number of Bernoulli observations per sampling occasion. Power is maximized with ratios weighted towards having more individuals (Figure 2), and quickly declines with alternative sampling ratios when total sample sizes and effect sizes are small. The analyses are however more robust to deviations from this ratio when TSS_T and ES are large (Figure 2C). Finally, of all the random effect parameters we analyzed, power to detect σ^2_{0k} is the most sensitive to the amount of "noise" present in the model, decreasing rapidly with increasing within-individual variation (Figure 6).

Power to detect σ^2_{1k} is also non-monotonic across ratios of individuals to sampling occasions, and is maximized with a ratio of individuals to sampling occasions ranging from 2:1 to 5:1 as TSS_T increases (Figure 3). On average, testing for differences in σ^2_{1k} are more powerful than for σ^2_{0k} across all sampling schemes and ES (Figures 2, 3), and requires fewer samples to obtain 80% power.

Finally, power to detect $\sigma^2_{vk}$is non-monotonic across ratios of individuals to Bernoulli observations and is an increasing function of the number of sampling occasions. Depending on sample size, sampling schemes ranging from maximizing Bernoulli observations to ratios of individuals to Bernoulli observations of 1:2 maximizes power (Figure 4). Unlike σ^2_{0k}, power to detect σ^2_{vk} is largely independent of additional variance in the model (Figure 6C), such that power to detect $\sigma^2_{vk}$is nearly equivalent at all levels of σ^2_{0k} except under the case of extreme values of $\sigma^2_{0k.}$

Collectively these results indicate the importance of clearly defining a biological question, designating the focal random effect, and knowing the expected magnitude of total variation when determining the appropriate experimental sampling design and TSS_T. Even at larger effect sizes, failure to account for system noise can lead to insufficient power and a failed experiment. Our findings should serve as a strong warning to empiricists interested in variance components that power analyses should be performed when designing experiments in order to overcome the problems of overall low power, large heterogeneity in power to detect different variance components, and heterogeneity in sampling scheme required to optimize power.

By introducing new strategies for analyzing variance among treatments we hope to inspire novel experimental designs in ecology and evolution. For example, the power analyses presented here can inform the design of experiments aimed at quantifying heterogeneous within-individual

variation by environment, which may lead to novel insights on the adaptive significance of within-individual variation (Westneat, Wright & Dingemanse, 2014).

In addition, these analyses answer the calls of researchers over the last decade for methods to investigate effects of treatment level variance on the variance of dependent variables (Benedetti-Cecchi, 2003). Transitions from one discrete environment to another (e.g. presence or absence of predators) are often classified as a form environmental variation, but switching between two distinct but relatively constant environments does not reflect environmental variation *per se*,such as temporal changes in the magnitude, pattern, and/or frequency of the environmental over time (Benedetti-Cecchi, 2003; Benedetti-Cecchi et al., 2006; Miner & Vonesh, 2004; Lawson et al., 2015). When this form of environmental variation is manipulated or natural variation exploited in an experimental context, within-individual variation can be described as the variable response of individuals to this variation in the environment. In this context, within-individual variation may itself be a form of phenotypic plasticity, and may have profound implications for understanding the evolution of environmentally induced plasticity, and the evolution of labile traits generally (Stamps, Briffa & Biro, 2012; Biro & Adriasenssens, 2013; Westneat, Wright & Dingemanse, 2014).

Further Considerations

Heterogeneous within-individual variation

In our power analyses we have made a few important simplifying assumptions. First, we assume that within-individual variation in both intercept and slope is homogenous among individuals within the same treatment. Additionally, we assume homogeneity of within-individual variance across an environmental gradient. However, these assumptions may not be true for some natural or experimental populations. In fact, it has recently been proposed that assessing the magnitude of variation in within-individual error variance within a single individual across an environmental gradient or among individuals exposed to the same environment/treatment is an important metric that may help to explain the evolution of plasticity (Cleasby, Nakagawa & Schielzeth, 2015; Westneat, Wright & Dingemanse, 2014). Power to detect differences in the magnitude of among-individual variation in within-individual variation by treatment (Cleasby, Nakagawa & Schielzeth, 2015) and heterogeneity of variance across an environmental gradient are interesting research questions that deserve attention, but are beyond the scope of this article. We also note that practicality limits exploration of increasingly complicated scenarios, despite their conceivable statistical feasibility and intrinsic charm due to complex novelty.

Covariance among intercept, slope, and variance components

All of our simulations assessed power to detect differences in a single target variance comparison between treatments, holding all other variance parameters constant (Table A1.1). However, manipulating non-target variation generates additional variation that is expected to decrease power to detect differences in the target variance parameter. Because we assumed no

slope variation in models where intercepts were allowed to vary and no intercept variation in the models focused on variation in slopes, we did not discuss power to detect covariance terms. However, these parameters can co-vary and the covariation among these parameters may contain a wealth of biologically relevant information. For example, covariation between phenotypic plasticity and within-individual variation may be tightly linked via developmental tradeoffs, which can lead to greater developmental instability in highly plastic individuals (Tonsor, Elnaccash & Scheiner, 2013). Indeed, it is not known whether an individual's reaction norm slope and within-individual variation around that reaction norm are always linked or if these relationships can be context-dependent. Similarly, we do not know if stronger behavioral responses lead to greater canalization of behavior. Understanding how to parameterize GLMM and how to optimize experiments to detect these covariances will be a useful step toward advancing evolutionary theory on adaptive, maladaptive and random patterns of variation.

Covariance between intercept and slope has been described extensively in theoretical papers and has been explored in earlier power analyses for LMM (Dingemanse & Dochtermann, 2013); however, empirical studies documenting significant covariance between these parameters remain rare (Mathot et al., 2011; Dingemanse et al., 2012). While covariance among these parameters may be uncommon, it is also likely that most experiments have insufficient power to detect such covariance. Additional analyses that determine power to detect significant differences in intercept and slope covariation for GLMMs is another important step considering the lack of current evidence for covariation reported in the literature.

Within-individual variation in slope

Research, including ours, on among-individual variation in plasticity assumes fully

repeatable plasticity within each individual, causing among-individual differences in phenotypic plasticity to be calculated using a single reaction norm for each individual (Dingemanse & Wolf, 2013). However, quantifying only a single reaction norm for each individual fails to capture any potential variation in plastic responses within an individual around its mean reaction norm, which may inflate estimates of among-individual variation and mask important variation that is subject to selection (Dingemanse & Wolf, 2013). Despite the reasonable assumption that each experimental individual would exhibit variation in their reaction norm if it were repeatedly measured, we are aware of no studies that demonstrate repeatable behavioral plasticity for a single individual when assessed multiple times.

Heterogeneity in sampling scheme and environment

In our simulations all individuals were measured an equal number of times and all treatments contained the same number of individuals, a luxury often not available to empiricists that often deal with missing data and unbalanced designs. Intuitively, unbalanced sampling schemes will lower the power to detect among-individual variation (Van de Pol, 2012); however we do not know the rate at which statistical power is lost with the magnitude of imbalance for a particular sampling design. Future research should follow the lead of Van de Pol, 2012 to determine how power to assess differences in variance for GLMM is affected by incomplete sampling, specifically when only a single measure is available for some individuals.

Experiments with more than two treatments

Finally, these power analyses were created for a two-treatment scenario--"homogenous" environmental variation treatment and a "variable" environmental variation treatment. However,

it is commonplace to have more than two treatments. Fortunately, our framework for conducting power analyses can be easily generalized for exploring power for experiments with more than two treatments (see supplemental material). In addition, syntax for the lme4package in R for specifying GLMM is highly flexible and can be written to restrict variance components to be the same in any number of treatments, while unique variance estimates can be obtained for any other given treatment. For example in a four treatment experiment composed of four levels of predator cue, two variance estimates could be obtained for among-individual variation (e.g. a single estimate for the three treatments with the lowest levels of predator cue and one estimate for the highest level of predator cue).

Conclusions

Random intercepts and slopes GLMMs are well established both ecology and evolution and behavioral ecology. Despite their ubiquity, the use of GLMMs to compare variance components among populations or among experimental treatments is rare. We call for future work analyzing the accuracy and precision of estimates comparing random effects by treatment for GLMMs (which our code facilitates) similar to the work of Moineddin, Matheson & Glazier, 2007 and Van de Pol, 2012 on the accuracy and precision of random effects estimates. As Van de Pol points out, just because power is high does not ensure the accuracy and precision of estimates. Finally, with expanding interest in a variety of variance parameters (e.g. heterogeneity in within-individual variation), we hope the power analyses presented here will spur novel empirical research and assist readers in constructing appropriate experimental designs and statistical models to test how variance components are shaped by ecological and evolutionary processes.

CHAPTER 2: THE EFFECTS OF ENVIRONMETNAL VARIATION ON INDIVIDUALVARIATION

Title: The effects of environmental variation on individual variation

Authors: Morgan P. Kain[1*], Michael W. McCoy[1]

[1]Department of Biology East Carolina University, N108 Howell Science Complex, Greenville, NC 27858

[*] Correspondence author. N108 Howell Science Complex East Carolina University Mail Stop 551 Greenville, NC 27858 E-mail: morganpkain@gmail.com

Abstract

Despite general consensus that both environmental and organismal variation impact ecological and evolutionary dynamics, few studies examine the interaction between these types of variation. As a result we know little about howenvironmental variation affectsorganismal variation at multiple levels and how variation among- and within-individuals in response to a variable environment influences organism fitness and population persistence. In this study weutilized experimental design that isolates changes in environmental variation from changes in the mean environment toevaluate the effects of environmental variation on among-individual variation in behavior, within-individual variation in behavior, and average behavioral responses. Specifically, we evaluated the effects of variation in predation risk by crayfish (*Procambarus sp*.) on among- and within-individual variationas well as average anti-predator behavior of freshwater snails (*Physa acuta*). In addition to our behavioral assays, we assessed the fitness consequences of behavioral and environmental variation by quantifying *Physa* survival and reproduction. We found that the effects of coarse-grain variation were primarily detectable in changes in the average behavioral responses of snails, whereas the effects of fine-grained variation were manifested in the magnitude of individual variation in behavioral responses. We also found that environmental variation affects survival and reproductive success in ways that are not fully explained by individual responses. In sum, our results suggests that variance estimates for different hierarchical levels (population, among, and within-individuals) may quantify alternative forms of environmentally induced plasticity that will have strong implications for predicting longer term evolutionary and ecological consequences of different forms and magnitudes of environmental variation.

Key-words: reaction norm, plasticity, individual variation, variance sensitivity, DHGLM, GLMM

Introduction

The key insights of Darwin and Wallace established variation as the central concept that makes evolutionary change possible and it has remained the core impetus for research in evolution and ecology every since. Yet, most studies have focused on estimating and comparing average responses of organisms in different environments, potentially masking important mechanisms that underlie ecological and evolutionary dynamics (Benedetti-Cecchi, 2003; Grimm & Railsback, 2005; DeAngelis & Mooij, 2005; Bolnick et al., 2011; Wolf &Weissig, 2012). Indeed, environmental variation (temporal variance in an environmental variable *sensu* Lawson et al., 2015) as well as variation among species, among populations and among individuals in their responses to environmental conditions have all been shown to affect population structure and the dynamics of ecological communities (e.g. Chesson, 1986; Butler, 1989; Miner & Vonesh, 2004; Benedetti-Cecchi et al., 2006; Bolnick et al., 2011; Wolf & Weissig, 2012, Lawson et al., 2015).

Despite the well-recognized ecological importance of both environmental and organismal variation, few studies examine the interaction between them (but see Mathot et al., 2012). As a result we know little about how environmental variation affects organismal variation at multiple levels. Yet, environmental variation is expected to produce variation both among and within individuals because individuals differ in physiological condition (Wolf & Weissig, 2010; Sih et al., 2015), cognition (Niemela et al., 2012;Stamps, Briffa & Biro, 2012), and in their ability to evaluate time intervals (Stamps, Briffa & Biro, 2012). Furthermore, while it is known that both environmental variation and individual variation independently affect organism survival (e.g. Wolf & Weissig, 2012;Stamps, Briffa & Biro, 2012), it is unknown how variation in traits (e.g. anti-predator behavior) among and within individuals in response to a variable environment

influences organism fitness and population persistence. Environmental variation may also indirectly impact population and community dynamics by increasing organismal variation, which in turn promotes the persistence of populations and stabilizes communities (Bolnick et al., 2011;Stamps, Briffa & Biro, 2012; Wolf & Weissing, 2012;Westneat, Wright & Dingemanse, 2014; Royaute & Pruitt, 2015).

Among-individual variation can expedites evolution in the presence of selection pressure (Wolf & Weissig, 2012), and intra-individual variation is required for the evolution of both phenotypic stability and flexibility (Westneat, Wright & Dingemanse, 2014). Thus, environmental variation may also influence evolutionary dynamics at multiple hierarchical levels. For example, a variable predator environment may result in individual prey that vary greatly around their own average behavioral response because they remain unpredictable to the predator, leading to greater magnitudes of within-individual variation across generations (Stamps, Briffa & Biro, 2012). Therefore, understanding how environmental variation affects variation in organismal responses at the population, individual and within-individual levels has important implications for understanding how variation affects ecological and evolutionary dynamics.

Here we utilize an experimental design that isolates changes in environmental variation from changes in the mean environment to evaluate the effects of environmental variation on average behavioral responses, among-individual variation in behavior, and within-individual variation in behavior. Specifically, we evaluate the effects of variation in predation risk by crayfish (*Procambarus sp*.) on among- and within-individual variationas well as average anti-predator behavior of freshwater snails (*Physa acuta*). We also assess the fitness consequences of behavioral and environmental variation by quantifying *Physa* survival and reproduction. We

manipulate variation in predation risk at both fine- and coarse-grained scales. Research on fine-grained variation focuses on quantifying responses to temporal variation in the magnitude of an environmental variable (e.g. Butler, 1989; Benedetti-Cecchi, 2003; Miner & Vonesh, 2004; Lawson et al., 2015), while coarse-grained temporal variation focuses on temporal shifts in an environment between a null state and an alternative state (e.g. Benedetti-Cecchi et al., 2006; Cifuentes et al., 2007; Bertocci et al., 2007).

In this study we manipulate both forms of variation because each can lead to unique organismal responses due to different mechanisms of action. Fine-grained environmental variation often results in differences between the mean response and responses in the mean environment due to nonlinear reaction norms, a phenomenon known as Jensen's inequality (Ruel & Ayres, 1999). For example, fine-grained variation in temperature can impact organism performance as well as long-term population growth rate even if the mean temperature is unchanged as a result of nonlinear thermal performance curves (e.g. Estay et al., 2011;Foray, Desouhant & Gibert, 2014; Paijmans et al., 2013; Vasseur et al., 2014). Conversely, under coarse-grained variation responses often cannot be predicted by a continuous function because shifts between discrete environmental states may trigger behaviorally or physiologically mediated shifts in the elevation of response curves (see Inouye, 2005; Benedetti-Cecchi, 2005). For example, an increase in the frequency of transitions between submergence and exposure to air in an intertidal community favored taxa that were able to retain moisture by growing in dense mats (Benedetti-Cecchi, 2006).

Pond snails (*Physa sp*.)are known todiffer in predator cue perception (DeWitt, Sih & Hucko, 1999), learning, memory and predator avoidance (Turner, Turner & Lappi, 2006), which makes them ideal models for investigating the effects of both fine- and coarse-grained environmental

variation on among- and within-individual variation. These experiments provide the first empirical evidence for environmental variation as an important driver of individual variation and one of the most comprehensive evaluations of the impacts of environmental variation in any system.

Methods

We conducted two experiments manipulating variation in predator cue environment. In our first experiment we manipulated temporal variation in predator risk by controlling the pattern of cue exposure across the full study duration (i.e. Benedetti-Cecchi et al., 2006; Figure 7).We call this the coarse-grained environmental variation experiment (*CGV*). To keep the full study duration mean cue exposure constant across all treatments each treatment received the same total days of cue. In the second experiment both the pattern of cue and cue concentration were manipulated in two-day segments (i.e. Benedetti-Cecchi, 2003; Miner & Vonesh, 2004). Here mean cue exposure in all treatments for each two-day segment was held constant (Figure 7). We refer to this as the fine-grained variation experiment (*FGV*).

Study System

Each experiment was modeled on a simplified pond food web composed of juvenile freshwater snails (*Physa acuta*)and crayfish (*Procambarus* sp.)—a common snail predator. *Procambarus* are known to induce "crawl-out" behavior (climbing to or above the waterline) and hiding behavior (Covich et al., 1994; Bernot & Turner, 2001) in *Physa*. These well-documented *Physa* behavioral responses to threat of crayfish predation, along with natural variation in predator presence and abundance in natural ecosystems (DeWitt, Sih & Hucko, 1999) make this system

ideal for testing questions about how temporal variability in predator presence impacts individual variation in prey behavior.

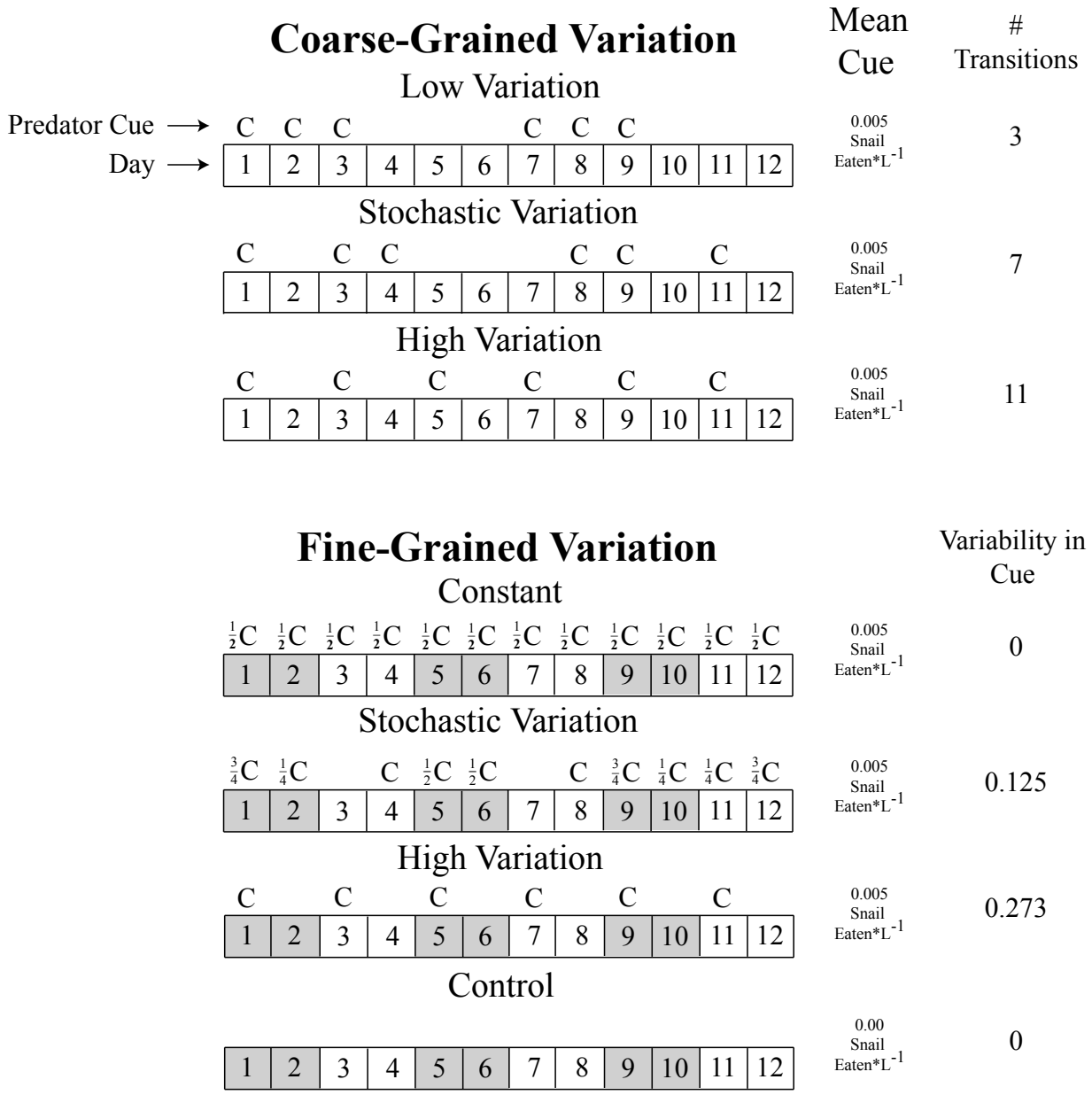

Figure 7: Treatment descriptions for both "Fine-Grained" and "Coarse-Grained" experiments. Each **C** corresponds to a *Procambarus* cue concentration of 0.01 snails consumed*L^{-1} of water. The absence of a **C** indicates no cue exposure. In all treatments in both experiments (apart from the Control treatment in the "Fine-Grained" experiment) *Physa* were exposed to a daily cue concentration of 0.005 snails*L^{-1}. In the "Coarse-Grained" experiment variation was determined by the number of environmental transitions. In the "Fine-Grained" experiment each treatment received a total of 0.01 snails consumed*L^{-1} in each two day period, which is denoted by the grey and white boxes. Here variation was determined by the change in cue magnitude between days. Adapted from Benedetti-Cecchi (2003).

All study organisms were obtained from ponds near Greenville, NC, USA. All experimental snails in *CGV* were laboratory reared offspring of wild caught *Physa*. For both experiments *Physa* were 7.0 ± 0.5mm long and approximately the same age at the start of the experiment. All *Procambarus* differed by less than 40mm in length but were of unknown age.

Experimental Design

Setup

Twelve 10-gallon glass aquaria were prepared with a 0.5 cm sand substrate and a 15x15 cm slate tile positioned 6cm from the long end of one side of the tank. A single AquaClear 20 filter was positioned above the tile shelter and set on ¾ power to cycle water without disturbing the sand bottom. Each tank was filled with tap water to a height of 20.2 cm above the sand bottom and treated with API water conditioner. Tanks were held in an aquarium room with controlled temperature (23^{o}C) and light-dark cycle (12h: 12h).

A total of sixty snails were randomly marked one of five colors using OPI© nail polish and allocated to the twelve tanks so that each tank received five snails of different colors. Snail length was measured using digital calipers at the time of initial painting. Following painting and initial measurements, snails were allowed to habituate to their new environment for 48 hours before experimentation began. Snails were repainted and measured on day 7 and measured again immediately following the behavioral trials on day 12. Snails were fed frozen spinach *ad libitum* throughout the experiment.

Prior to experimentation, three days of initial dose response trials were conducted to determine the dose response curve of snails to crayfish cue. A cue concentration window of 0.001—0.01 snails consumed*L^{-1} was determined to be optimal as it produced a large mean anti-

predator response (> 50% of snails at the water line after 20 min) and large variation in anti-predator response among snails (e.g. variability in time to return to tank bottom following addition of crayfish cue).

For both experiments predator cue was prepared by placing three randomly selected crayfish chosen from the laboratory stock of twelve crayfish in small plastic containers containing 0.5L of water and feeding each three 8mm *Physa*. The cue water from the three containers was combined to form 1.5L of concentrated stock cue water. Stock cue water was combined with water removed from crayfish holding tanks to produce 1L of cue water for each tank at the required cue concentration (0.0025 snails*L^{-1}–0.01 snails*L^{-1}; Figure 7). One liter of conditioned tap water was added to tanks when no cue was needed. One liter of either cue or tap water was used to bring the water line in experimental tanks from 20.2cm to 21.0cm.

Coarse-grained experiment unique design elements

For this experiment a total of three treatments (low variation, high variation, and stochastic) were used (Figure 7). Each treatment received a total of six days of cue at a concentration of 0.01 snails consumed*L^{-1} and six days of no cue arranged in different patterns. The low variation treatment received 3 days of a similar environment (cue or non-cue) before switching to a new environment, while the high variation environment alternated between cue days and no-cue days. The stochastic treatment was a randomized treatment with two restrictions: day 1 was restrained to be a cue day to correspond with the other two treatments, and the total number of cue days was restricted to six. Each treatment used a total of replicate 4 tanks, each containing 5 snails for a total of 20 snails per treatment.

Fine-grained experiment unique design elements

The fine-grained variation experiment reduced the total experimental duration into multiple 2-day time windows to isolate the effects of fine-grained temporal variation in predation risk. Variation was decoupled from changes in means by manipulating the timing and concentration of predator cue to hold total predator cue to 0.01 snails*L^{-1} in each 2-day time window. A total of four treatments (constant, high, stochastic, and control) were used (Figure 7). The constant treatment received 0.005 snails*L^{-1} of cue each day, while the high variation treatment received 0.01 snails*L^{-1} on odd-numbered days and no cue on even-numbered days. In the stochastic treatment the pattern of cue in each 2-day period was randomized but restricted to a total of 0.01 snails*L^{-1}. A control was also included for measures of *Physa* behavior in the absence of cue. Each treatment utilized a total of 3 replicate tanks, each containing 5 snails for a total of 15 snails per treatment.

Behavioral Trials

All behavioral trials for each experiment were conducted between 9:00am and 3:00pm for twelve consecutive days. For *CGV*, behavioral trials lasted 4h 40min, while the behavioral trials for *FGV* were extended to 6h 20min. Snail location (on sand bottom, hiding underneath the tile shelter or height on tank walls in cm) and behavior (stationary, moving, foraging, hiding) were recorded three times (20 minutes apart for the coarse-grain experiment and 40 minutes apart for the fine-grain experiment) prior to predator cue addition, and again a total of 7 times over 4h for *CGV* and 9 times over 5h for *FGV*. Carbon filters were removed from the aquarium filters prior to the beginning of behavioral trials on each day and replaced at the end of the day's behavioral trials. Two carbon filters were used (for at least 18 hours) to ensure that cue was removed from

the water between days. During the dose response trials, we confirmed that cue water filtered through two carbon filters for 18 hours resulted in *Physa* anti-predator behavior that did not differ from the control.

Predation Trials

Following the twelve days of non-lethal predator behavioral trials a single crayfish was introduced to each experimental tank. For *CGV* crayfish were added at 9:00am on day fourteen, 42 hours after the end of the final behavioral trial, and for *FGV* at 9:00am on day thirteen, 18 hours after the end of the final behavioral trial. Crayfish of similar size were selected at random and placed into each experimental tank. Following the addition of live crayfish, proportion and identity of surviving snails was recorded over 16 observations spanning 48 hours. After 48 hours all remaining snails and crayfish were removed from experimental tanks and the surviving snails were recorded.

Finally, to quantify snail reproductive success and oviposition behavior, we counted the number of egg masses in each tank and recorded their position (tank walls or underneath tile shelter) at the end of the lethal predator trials.

Statistical Analyses

Behavioral Trials

We analyzed the effects of variance in predator risk environment on two categories of snail anti-predator behavior: 1) climbing patterns and 2) behaviors independent of climbing patterns.For both experiments climbing patterns were obtained by fitting logistic functions to the climbing behavior of each snail on each day (snail:day) they climbed (Figure 8). The specific

form of the logistic function and procedure used to fit these functions is described in Supplement 1.From each fitted curve we extracted time at first response (the time the snail initiated a climbing response), time at max height (the time the snail reached their maximum height), max height (the maximum height the snail reached), and duration of response (the total time elapsed during the climbing event) (Figure 8). Behaviors independent of climbing patterns included the number of observations spent hiding, foraging, and climbing and were recorded for all individuals on all days.

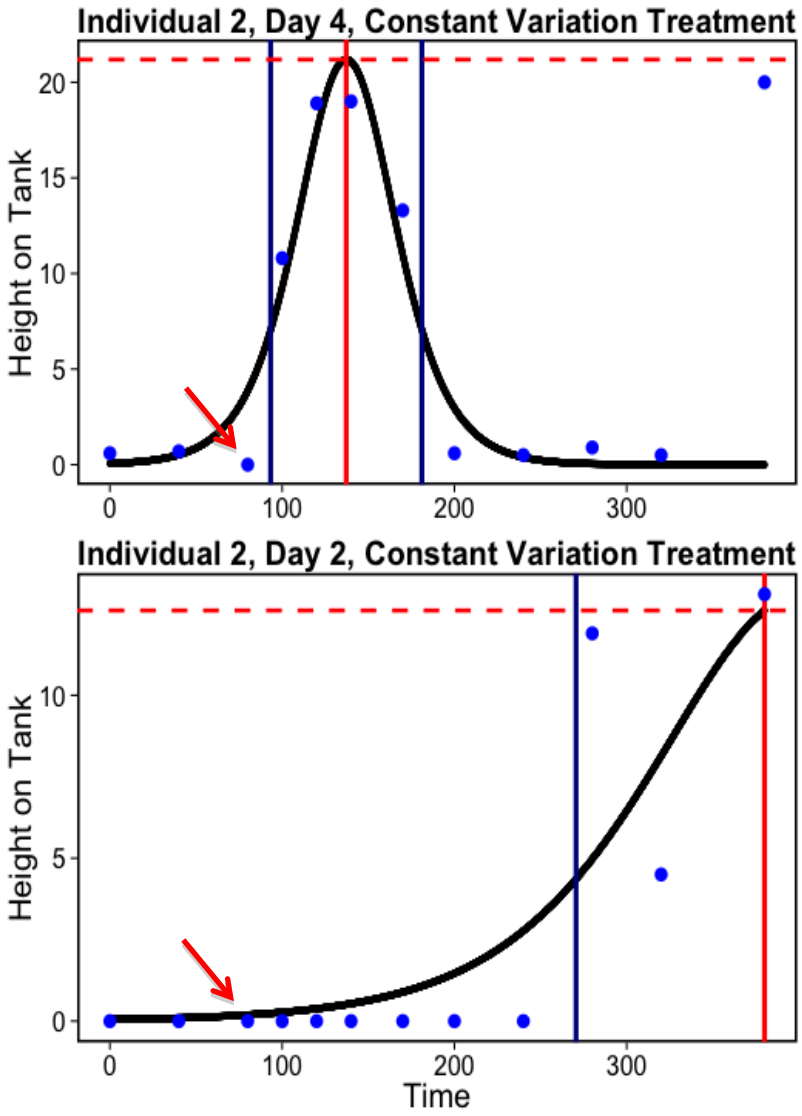

Figure 8: Example of two patterns of snail climbing behavior. Arrows indicate the point of cue addition. Vertical bold red lines indicate the time at max height, horizontal dashed red lines indicate max height, vertical bold blue lines indicate time at first response and time at last response. (A) illustrates a characteristic anti-predator climbing response occurring directly following cue addition. (B) Is an example of a snail:day that was excluded because the snail max height occurred on the last measurement.

Analysis of all behaviors was conducted using the Bayesian statistical software JAGS (Plummer, 2003) interfaced with the R statistical programming environment (R development core team, 2015) using the packages rjags (Plummer, 2014) and R2jags (Su, 2014). First, treatment effects on individual variation in behavior were analyzed using double hierarchical generalized linear models (DHGLM) (Lee & Nelder, 2006;Cleasby, Nakagawa & Schielzeth, 2015). DHGLM are hierarchical linear models that allow for explicit modeling of residual variance as a function of both fixed and random effects. Here we use DHGLMs to model additional structure in the dispersion (error) part of the model caused by treatment differences and individual variation. For a description of the DHGLM including a visual that illustrates each parameter of interest see Figure 9. For additional information on DHGLMs see Lee and Nelder (2006) and Cleasby, Nakagawa & Schielzeth(2015).

For all behaviors we fit four candidate models that assume a different variance structure for: 1) individual random effect (σ^2_{0k}); 2) treatment mean dispersion (β_{dk}) (Figure 9, Table 2). Specifically, these four candidate models are reduced DHGLMs with an experiment level individual variation in error variance (φ^2_d; Figure 9)estimated from all individuals (we include R code, sample output, and equations for full DHGLMs that include treatment unique φ^2_d in Supplement 1). All fixed effects (treatment, cue magnitude, and day) and a random effect for tank were included in each candidate model. For each behavior, the candidate model with the smallest DIC (deviance information criterion) was chosen as the most parsimonious model that best captured the structure of among- and within-individual variation.

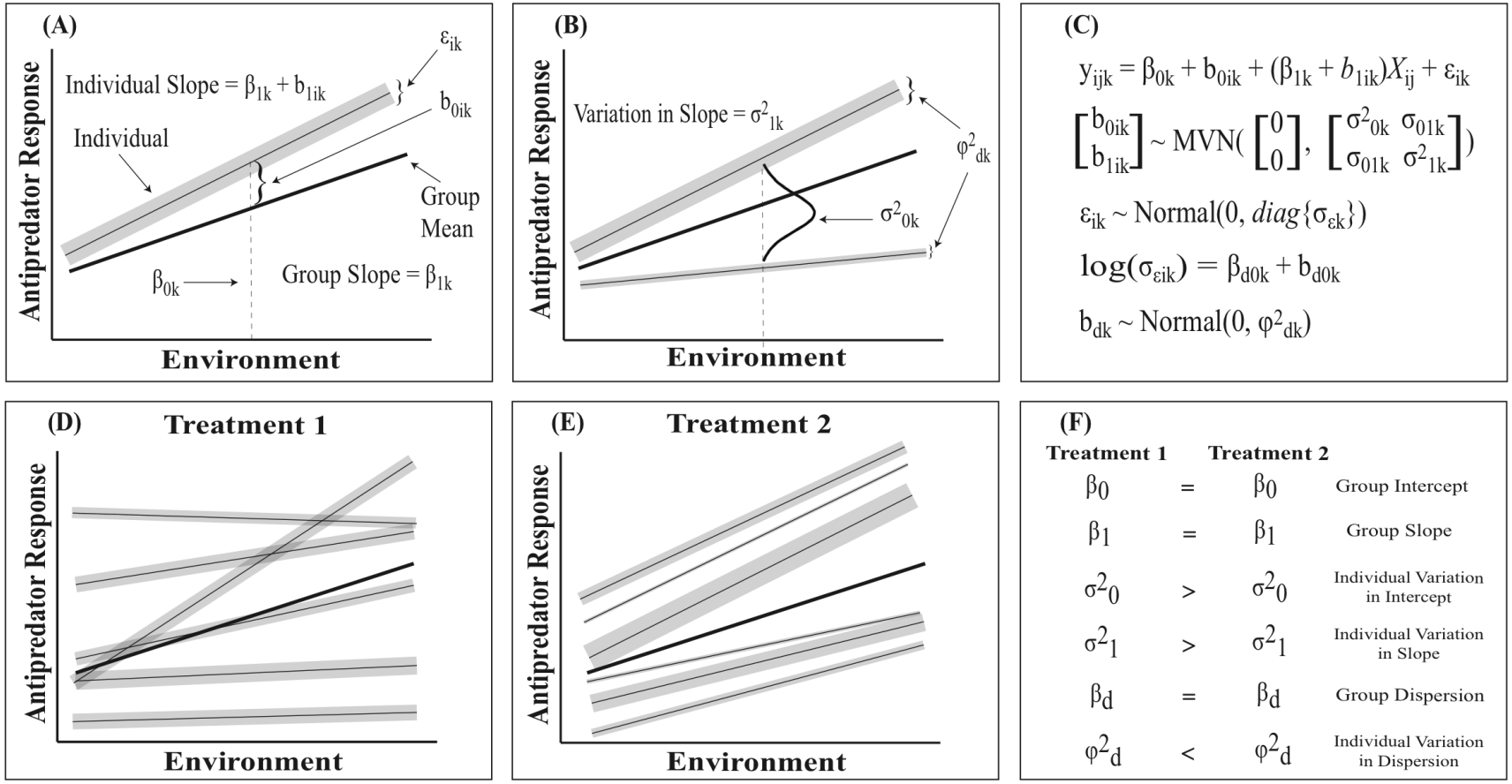

Figure 9: Illustration of the variance parameters of interest. Panels A depicts the deviation of a single individual from the group mean and illustrating each component of the linear predictor described in panel C. Panel B illustrates the random effects of: σ^2_{0k}: Individual variation in intercept; σ^2_{1k}: Individual variation in slope; φ^2_{dk}: Individual variation in error variance. Panel C contains a slightly truncated DHGLM model (for a more complete discussion of DHGLMs see Cleasby et al. 2015). Panels D and E provide an illustration of two experiments that differ in the magnitude of each random effect but have equivalent group mean intercept (β_0), slope (β_1) and dispersion (β_d). Panel F describes the relationship between treatments for each parameter. In this paper we test for differences in β_0, β_1, β_d, σ^2_{0k}, and $\varphi^2_{dk,}$ but include σ^2_{1k} for completeness.

To determine the fixed effects that contributed to predicting each snail behavior a second round of model selection was conducted. In this step we fit eight new candidate models for each behavior. Each model used the best variance structure determined from step one and one of the eight possible combinations of fixed effects (Table 2). For each behavior, the model with the smallest DIC was chosen as the overall most parsimonious model for describing both the variance structure and fixed effects. This procedure occasionally resulted in selecting fixed effects models with a ΔDIC that common rules of thumb would describe as being equivalent or indistinguishable from the next best model (Speigelhalter et al., 2002) (Table A2.4). Therefore the fixed effects retained in the final models (Table 3) should be cautiously interpreted in concert with an examination of the ΔDICs available in Table A2.4.

Table 2: Candidate models fit to determine the most parsimonious model for both the variance structure and fixed effects. A total of four models with different variance structures and eight models with all combinations of fixed effects were fit. An X indicates a variance structure or fixed effect included in the model. Here we highlight model 2.3 as the best overall model, highlighted in gray.

Error Variance	Heterogeneous Random Effect	Homogenous Random Effect	Heterogeneous Error Variance	Homogenous Error Variance	Homogenous Error Variance Random Effect
1	X		X		X
2	X			X	X
3		X	X		X
4		X		X	X

Error Variance	Fixed Effects	Day	Cue	Treatment
2	2.1	X	X	X
2	2.2	X	X	
2	2.3	X		X
2	2.4		X	X
2	2.5	X		
2	2.6		X	
2	2.7			X
2	2.8			

We adopted this two stage approach for two reasons: 1) A one step approach fitting models with all combinations of both variance structure and fixed effects simultaneously would have resulted in 768 models, which was not computationally feasible. In contrasts, this two-stage approach reduced the total to 288 models. 2) Given the lack of scientific consensus on the most appropriate metric for Bayesian model selection or model averaging (e.g. Gelman & Shalizi, 2013; Barker et al., 2015) we used an approach that allowed us to simply determine the model that best captured both the variance structure and the contribution of fixed effects.

Analyses of data from both experiments were conducted using this procedure with one exception because one of the treatments in the *FGV* was a control that never received predator cue (Figure 7). Thus, we analyzed all response variables for *FGV* with the control absent (FGV_{-C}) to isolate the effects of variation in predator cue on patterns of individual behavioral variation and also with the control present (FGV_{+C}) to determine if inferences changed when treatments also differed in mean cue.

All count data were analysed with DHGLMs assuming Poisson marginal distributions with Normally distributed random effects. To quantify treatment specific within-individual variance we used an additive overdispersion model which models increased Normally distributed variance in the linear predictor on the link scale (Browne et al., 2005; Nakagawa & Schielzeth, 2010). Data for all climbing behaviors were fit with DHGLMs assuming a Normal marginal distribution and Normally distributed random effects. Prior to analysis climbing behavior response variables were transformed with a box-cox data transformation to fit the assumptions of Normality.

Snail:days in which snails' max height occurred at the first or last measurement were removed prior to analysis (approximately 15-20% of all snail:days) (Figure 7). All excluded

snail:days were analyzed with a binomial GLMM in `lme4` (Bates et al., 2014) to determine if the probability of data exclusion differed by treatment, cue concentration or across days. Probability of data exclusion did not differ by treatment, thus inference on variance structure is assumed to be unaffected (see Supplement 1 for full results and implications of data censoring). Data exclusion was conducted for two reasons: 1) Each snail's movement on censored snail:days was too noisy to conform to a non-monotonic logistic function, resulting in a max height and time at max height that were not biologically informative (attributable to cue addition). Furthermore, on these snail:days duration was indeterminate because there was no calculable beginning and end to the climbing response; 2) A high concentration of data points at the limits of the data range resulted in model residuals that followed no known distribution. Due to the abundance of 0's any transformation resulted in substantial bias.

Lethal Predator Trials

We analyzed the effects of lethal predators on the probability of snail survival using a binomial GLMM fit in `lme4` (Bates et al., 2014) in the R statistical programming environment (R development core team, 2015). Treatment, time of predator exposure, snail size, and individual level behavioral responses during the twelve days of behavioral trials were used as predictors of the probability of mortality. Individual level predictors were defined *a priori* and included both mean behavior and variance in behavior for duration of response, proportion of time spent hiding, proportion of time spent foraging and total vertical movement during the twelve days of non-consumptive predator trials.

Snail Oviposition Behavior

To evaluate differences in egg laying behavior by treatment we analyzed the total number of egg masses laid and the number of egg masses laid underneath the tile shelter using Poisson GLMMs.

Results

To improve interpretation we organized our results by grouping the effects of environmental variation on mean behavior separately from its effects on variance in behavior. For both mean behavior and variance in behavior we describe the results from *CGV* prior to results from *FGV*. We mirror this layout in the discussion.

Our primary goals were to determine: a) whether the magnitude of among- and within-individual variation in *Physa* behavior differed by treatment; and b) if average *Physa* behavior differed by treatment, cue magnitude and day. As described in our methods and in supplement 1, we used a model selection approach using DIC as a selection criterion. Inferences about the effects of the magnitude of coarse- and fine-grained environmental variation on mean behavior and variance in behavior is based on the proportion of behavioral endpoints that differed by treatment in mean response, among-individual variation, and within-individual variation. We present effect sizes for both fixed effects and variance parameters for select cases that best highlight overall effects observed for each experiment. However, effect sizes for all variance parameters and fixed effects are available in Appendix 2:Table 4 and ΔDICs for all models are reported in Appendix 2:Table 3.

Behavioral Trials

Mean Behavior: Coarse-Grained Variation

Crayfish cue concentration was the most important fixed effect in *CGV*, retained in models for all eight response variables. Variation magnitude (treatment) was the second most important fixed effect, retained in models for five response variables. In *CGV* the high variation

treatment had the largest mean (from 7-62% greater than the next largest treatment mean, Table A2.5) for all five response variables for which treatment unique means were supported.

Mean Behavior: Fine-Grained Variation

Crayfish cue concentration and treatment were also the first and second most important fixed effects in *FGV,* respectively(Cue: FGV_{-C}: retained in four out of eight models; FGV_{+C}: 5/8, Treatment: FGV_{-C}: 4/8; FGV_{+C}: 3/8). Interestingly, in both FGV_{-C} and FGV_{+C} cue magnitude was not retained for time at max height or time at first response (Tables 3, A2.3). Whiletreatment unique means were supported in FGV_{-C} for four response variables, there was no apparent relationship between the magnitude of environmental variation and the degree of anti-predator responses. Hiding observations were highest in the stochastic variation treatment (60% greater than the next largest treatment mean, Table A2.5), climbing observations were highest in the constant variation treatment (16% greater), and max height was largest in the high variation treatment (19% greater). Interestingly, with the control included in the analyses for *FGV*, treatment unique means were retained in models for only three response variables (Tables 3, A2.3).

Variance Structure: Coarse-Grained Variation

In *CGV* the magnitude of individual variation around the treatment mean, determined by the standard deviation of the individual random effect (σ^2_{0k}), differed among treatments for five behaviors (Tables 3, A2.3). In *CGV* there was no clear pattern between the magnitude of environmental variation and the size of the individual random effect (σ^2_0) (Table A2.5). For

example, the high variation treatment had the largest σ^2_0 for three behaviors, but the smallest σ^2_0 for two others (Table A2.5). Treatment unique mean dispersion (β_{dk}) was included in models for only hiding observations.

Variance Structure: Fine-Grained Variation

In FGV_{-C} the magnitude of individual variation around the treatment mean (σ^2_{0k}), differed among treatments for four behaviors (Tables 3, A2.3). When the control was included in the analyses σ^2_0 differed by treatment for all eight behaviors. With the control absent the high variation treatment had the smallest σ^2_0 estimate for all behaviors, and when the control was included in analyses the control had the smallest σ^2_0 and the high variation treatment had the second smallest σ^2_0 estimate for all behaviors (Table A2.5).

In FGV_{-C} treatment unique mean dispersion (β_{dk}) was retained in models for hiding observations and response duration, and in FGV_{+C} β_{dk} was retained in models for response duration and total vertical movement (Tables 3, A2.3).

Table 3: The most parsimonious model for each response variable in both experiments. Table 1 provides a description of each model. ΔDIC values are the difference between the best and worst model. For all ΔDIC values for all models see Table S3.

Experiment 1 (Coarse-Grained Variation)				**Experiment 2 (Fine-Grained Variation)**					
				Control Absent			**Control Present**		
Response Variable	Model	ΔDIC		Model	ΔDIC		Model	ΔDIC	
		Variance	Mean		Variance	Mean		Variance	Mean
Hiding Observations	1.4	-15.56	-15.08	1.4	-8.85	-8.13	2.4	-4.10	-22.59
Foraging Observations	4.6	-14.79	-43.58	4.3	-8.50	-40.43	2.2	-22.30	-31.35
Climbing Observations	4.1	-35.37	-40.21	2.4	-15.06	-14.34	2.4	-15.18	-19.99
Max Height	2.4	-21.23	-41.81	4.1	-11.03	-3.61	2.2	-14.48	-5.47
Time at Max Height	2.6	-5.26	-17.80	2.8	-17.00	-6.01	2.8	-24.52	-6.78
Time at First Response	4.6	-4.09	-8.85	2.8	-16.44	-4.01	2.8	-19.61	-5.35
Response Duration	2.1	-8.91	-24.96	3.8	-3.91	-5.71	1.3	-4.12	-8.35
Total Vertical Movement	2.4	-18.03	-76.84	4.2	-6.45	-19.72	1.2	-25.41	-26.32

Lethal Predator Trials

The rate at which snails were eaten when exposed to lethal predators differed by treatment with marginal significance in *CGV* (Table 4; Binomial GLMM, LRT, $p = 0.071$) and significantly for *FGV* when analyzed with both the control removed (Table 4; Binomial GLMM, LRT, $p = 0.019$) and included (Table 4; Binomial GLMM, LRT, $p = 0.004$). However, neither snail size nor mean or variance in any of the *a priori* defined snail behaviors significantly predicted time to mortality (Table 4; Binomial GLMM, LRT, all $p > 0.05$).

Table 4: Snail survival during the lethal predator trials. Numbers listed are p-values from likelihood-ratio tests. Bolded values are p-values found to be significant at $\alpha = 0.05$. No p-values are listed for the predictors time or treatment for experiment 2 because of the significant treatment*time interaction term.

Experiment 1 (Coarse-Grained Variation)		**Experiment 2 (Fine-Grained Variation)**	
Predictor		Control Absent	Control Present
Treatment*Time	0.071	**0.019**	**3.6×10^{-3}**
Time	**2.2×10^{-16}**	-	-
Treatment	**0.016**	-	-
Mean Behavior, Variation in Behavior, body length	All $p > 0.05$	All $p > 0.05$	All $p > 0.05$

Oviposition Behavior

Oviposition behavior was unable to be traced to specific snails due to grouping of snails in tanks. However, analysis of treatment level effects on aggregate snail oviposition behavior revealed interesting patterns in *FGV*. In *FGV* mean number of egg masses laid differed by treatment, with snails in the high variation treatment laying the fewest eggs (Table 5; Poisson GLM, LRT, FGV_{-C}: Constant: 1.0x, Stochastic: 0.88x, High: 0.68x, $p = 2.79 \times 10^{-4}$; FGV_{+C}: Control: 1.0x, Constant: 0.83x, Stochastic: 0.73x, High: 0.58x $p = 2.41 \times 10^{-5}$). Additionally, after controlling for the total number of egg masses laid, the number of eggs laid underneath the tile shelter differed by treatment, with snails in the high variation treatment laying the lowest proportion underneath shelter (Table 5; Poisson GLM, LRT, FGV_{-C}: Stochastic: 1.0x, Constant: 0.78x, High: 0.28x, $p = 0.002$; FGV_{+C}: Stochastic: 1.0x, Constant: 0.78x, Control: 0.43x, High: 0.28x, $p = 0.057$). These patterns are especially intriguing because snail size did not differ by treatment in *FGV* (Constant: 9.97mm; High: 10.03mm; Stochastic: 9.66mm; Control: 9.92mm; LMM, LRT, $p = 0.196$).

Table 5: Snail oviposition behavior in all treatments for both experiments. Mean egg counts are presented in the top panels of the table (Coarse-Grain Variation Sample Size = 4; Fine-Grain Variation SS = 3). P-values from likelihood ratio tests between models with and without the predictor of Treatment are presented in the bottom panels.

Experiment 1 (Coarse-Grained Variation)			**Experiment 2 (Fine-Grained Variation)**				
Mean Egg Masses			**Mean Egg Masses**				
Treatment	**Total**	**Under Tile**	**Treatment**	**Total**		**Under Tile**	
Low	21.00	-	Constant	35.667		13.00	
High	19.25	-	High	25.00		3.33	
Stochastic	20.75	-	Stochastic	31.33		14.67	
			Control	43.00		8.67	
Treatment Significance				**Control Absent**	**Control Present**	**Control Absent**	**Control Present**
P-value	0.837	-		**2.79×10^{-4}**	**2.41×10^{-5}**	0.057	**0.0016**

Discussion

We found that both the type and magnitude of environmental variation cause changes in mean behavior and in variation in individual behavior. Specifically, we found that the effects of coarse-grain variation were primarily detectable in changes in the average behavioral responses of snails, whereas the effects of fine-grained variation were manifested in the magnitude of individual variation in behavioral responses (Tables 3, A2.4, A2.5). Changes in the average response versus changes in how variable individuals behave from both one another and from their own mean can have important implications for the evolution of behavioral reaction norms, and for the maintenance of variation in populations. We also found that environmental variation affects survival and reproductive success in ways that are not fully explained by individual responses, suggesting thatenvironmental variation is affecting survival on a level beneath detectable differences in behavior. In sum, our results suggests that variance estimates for different hierarchical levels (population, among, and within-individuals) may quantify an alternative form of environmentally induced plasticity that will have strong implications for the predicting longer-term evolutionary and ecological consequences of different forms and magnitudes of environmental variation.

Mean Behavior: Coarse-Grained Variation

In *CGV* environmental variation was adjusted by manipulating the frequency of environmental change (Figure 7). A high frequency of environmental change (high variation treatment) caused *Physa* to respond with the greatest average anti-predator response for each behavior whose mean differed by treatment (See Table A2.4 for effect sizes). This result is attributable to larger anti-predator responses of *Physa* on non-cue days that followed cue days

(Figure 10A). Because the high variation treatment received cue every other day, *Physa* in the high variation treatment exhibited the largest average anti-predator behavior on non-cue days relative to the other treatments (e.g. max height: LMM, LRT: treatment + cue vs. cue, treatment estimates: high: 10.722, low: 9.324, stochastic: 8.129, p = 0.036). An overall larger effect of cue in the high variation treatment also contributed to this result. For example, the difference between max height on cue days and non-cue days was largest in the high variation treatment (LMM, LRT: treatment*cue vs. treatment + cue, estimates: high: 3.34, low: 1.85, stochastic: 0.33, p = 0.017).

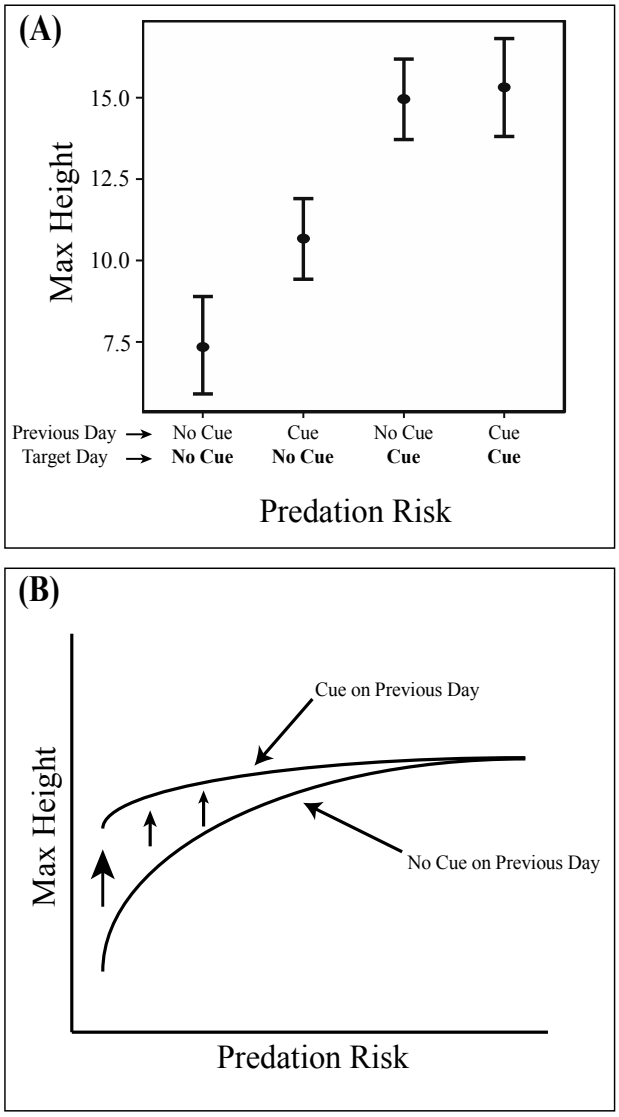

Figure 10: Illustration of the effects of coarse-grained variation on mean behavior. (A) depicts experimental data of average *Physa* max height on four types of target days. *Physa* mean behavior is greater on non cue days that follow cue days than on non-cue days that follow non-cue days. Error bars are 95% confidence intervals. (B) is a hypothetical illustration of two *Physa* response curves that differ in shape due to cue history. Previous exposure to cue increases max height the most at lower levels of predation risk.

These results suggest that individuals respond to the frequency of environmental change (e.g. Benedetti-Cecchi et al., 2006) through behaviorally mediated shifts in the elevation of individual anti-predator response curves (Figure 10B). However, in *CGV* we utilized only a single cue magnitude and thus were only able to capture shifts in the intercept of the response

curve. An important line of future research will be to design an experiment using methodology from both of our experiments, possibly using a response-surface design (Inouye 2001), to determine the effects of both shifts in the position of the response curve (*CGV*) and movement along the non-linear curve (*FGV*). Alternatively, lingering cue water from the previous day could have also led to these results, however this explanation is unlikely. Indeed, we conducted a dose response experiments conducted prior to the focal experiment that indicated *Physa* exposed to carbon-filtered cue water (0.01 snails*L^{-1}) responded equivalently to *Physa* exposed to water that never contained a predator.

Mean Behavior: Fine-Grained Variation

In *FGV*, cue magnitude was manipulated to alter environmental variance while keeping mean cue exposure constant over two-day periods (Figure 7). In FGV_{-C} there was no discernable pattern between the amount of environmental variation and mean anti-predator response, implying an approximately linear aggregate anti-predator response across dose. A linear response curve with no variation would result in perfectly equal mean behavior in all treatments because each treatment was exposed to the same total amount of cue. However, due to the presence of substantial variation among individuals, treatments and behaviors, treatment unique means were supported for three response variables, each of which had the largest mean in a different treatment (Table A2.5).

Among-Individual Behavioral Variation: Coarse-Grained Variation

In *CGV* unpredictability in the environment had no clear effects on among-individual variation. Time at max height, max height and hiding observations were the most variable among individuals in the high variation treatment, while both response duration and total vertical

movement were the least variable among individuals in the high variation treatment (Table A2.5). However, more among-individual variation in the high treatment for three behaviors suggests that *Physa* may differ in their inclination to reduce anti-predator behavior following the removal of predator cue. For example, among-individual variation in the ability to reduce responses on non-cue days following cue days will result in a larger potential for among-individual variation in the high variation treatment because there are more environmental transitions. To effectively evaluate the amount of individual variation in the ability to reduce responses on non-cue days following cue days, a modified experimental design with a greater number of transitions between cue and non-cue days for each individual would be required to obtain sufficient power. Additionally, future work should also explore the strength of selection for individuals that can optimize behavioral choices at this resolution of environmental variation. Individuals that are able to quickly evaluate cue levels and respond with appropriate anti-predator behavior, but also reduce their response when cue has dissipated are expected to be selected for in a variableenvironment. With some evidence for heritability of both plasticity and intra-individual variation (Johnstone & Manica, 2011;Wolf, Van Doorn & Weissing, 2011; Kortet et al., 2014;Westneat, Wright & Dingemanse, 2014) evolution for smaller adjustments to non-linear response curves may be detectable.

Among-Individual Behavioral Variation: Fine-Grained Variation

In *FGV* among-individual variation was highest under intermediate levels of environmental variation and intermediate levelsof cue exposure (stochastic and constant treatments) and lowest under high variation and a combination of no variation and no cue exposure (control treatment) (Figure 11A). Low among-individual variation in the high and control treatments suggest that under large amounts of crayfish cue (0.01 snails/L) individuals

responded with strong anti-predator behavior, and when exposed to no cue all individuals climbed infrequently and spent the majority of their time foraging. However, large among-individual variation in the constant and stochastic treatments suggests large variation in the shapes of individual anti-predator response curves at intermediate cue levels (Figure 11B). For example, this pattern of variation can emerge when highly cue-sensitive *Physa* perceive 0.005 snails/L as ahighlyriskyenvironment, leading to quickly saturating convex response curves, and highly cue-insensitive *Physa* perceive the same 0.005 snails/L as a low risk environment, leading to accelerating highly concave functions (Figure 5B).

While our study was not designed to quantify heterogeneity in *Physa* sensitivity to predator cue, *Physa* that were exposed to all cue magnitudes (stochastic treatment) displayed highly diverse response curves across cue magnitudes. Previous work has

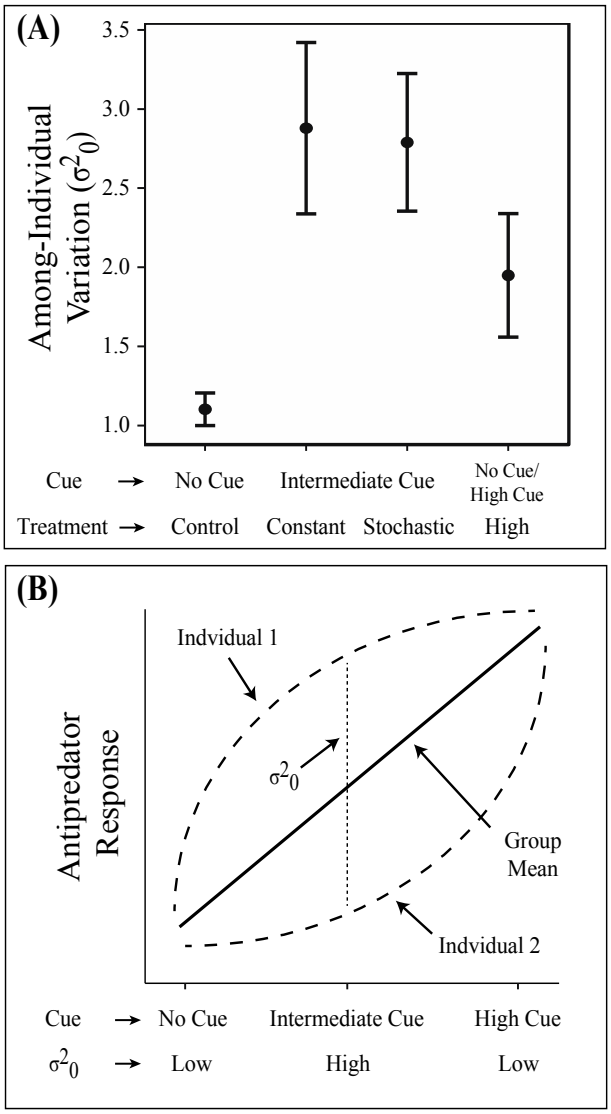

Figure 11: Illustration of the effects of fine-grained variation on among-individual variation in behavior. (A) depicts experimental data of average effect size of σ^2_0 across all eight behaviors. A single effect size is calculated as the ratio between the standard deviation of σ^2_0 between a single treatment and the smallest σ^2_0 among all treatments for a single response behavior. Error bars are 95% confidence intervals. (B) is a hypothetical illustration of a linear group mean response and two non-linear *Physa* response curves. Intermediate cue magnitudes lead to high σ^2_0,as seen by the large effect sizes in the Constant and Stochastic treatments in (A). No cue and a high cue magnitude lead to low σ^2_0, as seen by the small effect sizes in the Control and High variation treatments in (A).

suggested that intermediate levels of environmental variation will produce the largest amount of among-individual variation due to differences in animal cognition (Niemela et al., 2012). All individuals are able to perceive environments at the extremes, but differences in sensory mechanisms and cognitive behavior (e.g. risk assessment) results in a large amount of variation at intermediate environmental levels. These results reveal a promising line of future research to quantify the strength of selection at intermediate levels of fine-grained environmental variation. Additionally, because genes controlling stress and defense in *Physa* have recently been identified (Seok Lee et al., 2011), a combination of genetic work and experimental manipulations maybe able to tie magnitudes of environmental variation to genetic and physiological mechanisms that control the shape of anti-predator response curves.

Within-Individual Behavioral Variation: Coarse- and Fine-Grained Variation

Results from both experiments provide a lack of support for differences in error variance due to environmental variation (Tables 3, A2.3, A2.4). This is unexpected as greater variability in the environment is anticipated to induce more diverse behavior. For example, in *FGVPhysa* exposed to the four different cue concentrations in the stochastic treatment were expected to exhibit greater within-individual variation than *Physa* exposed only to a single cue magnitude in the constant treatment. These results are likely the outcome of a lack of statistical power to detect differences in treatment mean dispersion (β_d). Indeed, averaging within-individual variation across 12-15 highly variable individuals produced similar estimates for β_{dk}, resulting in small effect sizes and low support for treatment unique β_{dk}(Table A2.5).

However, homogenous mean dispersion (β_d) for nearly all response variables may also have a biological basis. For example, if individuals' sensitivity to cue fluctuates greatly throughout the day or is tied to a physiological state such as hunger or fatigue, within-individual

variation may be large and homogenous among individuals. Alternatively, unobserved overnight behavioral choices may have affected the strength of responses observed the following day (e.g. climbing all night could lead to the need to forage regardless of cue environment).

Given the support for heterogeneity in individual anti-predator response curves, environmental variation should also produce heterogeneity among individuals in within-individual variation (φ^2_{dk}). However, we lacked statistical power to model treatment unique individual random effects in the dispersion part of the model (φ^2_{dk}). This is unsurprising given the large sample sizes required to estimate φ^2_{dk} (Cleasby, Nakagawa & Schielzeth, 2015). We provide R scripts in Supplement 1 for modeling treatment unique φ^2_{dk} that may ultimately be important in the evolution of novel behavioral responses to environmental variation (Westneat, Wright & Dingemanse, 2014).

Lethal Predator Trials

On average, snails had the lowest survival after 48 hours in the high variation treatment in both *CGV* (after 48 hours proportion of surviving snails in high: 0/17; low: 5/17; stochastic: 3/17) and *FGV* (constant: 5/13; high: 0/15; stochastic: 0/13; control: 4/14). Additionally, probability of survival decreased most rapidly in the high variation treatment in *FGV* (1.68x greater decrease per minute than the next fastest rate, $p = 0.010$). Yet, survival was not impacted by mean behavior or variation in behavior for any of our *a priori* defined behaviors. This combination of results is highly unexpected because mean anti-predator behavior in *CGV* was highest in the high variation treatment for all behaviors and variance in behavior in the high variation treatment in *FGV* was lowest. These results may stem from one of two alternative scenarios. First, *Physa* that exhibited large anti-predator responses in the high variation

treatments were lower quality with fewer energy reserves. Consequently, after two days of constant predation risk individuals in the high variation treatment faced higher probability of desiccation when above the water line and had to return to a vulnerable height, or had to return to vulnerable heights to forage. Second, there may be a fundamental difference between exposure to cue and exposure to live predators that resulted in no relationship between non-lethal behavioral trials and lethal predator trials.

Despite no evidence for a connection between individual behavioral choices during the twelve days of behavioral trials and mortality, our results show that the magnitude of environmental variation profoundly affected the rate of *Physa* mortality. This result might be important for understanding population dynamics in the face of increasing environmental variation such as predicted to occur with the impending effects of climate change. For example, increased fine-grain variation in weather patterns may drive greater variation in resource availability or predator risk, leading to population and community level effects.

Oviposition Behavior

High fine-grained variation caused *Physa* to lay fewer eggs on average and a smaller proportion of eggs underneath the tile shelter (Table 5), while stochastic variation caused individuals to lay the highest proportion of eggs underneath the tile shelter. *Physa* in the high variation and stochastic variation treatments spent the least and most amount of time underneath the tile shelter, respectively (Table A2.5), suggesting that *Physa* laid their egg masses in their preferred predator avoidance locations. Because snailsdid not differ in size by treatment, fewer egg masses laid by snails in the high variation treatment is suggestive of lower energy reserves or a tradeoff for growth and survival (predator avoidance) over reproduction.

The dual dependence of survival and reproductive potential on the magnitude of environmental variation has large implications for the persistence of *Physa* populations in the face of increasing environmental variation. For example, if the flooding/drying cycle of a rock pool system corresponds to the influx of predators from a neighboring body of water, increasing variation in flooding/drying cycles would cause *Physa* to expose both themselves and their egg masses to increased probability of desiccation.

Conclusion

In this study we show that coarse-grain variation primarily affects average behavior and fine-grain variation drives individual variation in behavior, indicating that both the magnitude of environmental variation as well as subtle differences in the type of variation can have different effects on organism responses. Moreover, we show that environmental variation affects survival and reproductive success in ways that are not fully explained by individual responses. Significant differences in survival and reproductive strategy indicate that increased research effort should be spent on ecological responses to environmental variation, especially in light of substantial theoretical results demonstrating increased prey suppression under pulsed predation regimes (Sih & McCarthy, 2002; Liu & Chen, 2003;Lie, Teng & Chen, 2006; Qian et al., 2009) and projections of changes in environmental variation due to climate change (Highes, 2000; Muller & Sotne, 2001; Luterbacher, 2004). These results also supply critical empirical support for the need to expand the focus in the field of behavioral ecology to include variation *per se* as a predictor of individual variation in behavior and towards quantifying and contrasting the magnitude of among- and within-individual variation among multiple groups of individuals.

CONCLUSION AND FUTURE DIRECTIONS

Variation *per se* has been suggested to play an important role in many ecological responses (e.g. Butler, 1989; Benedetti-Cecchi, 2003; Miner & Vonesh, 2004) including shaping individual-level phenotypic variation (Mathot et al., 2012). Empirical support, however, remains scarce due to difficulties in experimental design conflating mean and variance (e.g. Navarrete, 1996; McCabe & Gotelli, 2000; Sih & McCarthy, 2002; Van Buskirk et al., 2002; Pecor & Hazlett, 2003), and for more recent studies on phenotypic variance, experiments requiring large sample sizes and complex and computationally intensive statistics (e.g. [G]LMM and [D]HGLM).

In this thesis I presented a practical guide and power analysis for GLMMs and advances in experimental methodology that remedy deficiencies in current approaches used to study variation, opening the door for future work investigating novel hypotheses on the effects and consequences of variation at multiple levels. Furthermore, I presented experimental evidence for changes in mean behavior and individual behavioral variance in response to environmental variation that reveals the need for increased research focused on variance as both a predictor and response variable, especially in the field of behavioral ecology.

In chapter one I presented parameterizations and power analyses for GLMM that evaluated power to detect differences in variance by treatment for among-group variation in intercept, within-group variation in intercept, and among-group variation in slope. My results indicate heterogeneity in power across ratios of individuals to repeated measures with an optimal ratio determined by both the target variance parameter and total sample size. Additionally, power to detect each variance parameter was low overall (in most cases >1,000 total observations per

treatment needed to achieve 80% power) and decreased with increasing variance in non-target random effects.

These power analyses assist empiricists in designing novel experiments focused on detecting differences in variance by treatment. However, substantial work is still required to increase the availability and use of existing statistical tools to non-statisticians. First, my power analyses make a series of simplifying assumptions and do not explore many levels of individual variation of interest to behavioral ecologists. For example, I assume no variance in intercept for analysis of variation in slope and a fully balanced design for all analyses. I also do not discuss power to detect covariance between intercept and slope and among-individual variation in intra-individual variation (within-individual "error" variation; Westneat, Wright & Dingemanse, 2014; Cleasby, Nakagawa & Schielzeth, 2015).

Additionally, I conducted my power analyses for GLMM, which is but one approach for modeling hierarchical variance. For example, I analyzed my empirical data presented in Chapter 2 with DHGLMs using a Bayesian approach that allows for the incorporation of treatment unique random effects in the dispersion (error) part of the statistical model. I am aware of only a few resources that assist ecologists in using DHGLMs (Westneat, Schofield & Wright, 2013; Cleasby, Nakagawa & Schielzeth, 2015), and only surface-level power analyses for DHGLMs (Cleasby, Nakagawa & Schielzeth, 2015). Finally, while resources for empiricists are increasing in abundance, many R packages are in their infancy and remain difficult to implement for many statistical problems. For example, the use of DHGLM to model treatment unique random effects is not currently possible in R using a Frequentist approach without substantial programming.

The experiments described in Chapter 2 reveal a relationship between environmental variation, mean behavior, and behavioral variation that is dependent on the type of

environmental variation. As coarse-grained variation increased mean behavior increased; however, coarse-grained variation had no strong relationship with among- or within-individual variation. Conversely, an increase in fine-grained variation led to an increase in among-individual variation but had no discernable impact on mean behavior or within-individual variation. Additionally, I showed that fine-grained variation affected snail reproductive strategy and both types of variation affected snail survival.Changes in the average response versus changes in how variable individuals behave from both one another and from their own mean can have important implications for the evolution of behavioral reaction norms, and for the maintenance of variation in populations. My results suggest that variance estimates for different hierarchical levels (population, among, and within-individuals) may quantify an alternative form of environmentally induced plasticity that will have strong implications for the predicting longer-term evolutionary and ecological consequences of different forms and magnitudes of environmental variation.

The results obtained from this study also supply critical empirical support for the need to extend current approaches used in behavioral ecology for quantifying among-individual variation away from simply testing whether there is significant deviation from a null model of no variation (Martin et al., 2011; Van de Pol et al., 2012; Dingemanse & Dochtermann, 2013) toward quantifying and contrasting the magnitude of among- and within-individual variation among multiple groups of individuals. Furthermore, due to significant our results indicate that increased research effort should be placed on understanding ecological responses to environmental variation, especially in light of substantial theoretical results demonstrating increased prey suppression under pulsed predation regimes (Sih & McCarthy, 2002; Liu & Chen, 2003; Lie et al., 2006; Qian et al., 2009) and projections of increased variation due to climate change.

Yet, these experiments merely scratch the surface of variance focused empirical research. Exceedingly little is known about the interaction of variance at multiple hierarchical levels and no empirical work has evaluated treatment/group differences in the amount of among-individual variation in within-individual "error" variation. I am also aware of no work that quantifies variation in nonlinear reaction norms among-individuals. My research indicates that *Physa* differ in nonlinear anti-predator behavior to predator cue concentration, producing variation among-individuals in responses to variation in predator risk environment. Work quantifying the magnitude of variation in non-linear response curves could help to predict responses to the magnitude of variation in the environment. Additionally, because genes controlling stress and defense in *Physa* have recently been identified (Seok Lee et al., 2011), a combination of genetic work and experimental manipulations may be able to tie magnitudes of environmental variation to rates of gene expression that control the shape of anti-predator response curves. Further experimental work could then determine the fitness consequences of specific levels of mRNA expression and the ability to regulate gene expression on a fine-grained scale in response to environmental variation.

REFERENCES

Barker, R.J. & Link, W.A. 2015. Truth, Models, Model Sets, AIC, and Multimodel Inference: A Bayesian Perspective. *The Journal of Wildlife Management,* 79, 730–738.

Bates, D., Maechler, M., Bolker, B., & Walker S. 2014. Fitting linear mixed-effects models using lme4. arXiv reprint. arXiv: 1406.5823.

Bell, A.M., Hankison, S.J. & Laskowski, K.L. 2009. The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, *77*, 771–783.

Benedetti-Cecchi, L. 2003. The importance of the variance around the mean effect size of ecological processes. *Ecology*, 84, 2335–2346.

Benedetti-Cecchi, L. 2005. The importance of the variance around the mean effect size of ecological processes: Reply. *Ecology*, 86, 265–268.

Benedetti-Cecchi, L., Bertocci, I., Vaselli, S. & Maggi, E. 2006. Temporal variance reverses the impact of high mean intensity of stress in climate change experiments. *Ecology*, 87, 2489-2499.

Bernot, R.J. & Turner, A.M. 2001. Predator identity and trait-mediated indirect effects in a littoral food web. *Oecologia*, 129, 139–146.

Bertocci, I., Vaselli, S., Maggi, E. & Benedetti-Cecchi, L. 2007. Changes in temporal variance of rocky shore organism abundances in response to manipulation of mean intensity and temporal variability of aerial exposure. *Marine Ecology Progress Series*, 338, 11–20.

Biro, P.A. & Adriaenssens, B. 2013. Predictability as a personality trait: consistent differences in intra-individual behavioral variation. *The American Naturalist*, 182, 621–629.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.S.S. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24, 127–135.

Bolker, B.M. 2015. Linear and generalized linear mixed models. *Ecological Statistics: Contemporary theory and application* (eds G.A. Fox, S. Negrete-Yankelevich & V.J. Sosa), pp. 309–334. Oxford University Press, USA.

Bolnick, D.I., Amarasekare, P., Araujo, M.S., Burger, R., Levine J.M., Novak M., Rudolf, V.H.W., Schreiber S.J., Urban, M.C. & Vasseur, D.A. 2011. Why intraspecific trait variation matters in community ecology. *Trends in Ecology and Evolution*, 26, 183–192.

Browne, W.J., Subramanian, S.V., Jones, K. & Goldstein, H. 2005. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 168, 599–613.

Butler, M.J. 1989. Community responses to variable predation: Field studies with sunfish and freshwater macroinvertebrates. *Ecological Monographs*, 59, 311–328

Cernicchiaro, N., Renter, D.G., Xiang, S., White, B.J. & Bello, N.M. 2013. Hierarchical Bayesian modeling of heterogeneous variances in average daily weight gain of commercial feedlot cattle. *Journal of animal science*, 91, 2910–2919.

Chesson, P.L. 1986. Environmental variation and the coexistence of species. Community Ecology, Chapter 14. Harper and Row, 246–256.

Cifuentes, M., Kamlah, C., Thiel, M., Lenz, M. & Wahl, M. 2007. Effects of temporal variability of disturbance on the succession in marine fouling communities in northern-central Chile. *Journal of Experimental Marine Biology and Ecology*, 352, 280–294.

Cleasby, I.R., Nakagawa, S. & Schielzeth, H. 2015. Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, 6, 27–37.

Covich, A.P., Crowl, T.A., Alexander Jr., J.E. & Caryn C. Vaughn. 1994. Predator-avoidance responses in freshwater decapod-gastropod interactions mediated by chemical stimuli.

The North American Benthological Society, 12, 283–290.

DeAngelis, D.L. & Mooij, W.M. 2005. Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution, and Systematics*, 147–168.

DeWitt, T.J., Sih, A. & Hucko, J.A. 1999. Trait compensation and cospecialization in a freshwater snail: size, shape and antipredator behavior. *Animal Behavior*, 58, 397–407.

Dingemanse, N.J., Kazem, A.J.N., Réale, D. & Wright, J. 2010. Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology and Evolution*, 25, 81–89.

Dingemanse, N.J., Bouwman, K.M., van de Pol, M., van Overveld, T., Patrick, S.C., Matthysen, E. & Quinn, J.L. 2012. Variation in personality and behavioural plasticity across four populations of the great tit Parus major. *Journal of Animal Ecology*, 81, 116–126.

Dingemanse N.J. & Dochtermann N.A. 2013. Quantifying variation in behaviour: mixed-effect modeling approaches. *Journal of Animal Ecology*, 82, 39–54.

Dingemanse, N.J. & Wolf, M. 2013. Between-individual differences in behavioural plasticity within populations: causes and consequences. *Animal Behaviour*, 85, 1031–1039.

Estay, S.A., Clavijo-Baquet, S., Lima, M. & Bozinovic, F. 2011. Beyond average: an experimental test of temperature variability on the population dynamics of Tribolium confusum. *Population Ecology*, 53, 53–58.

Fitzmaurice, G.M., Laird, N.M. & Ware, J.H. 2004. *Applied Longitudinal Analysis*, 2nd edn. Wiley, New Jersey, USA.

Field, C.B., Chapin III, F.S., Matson, P.A. & Mooney, H.A. 1992. Responses of terrestrial ecosystems to the changing atmosphere: a resource-based approach. *Annual Review of Ecology and Systematics*, 23, 201–235.

Foray, V., Desouhant, E. & Gibert, P. 2014. The impact of thermal fluctuations on reaction

norms in specialist and generalist parasitic wasps. *Functional Ecology*, 28, 411–423.

Gelman, A. & Hill, J. 2006. *Data Analysis Using Regression and Multi- level/Hierarchical Models*. Cambridge University Press, England.

Gelman, A. & Shalizi, C.R. 2013. Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66, 8–38.

Ghalambor, C.K., Angeloni, L.M. & Carroll, S.P. 2010. Behavior as phenotypic plasticity. *Evolutionary behavioral ecology* (eds D.F. Westneat & C.W. Fox), pp. 90–197. Oxford University Press, USA.

Grimm, V. & Railsback, S.F. 2005. Individual-based Modeling and Ecology: Princeton Series in Theoretical and Computational Biology. Princeton University Press.

Hill, W.G. & Zhang, X.S. 2004. Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genetical Research*, 83, 121–132.

Hinde, J. 1982. Compound Poisson regression models. *GLIM 82: Proceedings of the International Conference on Generalised Linear Models* (ed R. Gilchrist), pp. 109–121. Springer, New York.

Hooper, D.U., Chapin III, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, Naeem, S., Schmid, B., Setala, H., Symstad, A.J., Vandermeer, J. & Wardle, D.A. 2005. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. Ecological Monographs, 75, 3–35.

Hughes, L. 2000. Biological consequences of global warming: is the signal already apparent?. *Trends in Ecology and Evolution*, 15, 56–61.

Inouye, B.D. 2001. Response surface experimental designs for investigating interspecific competition. *Ecology*, 82, 2696–2706.

Inouye, B.D. 2005. The importance of the variance around the mean effect size of ecological

processes: comment. *Ecology*, 86, 262–265.

Jennions, M. & Møller, A. 2003. A survey of the statistical power of research in behavioral ecology and animal behavior. *Behavioral Ecology*, 14, 438–445.

Johnson, P.C., Barry, S.J., Ferguson, H.M. & Müller, P. 2014. Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution*, 6, 133–142.

Johnstone, R.A. & Manica, A. 2011. Evolution of personality differences in leadership. *Proceedings of the National Academy of Sciences*, 108, 8373–8378.

Kizilkaya, K. & Tempelman, R.J. 2005. A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetics Selection Evolution*, 37, 31–56.

Kortet, R., Vainikka, A., Janhunen, M., Piironen, J. & Hyvärinen, P. 2014. Behavioral variation shows heritability in juvenile brown trout Salmo trutta. *Behavioral Ecology and Sociobiology*, 68, 927–934.

Lawson, C.R., Yngvild V., Bailey, L. & van de Pol, M. 2015. Environmental variation and population responses to global change. *Ecology Letters*, doi: 10.1111/ele.12437

Lee, Y.S., Lee, S.G., Kang, S.W., Jeong, J.E., Baek, M K., Choi, S.H., Chae, S.H., Jo, Y.H. & Park, H.S. 2011. Expressed sequence tag analysis of Physa acuta: a freshwater pulmonate in Korea. *Journal of Shellfish Research*, 30, 127–132.

Lee, Y. & Nelder, J.A. 2006. Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series C*, 55, 139–185.

Levin, S.A. 1998. Ecosystems and the biosphere as complex adaptive systems. *Ecosystems*, 1, 431–436.

Lie, B., Teng, Z. & Chen, L. 2006. Analysis of a predator-prey model with Holling II functional

response concerning impulsive control strategy. *Journal of Computational and Applied Mathematics*, 193, 347–362.

Liu, X. & Chen, L. 2003. Complex dynamics of Holling type II Lotka-Volterra predator-prey system with impulsive perturbations of the predator. *Chaos, Solitons and Fractals*, 16, 311–320.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M. & Wanner, H. 2004. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science*, *303*, 1499–1503.

Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*, 1st edn. Sinauer Press, Sunderland.

Martin, J.G.A. & Réale, D. 2008. Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. *Animal Behaviour*, 75, 309–318.

Martin, J.G.A., Nussey, D.H., Wilson, A.J. & Réale, D. 2011. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. *Methods in Ecology and Evolution*, 2, 362–374.

Mathot, K.J., van den Hout, P.J., Piersma, T., Kempenaers, B., Réale, D. & Dingemanse, N.J. 2011. Disentangling the roles of frequency-vs. state-dependence in generating individual differences in behavioural plasticity. *Ecology letters*, 14, 1254–1262.

Mathot, K.J., Wright, J., Kempenaers, B. & Dingemanse, N.J. 2012. Adaptive strategies for managing uncertainty may explain personality-related differences in behavioural plasticity. *Oikos*, 121, 1009–1020.

McCabe, D.J. & Gotelli, N.J.. 2000. Effects of disturbance frequency, intensity, and area on assemblages of stream macroinvertebrates. *Oecologia*, 124, 270–279.

Miner, B.G. & Vonesh, J.R. 2004. Effects of fine grain environmental variability on

morphological plasticity. *Ecology Letters*, 7, 794–801.

Menge, B.A. & Sutherland, J.P. 1987. Community regulation: variation in disturbance, competition, and predation in relation to environmental stress and recruitment. *The American Naturalist*, 130, 730–757.

Moineddin, R., Matheson, F.I. & Glazier, R.H. 2007. A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, 7, 1–34.

Moore, A.J., Brodie, E.D. III & Wolf, J.B. 1997. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution*,51, 1352–1362.

Nakagawa, S. & Schielzeth, H. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, 85, 935–956.

Navarrete, S.A. 1996. Variable predation: effects of whelks on a mid-intertidal successional community. *Ecological Monographs*. 66, 301–321.

Niemelä, P.T., Vainikka, A., Forsman, J.T., Loukola, O.J. & Kortet, R. 2013. How does variation in the environment and individual cognition explain the existence of consistent behavioral differences?. *Ecology and Evolution*, 3, 457–464.

Nussey, D.H., Wilson, A.J. & Brommer, J.E. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*,20, 831–844.

Paaijmans, K.P., Heinig, R.L., Seliga, R.A., Blanford, J.I., Blanford, S., Murdock, C.C. & Thomas, M.B. 2013. Temperature variation makes ectotherms more sensitive to climate change. *Global Change Biology*, 19, 2373–2380.

Peacor, K.W. & Hazlett, B.A. 2003. Frequency of Encounter with Risk and the Tradeoff between Pursuit and Antipredator Behaviors in Crayfish: A Test of the Risk Allocation Hypothesis. *Ethology*, 109, 97–106.

Pigliucci M. 2001. *Phenotypic Plasticity: Beyond Nature and Nurture*. John Hopkins University Press, Baltimore, MD.

Pinheiro, J. & Bates, D. 2001. *Mixed-Effects Models in S and S-PLUS*. Springer, New York.

Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, 2003. URL http://sourceforge.net/projects/mcmc-jags/.

Plummer, M. 2014. rjags: Bayesian graphical models using MCMC. R package version 3-14. http://CRAN.R-project.org/package=rjags

Qian, L., Lu, Q., Meng, Q. & Feng, Z. 2010. Dynamical behaviors of a prey–predator system with impulsive control. *Journal of Mathematical Analysis and Applications*, 363, 345–356.

R Development Core Team. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org

Royauté, R. & Pruitt, J.N. *In Press*. Varying predator personalities generates contrasting prey communities in an agroecosystem. *Ecology*. http://dx.doi.org/10.1890/14-2424.1

Ruel, J.J. & Ayres, M.P. 1999. Jensen's Inequality predicts effects of environmental variation. *Trends in Ecology and Evolution*, 14, 361–366.

Scherbaum, C.A. & Ferreter, J.M. 2009. Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, 12, 347–367.

Schielzeth, H. & Nakagawa, S. 2013. Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, 4, 14–24.

Schoeppner, N.M. & Relyea, R.A. 2009. Phenotypic plasticity in response to fine-grained environmental variation in predation. *Functional Ecology*, 23, 587–594.

Sih, A. & McCarthy, T.M. 2002. Prey responses to pulses of risk and safety: testing the risk

allocation hypothesis. *Animal Behavior*, 63, 437–443.

Sih, A., Mathot, K.J., Moirón, M., Montiglio, P.O., Wolf, M. & Dingemanse, N.J. 2015. Animal personality and state–behaviour feedbacks: a review and guide for empiricists. *Trends in Ecology and Evolution*, 30, 50–60.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.

Stacey, P.B. & Taper, M. 1992. Environmental variation and the persistence of small populations. *Ecological Applications*, 2, 18–29.

Stamps, J.A., Briffa, M. & Biro, P.A. 2012. Unpredictable animals: individual differences in intraindividual variability (IIV). *Animal Behaviour*, 83, 1325– 1334.

Stroup, W.W. 2012. *Generalized linear mixed models: modern concepts, methods and applications*. CRC Press, USA.

Swetnam, T.W. & Betancourt, J.L. 1997. Mesoscale disturbance and ecological response to decadal climatic variability in the American Southwest. *Journal of Climate*, 11, 3128–3147.

Su, Y. & Yajima, M. 2015. R2jags: A Package for Running 'JAGS' from R. R package version 0.05-03. http://CRAN.R-project.org/package=R2jags

Tonsor, S.J., Elnaccash, T.W. & Scheiner, S.M. 2013. Developmental instability is genetically correlated with phenotypic plasticity, constraining heritability, and fitness. *Evolution*,67, 2923–2935.

Turner, A.M., Turner, S.E. & Lappi, H.M. 2006. Learning, memory and predator avoidance by freshwater snails: effects of experience on predator recognition and defensive strategy. *Animal Behavior*, 72, 1443–1450.

van de Pol, M. & Wright, J. 2009. A simple method for distinguishing within-versus between-subject effects using mixed models. *Animal Behaviour*, 77, 753–758.

Van Buskirk, J., Muller, C., Portmann, A. & Surbeck, M. 2002. A test of the risk allocation hypothesis: tadpole responses to temporal change in predation risk. *Behavioral Ecology*, 13, 526–530.

Vasseur, D.A., DeLong, J.P., Gilbert, B., Greig, H.S., Harley, C.D., McCann, K.S., Savage, V., Tunney, T.D. & O'Connor, M.I. 2014. Increased temperature variation poses a greater risk to species than climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*, 20132612.

Verbeke, G. & Molenberghs, G. 2000. *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Westneat, D.F., Hatch, M.I., Wetzel, D.P. & Ensminger, A.L. 2011. Individual variation in parental care reaction norms: integration of personality and plasticity. *TheAmerican Naturalist*, 178, 652–667.

Westneat, D.F., Schofield, M. & Wright, J. 2013. Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, 24, 598-604.

Westneat, D.F., Wright, J. & Dingemanse, N.J. 2014. The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*. doi:10. 1111/brv.12131

Wolf, M. & Weissing, F.J. 2010. An explanatory framework for adaptive personality differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 3959–3968.

Wolf, M., Van Doorn, G.S. & Weissing, F.J. 2011. On the coevolution of social responsiveness and behavioural consistency. *Proceedings of the Royal Society of London B: Biological Sciences*, 278, 440–448.

Wolf, M. & Weissing, F.J. 2012. Animal personalities: consequences for ecology and evolution. *Trends in Ecology & Evolution*, 27, 452-461.

Zuur, A.F., Leno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. 2009. *Mixed effects models and extensions in ecology with R*. Springer, New York.

Zuur, A.F., Hilbe, J.M. & Leno, E.N. 2013. *A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists*. Highland Statistics Ltd, Newburgh, UK.

APPENDIX A: SUPPLEMENTARY TUTORIAL FOR " A PRACTICAL GUIDE AND POWER ANALYSES FOR GLMMs: DETECTING AMONG TREATMENT VARIATION IN RANDOM EFFECTS"

Introduction

The goal of this document is to provide the R code necessary to conduct the power analyses described in the main paper and instructions on how to sculpt simulations for random intercepts and random slopes GLMM for your specific research question.

Here we present simulations for a GLMM object using newly developed simulation capabilities of the lme4 package. We provide a worked example for simulating data and running a power analysis for detecting differences in variation (among-individual variation in reaction norm intercept) by treatment. For each of the other target reaction norm variance parameters we provide the lme4 syntax required to simulate and analyze data for comparisons of that variance parameter.

This tutorial is broken into five main segments:

1) Worked example for among-individual variation in intercept (σ^2_{0k})
    a) Data simulation and power analysis for a specific sampling scheme and effect size
    b) Quick extraction and visualization of results
2) Parameterization for the other two variance comparisons
    a) Among-individual variation in slope (σ^2_{1k})
    b) Within-individual variation in intercept (σ^2_{vk})
3) Syntax for a three-treatment model
4) Example nested loop structure for full simulation
5) Supplemental Table 1: Full parameter values for all simulations presented in the main text

1) Worked example for among-individual variation in intercept (σ^2_{0k})

For this worked example we discuss a two-treatment experiment whose goal is to determine if the magnitude of among-individual variation differs by treatment. We describe an experiment with a binomially distributed response variable. An example of an experiment that fits these criteria is an experiment testing if individual freshwater snails behave more/less variably from one another when exposed to predator cue than when exposed to no predator cue. Freshwater snails respond to predation risk by pumpkinseed sunfish by hiding under debris. The proportion of time individuals are observed to be hiding underneath a tile shelter following exposure to cue from a sunfish is one possible binomial response variable.

To conduct these simulations and power analyses you will need a recent version of `lme4`.

1) Step one is to set up a data frame containing information about the experimental design (number of treatments, number of individuals per treatment, number of observations per individual): expand.grid() is useful. However, note that expand.grid() results in a fully balanced design. To allow for some variation in the number of observations per individual, some modification will be needed.

If you have any continuous predictors or covariates, then you need to figure out what the distribution of that is going to be: is it regular/linear, or Normally distributed, or uniform? To include observation-level random effects/overdispersion (within-individual variation), then add an obs variable to the data frame which is defined simply as factor(seq(nrow(your_data))).

For this worked example we only use a single value for each experimental design variable. Once you have a simulation/estimation procedure working for a particular set of variables, you can

embed it in a large, nested for loop that explores the whole range of experimental design variables you are interested in (e.g. effect size, variance, number of blocks, samples per block, etc.). An example nested for loop is provided in section 4 of this supplement.

Each variable value stored below corresponds to a single treatment. For example rep_vec sets the number of individuals sampled in *each treatment.* In the two treatment model presented below TSS for experiment is (rep_vec)*(repeat_vec)*2 = 180

Number of individuals

```
> rep_vec <- 15
```

Number of repeated measures within each individual

```
> repeat_vec <- 6
```

Among-individual variation in treatment 1 – Hereby referred to as the homogeneous (low-variation) treatment

```
> theta_among.homog_vec <- 0.2
```

Among-individual variation in treatment 2 – Hereby referred to as the variable (high-variation) treatment

```
>theta_among.var_vec <- 0.4
```

Observation level variation/overdisperison (within-individual variation)

```
>theta_obs_vec <- 0.2
```

Set up data frame:

```
> expdat <- expand.grid(indiv =factor(seq(rep_vec)),
+ obs =seq(repeat_vec), ttt = c("homog", "var"))

> expdat <- transform(expdat, homog = as.numeric((ttt ==
+     "homog")), var = as.numeric((ttt == "var")))

> expdat$total_obs <- factor(seq(nrow(expdat)))
```

We generally find it most convenient to store everything in a large multi-dimensional array, with one dimension for each experimental design variable, a dimension for replicates, and as many dimensions as necessary for the information you want to save about each replicate. For example, if you will be considering 10 possible numbers of samples per block, 10 possible numbers of blocks, and 5 possible effect sizes, doing 1000 replicates for each combination, and you wanted to keep information about the mean and standard deviation of 3 different parameters, you would end up with a 10 x 10 x 10 x 5 x 1000 x 3 x 2 array (Note this is an array of 30 *million* elements, representing 5,000,000 simulation runs – it's easy to get carried away with this sort of experiment!). Make sure to give dimnames to the array, where each element in the list itself has a name. For example, the full array used in the simulation run for the power analyses results presented in the paper was as follows:

```
> power_sim <- array(NA,dim=c(length(theta_among.homog_vec),
+     length(theta_among.var_vec), length(theta_obs_vec),
+     length(rep_vec), length(repeat_vec),nsim, 9),
+     dimnames=list(theta_h=theta_among.homog_vec,
+     theta_var=theta_among.var_vec,
+     theta_obs=theta_obs_vec,num.indiv=rep_vec,
+     repeat.meas=repeat_vec, sim.count=seq(nsim),
+     var=c("est","stderr","zval","ttt.pval","obsvar",
+ "indivvar.homog","indivvar.var","devdiff","var.pval")))
```

Keeping the data in an array this way makes it easy to select and/or average across the slices you want; when you want to convert the data to long format for analysis or plotting with lattice or ggplot, just use reshape2::melt(). A brief example using melt() and ggplot to visualize results is provided in supplement section 1b.

2) Specify the parameters: "theta" - in the case of single variance terms, which is just the standard deviation of each random effect: e.g. theta=c(1, 0.5) (among-individual variation is 4x among-observation variance). "beta" is the fixed-effects parameters on the logit scale, in this case (intercept, treatment).

```
>nsim <- 20
> beta <- c(0.5, 0)
> theta <- c(theta_obs_vec, theta_among.var_vec,
+     theta_among.homog_vec)
>.progress <- "text"
>verbose <- TRUE
```

Set up a matrix of NA in case of failure in model fitting

```
>errmat <- matrix(NA,nrow=nsim,ncol=9)
```

3) Set up the formula corresponding to the model you want to fit: For this example we want to simulate and estimate a case where the among-individual variation in intercept differs by treatment. See the posts on r-sig-mixed-models by David Afshartous <http://thread.gmane.org/gmane.comp.lang.r.lme4.devel/214> for additional information on how this works. Essentially, you have to construct your own numeric dummy variables. Simulate on this basis: simulate(formula, newdata, parameters, family = binomial) returns a response vector. See ?simulate.merMod for an additional example.

```
> ss2 <-simulate(~ttt+(0+homog|indiv:ttt)+(0+var|indiv:ttt)+
+     (1|total_obs),nsim=nsim,family=binomial,
+     weights=rep(5,nrow(expdat)),
+     newdata=expdat,newparams=list(theta=theta2,beta=beta))

>expdat$resp <- ss2[[1]]
```

4) Run the glmer and extract p-values and other model output.

Model for heterogeneous among-individual variation by treatment:

```
>fitfun <- function(expdat,i,.progress="text",verbose=TRUE){
+     fit2 <- try(glmer(resp~ttt+(0+homog|indiv:ttt)+          +
(0+var|indiv:ttt)+(1|total_obs),family=binomial,
+     weights=rep(5,nrow(expdat)),
+     data=expdat),silent=TRUE)
+           if(is(fit2,"try-error")) return(errmat)
+     fit2B <- try(update(fit2,.~ttt+(1|indiv:ttt)+(1|total_obs)),
+     silent=TRUE)
+      if (is(fit2B,"try-error")) return(errmat)
+
```

```
+    fits=laply(seq(nsim),function(i)        +
fitsim2(i,models=list(fit2,fit2B)),.progress=.progress)
+    return(fits)
+}
```

Fitting function to compare models:

```
>fitsim2 <- function(i,models=list(fit2,fit2B)) {
+    r1 <- try(refit(models[[1]],ss2[[i]]),silent=TRUE)
+    if (is(r1,"try-error")) return(rep(NA,9))
+    r1B <- try(refit(models[[2]],ss2[[i]]),silent=TRUE)
+    ss <- try(summary(r1))
+    cc <- if (is(ss,"try-error")) rep(NA,4) else             +
coef(summary(r1))["tttvar",]
+    res <- c(cc,unlist(VarCorr(r1)))
+ if (is(r1B,"try-error")) return(c(res,rep(NA,2)))
+ aa <- anova(r1,r1B)[2,]
+ devdiff <- unlist(c(aa["Chisq"]))
+var.pval <- unlist(c(aa["Pr(>Chisq)"]))
+return(c(res,devdiff,var.pval))
+}
```

Run the power analysis and save the results. Here our power_sim array only has a single dimension for each experimental design variable.

```
> power_sim[1,1,1,1,1,,] <- fitfun(expdat, i = i)
```

Obtain power using reshape2::melt().

```
>d2 <- melt(power_sim, id.var = "var")
>d3 <- data.frame(d2[1:20, 1:6],
+    est = d2[d2$var=="est", ]$value,
+    stderr = d2[d2$var=="stderr", ]$value,
+    var.pval = d2[d2$var=="var.pval", ]$value)
>with(d3, mean(var.pval < 0.05))
```

Estimates:

```
> ggplot(arrange(d3, est), aes(x = seq(nsim), y = est,
+    ymin = est - 1.96 * stderr, ymax = est + 1.96 *               +
stderr)) + geom_pointrange() +
+    geom_hline(yintercept = mean(d3$est), colour = "red") +
+    theme_bw()
```

```
>truevals <- data.frame(variable = c("indivvar.homog",
+    "indivvar.var"), trueval = c(mean(d3$indivvar.homog),        +
     mean(d3$indivvar.var)))

>d4 <- d2[d2$var=="indivvar.homog" | d2$var=="indivvar.var", ]

>ggplot(d4, aes(x = sim.count, y = value, colour = var)) +
+    geom_point(lwd = 3) + geom_hline(data = truevals,             +
aes(yintercept = trueval, colour = variable), lty = 2) +
+theme_bw()
```

2) Parameterization for the other two variance comparisons

a) Among-Individual Variation in Slope

To model among-individual variation in slope the expand.grid() statement must be modified to treat obs as a sequential measurement.

```
> expdat <- expand.grid(indiv=factor(seq(rep_vec[i])),
+ obs=seq(repeat_vec[j]), ttt=c("homog","var"))

> expdat <- transform(expdat,
+ homog=as.numeric((ttt=="homog")),
+ var=as.numeric((ttt=="var")))

> expdat$total_obs <- factor(seq(nrow(expdat)))

> expdat <- transform(expdat, homogobs = homog*obs, varobs =
+ var*obs)
```

glmer code to model among-individual variation in slope:

```
> fit2 <- try(glmer(resp~ttt+(0+homogobs|indiv:ttt)+          +
(0+varobs|indiv:ttt)+(1|total_obs),
+    weights=rep(5,nrow(expdat)),
+    family=binomial,
+    data=expdat),silent=TRUE)
```

```
> fit2B <- try(update(fit2,.~ttt+(0+obs|indiv:ttt)+
+ (1|total_obs)),silent=TRUE)
```

b) Within-Individual Variation in Intercept

To model within-individual variation in intercept a new parameter is needed for the number of Bernoulli observations per sampling occasion (weight)

```
> fit2 <-try(glmer(resp~ttt+(1|indiv)+(0+homog|total_obs)+
      + (0+var|total_obs),family=binomial,weights=rep(weight[j],
+ nrow(expdat)),data=expdat),silent=TRUE)

> fit2B <- try(update(fit2,.~ttt+(1|indiv)+
+     (1|total_obs)),silent=TRUE)
```

3) Syntax for a three-treatment model

We will call our three treatments for the three-treatment model low, mid, and high

expand.grid() statement for a three treatment model:

```
> expdat <- expand.grid(indiv =factor(seq(rep_vec)),
+ obs =seq(repeat_vec), ttt = c("low", "mid", "high"))

> expdat <- transform(expdat,
+      low = as.numeric((ttt == "low")),
+      mid = as.numeric((ttt == "mid")),
+      high = as.numeric((ttt == "high")))

> expdat$total_obs <- factor(seq(nrow(expdat)))
```

lme4 syntax can be manipulated to provide estimates for any combination of treatments for any or all of the random effects included in the model. For a three treatment model for the random-intercept model presented previously, unique random effects can be estimates for each treatment for either among-individual or within-individual variation or both.

```
> glmer(resp~ttt+(0+low|indiv:ttt)+                              +
(0+med|indiv:ttt)+(0+high|indiv:ttt)+(1|total_obs),              +
family=binomial,
```

```
+     weights=rep(repeat_vec,nrow(expdat)),
+     data=expdat),silent=TRUE)
```

For a unique random effect estimate for a single treatment (here the high treatment) and a single random effect estimate for the other two treatments (low and med treatments), a new "dummy variable" column is needed (containing a 1 for both low and med treatments and a 0 for the high treatment). Here we call this new column low_med

```
> glmer(resp~ttt+(0+high|indiv:ttt)+                                    +
(0+low_med|indiv:ttt)+(1|total_obs),                                          +
family=binomial,
+     weights=rep(repeat_vec,nrow(expdat)),
+     data=expdat),silent=TRUE)
```

Note that under LRTs where the difference in degrees of freedom between models > 1, more complicated adjustments are needed to correct for the conservative nature of the LRT when parameters are at the boundary of conceivable space (See Zuur et al., 2009)

4) Example Nested Loop Structure for Full Simulation

```
>for (t1 in seq_along(theta_among.homog_vec)) {
+  for (t2 in seq_along(theta_among.var_vec)) {
+    for(t3 in seq_along(theta_among.obs_vec)){
+      for (i in seq_along(rep_vec)) {
+        for (j in seq_along(repeat_vec)){
+
+      if (verbose) cat(t1,"/",length(theta_among.homog_vec)," ",
+                       t2,"/",length(theta_among.var_vec)," ",
+                       t3,"/",length(theta_among.obs_vec)," ",
+                       i,"/",length(rep_vec)," ",
+                       j,"/",length(repeat_vec)," ",
+                       "\n",
+                       sep="")
+
+     expdat <- expand.grid(indiv=factor(seq(rep_vec[i])),
+                           obs=seq(repeat_vec[j]),
+                           ttt=c("homog","var"))
+
+expdat <- transform(expdat,
+                           homog=as.numeric((ttt=="homog")),
+                           var=as.numeric((ttt=="var")))
+     expdat$total_obs <- factor(seq(nrow(expdat)))
+
```

```
+     theta2 <- c(theta_among.obs_vec[t3],
+theta_among.var_vec[t2],
+theta_among.homog_vec[t1])
+
+     ss2 <- simulate(~ttt+(0+homog|indiv:ttt)+(0+var|indiv:ttt)+  +
                (1|total_obs),
+                         nsim=nsim,
+                      family=binomial,       +
weights=rep(5, nrow(expdat)),
+                         newdata=expdat,
+                      newparams=list(theta=theta2,                 +
                beta=beta))
+
+expdat$resp <- ss2[[1]]
+
+            power_sim[t1,t2,t3,i,j,,] <- fitfun(expdat, i = i)
+
+ }
+}
+}
+}
+ }
+    path <-"your_path.Rdata"
+    save("power_sim", file=path)
```

APPENDIX B: SUPPLEMENTARY TUTORIAL FOR "THE EFFECTS OF ENVIRONMENTAL VARIATION ON INDIVIDUAL VARIATION USING R AND JAGS"

Introduction

The goal of this document is to present details on the statistical methodology used beyond what is provided in the primary manuscript. This supplement is broken into five primary sections:

1) Description and R code for the method used to fit logistic curves to snails' climbing behavior

2) Sample JAGS model code

A) Table S1: Description of all candidate JAGS models used to determine the most parsimonious variance and fixed effects structure for each response variable

B) JAGS code for all 8 candidate models used to determine the most parsimonious variance structure of the Normally distributed "time at max height" snail behavior. Models 2, 4, 6 and 8 correspond to the 4 models fit in the primary manuscript

C) JAGS code for the 4 candidate models used in the primary manuscript to determine the most parsimonious variance structure of the Poisson distributed "observations spent foraging" snail behavior

4) Results of Snail:Day Censoring

A) Table S4

5) Summary of all ΔDIC values for all models and summary of effect sizes for most parsimonious models

6) Additional analysis for the "Coarse-Grained" variation experiment evaluating the significance of treatment level differences in behavior in the three measurements prior to cue addition

1) Fitting algorithm for snail climbing curves

To fit logistic curves of the form: [(b*k*r*exp(r*x))/((b+exp(r*x))^2)] to the climbing patterns of each snail on each day (snail:day) we used a two step approach.

1) First, we summed the snails height across time and fit the integral of the desired logistic: [k/(1+b*exp(-r*x))] to the cumulative snail height

2) Using the parameters from the fitted cumulative logistic curve as starting parameters, we fit logistic curves: [(b*k*r*exp(r*x))/((b+exp(r*x))^2)] to the raw snail climbing data

We used the package `nlmrt` (Nash 2014) to fit logistic curves using non-linear least squares. For some snail:days, `nlmrt` produced fitted curves that greatly overestimated the height to which snails climbed (~ 8% of all snail:days). For these snail:days small adjustments were made to reduce the peak of the fitted logistic curve to the actual max height snails reached. In all cases only the magnitude of the peak was adjusted. Time at max height, time at first response, and the duration of response remained unaltered. To ensure that these adjustments were not differentially affecting the results of these snail:days relative to snail:days where no adjustments were made (~ 92% of all snail:days), we verified that for all adjusted snail:days the parameters given by `nlxb` and our adjusted parameters resulted in undetectable differences in likelihood using `mle2` (Bolker, 2014).

R code for step 1:

```
# fit cumulative curves with nlxb (robust nls)

for (i in unique(experiment_2$individual)){
    for (j in unique(experiment_2$day)){

 temp_dat <- experiment_2[experiment_2$individual == i &
                               experiment_2$day == j, ]

 # The cumulative logistic curves fit successfully with simply
      # derived starting parameters

 nls_logistic <- try(nlxb(location ~ k / (1 + b * exp(-r * time)),
                       data = temp_dat, start = list(k =
                        max(temp_dat$location), b = 3, r = .01)), silent
                        = TRUE)

 # Results are stored in an array named param_array

    param_array[i, j, ] <- c(coef(nls_logistic)[[1]],
      coef(nls_logistic)[[2]], coef(nls_logistic)[[3]])

        }
      }
```

param_array is melted using `melt::reshape2` and arranged into a data frame named coef_exp_2 to facilitate fitting in step 2

R code for step 2:

```
# counter for adjustment_check

z = 1

for (i in unique(experiment_2_2$individual)){
for (j in unique(experiment_2_2$day)){

# set up temporary data frame with a single snail:day

temp_dat <- experiment_2_2[experiment_2_2$individual == i &
      experiment_2_2$day == j, ]

# attempt to fit a logistic curve using the starting parameters from
    # step 1

nls_logistic <- try(nlxb(location ~ ((b*k*r*exp(r*time))/
((b+exp(r*time))^2)),
```

```
data = temp_dat,start = list(k =
(coef_exp_2[coef_exp_2$individual == i & coef_exp_2$day == j,4][1]*
(max(temp_dat$time)/length(unique(temp_dat$time)))),
                                b = coef_exp_2[coef_exp_2$individual ==
                              i &coef_exp_2$day == j,4][2],
r = coef_exp_2[coef_exp_2$individual ==
                                   i &coef_exp_2$day == j,4][3])),
                              silent = TRUE)

# first deal with the situation where nlxb doesnt converge

if (inherits(nls_logistic, "try-error")) {

# if the snail never climbs simply fit a flat line at 0

if (sum(temp_dat[ , 9]) == 0) {

param_array[i, j, ] <- c(0, 0, 0)

} else if ((sum(temp_dat[ , 9]) > 0)){

# if the snail did climb, use the parameters from the fitted cumulative
  # curveadjusted for height

temp_max_height <- max(log_cur_fd(seq(0,max(temp_dat$time), by =0.1),
                         k = coef_exp_2[coef_exp_2$coefs == "k" &
                               coef_exp_2$individual == i &
                               coef_exp_2$day == j, 4],
                         b = coef_exp_2[coef_exp_2$coefs == "b" &
                               coef_exp_2$individual == i &
                               coef_exp_2$day == j, 4],
                         r = coef_exp_2[coef_exp_2$coefs == "r" &
                               coef_exp_2$individual == i &
                               coef_exp_2$day == j, 4]))

q <- 1

# using small incremental adjustments determine the point at which the
# curve aligns better with the snail max height and adjust the k
# parameter accordingly

while (temp_max_height < max(temp_dat$location)){

temp_max_height <- max(log_cur_fd(seq(0,max(temp_dat$time), by =0.1),
                              k = (coef_exp_2[coef_exp_2$coefs == "k" &
                                    coef_exp_2$individual == i &
                                    coef_exp_2$day == j, 4])*q,
                              b = coef_exp_2[coef_exp_2$coefs == "b" &
                                    coef_exp_2$individual == i &
                                    coef_exp_2$day == j, 4],
                              r = coef_exp_2[coef_exp_2$coefs == "r" &
                                    coef_exp_2$individual == i &
                                    coef_exp_2$day == j, 4]))

q <- q + .1

}
```

```
adjustment_check[z] <- c(i, j)

param_array[i, j, ] <- c((coef_exp_2[coef_exp_2$coefs == "k" &
                              coef_exp_2$individual == i &
                              coef_exp_2$day == j, 4])*q,
                         coef_exp_2[coef_exp_2$coefs == "b" &
                         coef_exp_2$individual == i &
                              coef_exp_2$day == j, 4],
                         coef_exp_2[coef_exp_2$coefs == "r" &
                              coef_exp_2$individual == i &
                              coef_exp_2$day == j, 4])

}

# if nlxb does converge

} else if (inherits(nls_logistic, "nlmrt")) {

# if the snail never climbs simply fit a flat line at 0

if (sum(temp_dat[ , 9]) == 0) {

param_array[i, j, ] <- c(0, 0, 0)

# if the snails did climb but the nlxb object greatly over-estimates
   # the height climbed to
# determined by a height that is higher than the take allows (>
       # 30cm)

} else if (

max(log_cur_fd(seq(0,max(temp_dat$time), by = 0.1),
      k = coef(nls_logistic)[[1]],
      b = coef(nls_logistic)[[2]],
      r = coef(nls_logistic)[[3]]))> 30) {

temp_max_height <-  max(log_cur_fd(seq(0,max(temp_dat$time), by = 0.1),
      k = coef(nls_logistic)[[1]],
      b = coef(nls_logistic)[[2]],
      r = coef(nls_logistic)[[3]]))

q <- 1

# reduce the height until it is at the height the snail obtained

while (temp_max_height > max(temp_dat$location)){

temp_max_height <-  max(log_cur_fd(seq(0,max(temp_dat$time), by = 0.1),
      k = coef(nls_logistic)[[1]]/q,
      b = coef(nls_logistic)[[2]],
      r = coef(nls_logistic)[[3]]))

      q <- q + .1

}
```

```
adjustment_check[z] <- c(i, j)

param_array[i, j, ] <- c(
      coef(nls_logistic)[[1]]/q,
      coef(nls_logistic)[[1]],
      coef(nls_logistic)[[1]])

# if no adjustments are needed simply store the parameters from the
   # fitted nlxb object

} else {

param_array[i, j, ] <- c(coef(nls_logistic)[[1]],
      coef(nls_logistic)[[2]],
      coef(nls_logistic)[[3]])

}
}
z = z + 1
}
}
```

2) JAGS models for variance structure for the snail behavior "time at max height"

Here we present JAGS code for all 8 candidate models that can be used to determine the most parsimonious variance structure when homogenous or heterogeneous by treatment individual random effect, treatment mean error variance and individual random effect for error variance is being investigated.

Table A2.1: Candidate models needed to determine to most parsimonious model for the variance structure when random effects in the dispersion part of the model (error variance) is also in question. A total of eight models with different variance structures are needed in this case. An X indicates a variance structure or fixed effect included in the model. Here we depict model 2 as the model that best describes the variance structure.

Error Variance Structure	Heterogeneous Random Effect	Homogenous Random Effect	Heterogeneous Error Variance	Homogenous Error Variance	Homogenous Error Variance Random Effect	Heterogeneous Error Variance Random Effect
1	X		X		X	
2	**X**			**X**	**X**	
3		X	X		X	
4		X		X	X	
5	X		X			X
6	X			X		X
7		X	X			X
8		X		X		X

A) These models are described in the following table

Note: For the primary manuscript only models 2, 4, 6, 8 were fit (which correspond to models 1, 2, 3, and 4 respectively in Table 3) due to insufficient data to obtain accurate/precise estimates of treatment specific individual random effects for error variance. For our specific experimental design and sample size only models 2, 4, 6 and 8 could be reliably fit; however we provide code for all 8 models to provide guidance for other researchers interested in variance structure that has an appropriate design and sufficient sample size to model variance structure at this fine resolution. We include only a single example of the data, initial values, and model initiation arguments required due to their semi-conserved nature across all 8 models. A description of each

vector JAGS requires as well as each model parameter is also included. For all analyses convergence and sufficient mixing of Markov chains was obtained by reducing the potential scale reduction factor (R hat) below 1.01 for all parameters and using a total number of iterations greater than suggested by the raftery diagnostic (following Plummer et al., 2006).

B) JAGS models for the snail behavior "time at max height"

```
max_height_time_JAGS.data_model_1 <- list("y", "n", "indiv", "TTT",
      "cue", "day", "tank", "n.ids", "n.tank, n.ttt)

  # y       = data vector
  # n       = total sample size
  # indiv   = individual vector
  # TTT     = vector recording the treatment each individual is found in
  # cue     = cue vector
  # day     = day vector
  # tank    = tank vector
# n.ids     = number of individuals
  # n.tank  = number of tanks
# n.ttt     = number of treatments

max_height_time_JAGS.inits_model_1  <- function() {list(
                              alpha = rnorm(n.ids),
                              alpha.tank = rnorm(n.tank),
            alpha.e = rnorm(n.ids),
                              beta = rnorm(2),
mu.a = rnorm(n.ttt),
                              mu.e = rnorm(n.ttt),
sig.a = runif(n.ttt),
                              sigma.y = runif(n.ttt),
                              sig.tank = runif(1))}

  # alpha         = individual intercept
  # alpha.tank    = tank intercept
  # alpha.e       = individual error variance
  # beta          = fixed effects coefficient vector
  # mu.a          = treatment mean
  # mu.e          = treatment mean error variance
  # sig.a         = individual random effect
# sigma.y         = individual error variance random effect
  # sig.tank      = tank random effect

max_height_time_JAGS.param_model_1 <- c("sig.a", "sigma.y", "sig.tank",
      "mu.a", "mu.e", "beta")

max_height_time_JAGS.output_model_1 <-
            jags(max_height_time_JAGS.data_model_1,
                  max_height_time_JAGS.inits_model_1,
                  max_height_time_JAGS.param_model_1,
```

```
n.chains=6, n.iter=400000, n.burnin=40000,
                  n.thin=20, model.file = max_height_time_JAGS.model_1)


# model 1
  # Heterogeneous random effect + Heterogeneous error variance by
    # treatment +Heterogeneous error variance random effect


  max_height_time_JAGS.model_1 <- function() {
1
2     for(i in 1:n) {
3
4       y[i] ~ dnorm(y.hat[i], tau.y[i])
5
6       y.hat[i] <- alpha[indiv[i]] + alpha.tank[tank[i]] +
7             beta[1]*cue[i] + beta[2]*day[i]
8
9       tau.y[i] <- 1/pow(sigma[i], 2)
10
11      log(sigma[i]) <- alpha.e[indiv[i]]
12
13    }
14
15    for(k in 1:2) {
16      beta[k] ~ dnorm(0, 0.0001)
17    }
18
19    for(q in 1:n.tank) {
20      alpha.tank[q] ~ dnorm(0, tau.tank)
21    }
22
23    tau.tank <- pow(sig.tank, -2)
24      sig.tank ~ dunif(0, 100)
25
26    for(j in 1:n.ids) {
27      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a[TTT[j]])
28      alpha.e[j] ~ dnorm(mu.e[TTT[j]], tau.e[TTT[j]])
29    }
30
31    for(q in 1:n.ttt) {
32      tau.e[q] <- pow(sigma.y[q], -2)
33      sigma.y[q] ~ dunif(0, 100)
34      mu.e[q] ~ dnorm(0, 0.0001)
35      mu.a[q] ~ dnorm(0, 0.0001)
36      tau.a[q] <- pow(sig.a[q], -2)
37      sig.a[q] ~ dunif(0, 100)
38    }
39
40  }
```

In all subsequent models (2-8) lines 1-24 are conserved so we do not restate this part of the model. All model changes occur in the lines 27-28 (and with the priors connected to these

parameters on lines 32-37). For the remainder of the models we provide solely the lines that change from model to model.

```
# model 2
# Heterogeneous random effect + Heterogeneous error variance by
    # treatment + Homogeneous error variance random effect


    for(j in 1:n.ids) {
      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a[TTT[j]])
      alpha.e[j] ~ dnorm(mu.e[TTT[j]], tau.e)
    }

    for(q in 1:n.ttt) {
      mu.e[q] ~ dnorm(0, 0.0001)
      mu.a[q] ~ dnorm(0, 0.0001)
      tau.a[q] <- pow(sig.a[q], -2)
      sig.a[q] ~ dunif(0,100)
    }

    tau.e <- pow(sigma.y, -2)
    sigma.y ~ dunif(0, 100)


# model 3
# Heterogeneous random effect + Homogeneous error variance by
    # treatment + Heterogeneous error variance random effect


    for(j in 1:n.ids) {
      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a[TTT[j]])
      alpha.e[j] ~ dnorm(mu.e, tau.e[TTT[j]])
    }

    for(q in 1:n.ttt) {
      tau.e[q] <- pow(sigma.y[q], -2)
      sigma.y[q] ~ dunif(0, 100)
      mu.a[q] ~ dnorm(0, 0.0001)
      tau.a[q] <- pow(sig.a[q], -2)
      sig.a[q] ~ dunif(0,100)
    }

    mu.e ~ dnorm(0, 0.0001)


# model 4
 # Heterogeneous random effect + Homogeneous error variance by
   # treatment + Homogeneous error variance random effect

    for(j in 1:n.ids) {
      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a[TTT[j]])
      alpha.e[j] ~ dnorm(mu.e, tau.e)
    }

    for(q in 1:n.ttt) {
mu.a[q] ~ dnorm(0, 0.0001)
      tau.a[q] <- pow(sig.a[q], -2)
      sig.a[q] ~ dunif(0,100)
    }
```

```
    mu.e ~ dnorm(0, 0.0001)
    tau.e <- pow(sigma.y, -2)
    sigma.y ~ dunif(0, 100)


# model 5
# Homogeneous random effect + Heterogeneous error variance by
    # treatment + Heterogeneous error variance random effect

    for(j in 1:n.ids) {
      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a)
      alpha.e[j] ~ dnorm(mu.e[TTT[j]], tau.e[TTT[j]])
    }

    for(q in 1:n.ttt) {
      tau.e[q] <- pow(sigma.y[q], -2)
      sigma.y[q] ~ dunif(0, 100)
      mu.a[q] ~ dnorm(0, 0.0001)
  mu.e[q] ~ dnorm(0, 0.0001)
    }

tau.a <- pow(sig.a, -2)
    sig.a ~ dunif(0,100)


# model 6
# Homogeneous random effect + Heterogeneous error variance by
     # treatment + Homogeneous error variance random effect

    for(j in 1:n.ids) {
      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a)
      alpha.e[j] ~ dnorm(mu.e[TTT[j]], tau.e)
    }

    for(q in 1:n.ttt) {
mu.a[q] ~ dnorm(0, 0.0001)
mu.e[q] ~ dnorm(0, 0.0001)
}

    tau.e <- pow(sigma.y, -2)
    sigma.y ~ dunif(0, 100)
tau.a <- pow(sig.a, -2)
    sig.a ~ dunif(0,100)


# model 7
   # Homogeneous random effect + Homogeneous error variance by treatment
      # + Heterogeneous error variance random effect

    for(j in 1:n.ids) {
      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a)
      alpha.e[j] ~ dnorm(mu.e, tau.e[TTT[j]])
    }

    for(q in 1:n.ttt) {
      tau.e[q] <- pow(sigma.y[q], -2)
      sigma.y[q] ~ dunif(0, 100)
```

```
      mu.a[q] ~ dnorm(0, 0.0001)
}

    mu.e ~ dnorm(0, 0.0001)
tau.a <- pow(sig.a, -2)
    sig.a ~ dunif(0,100)


# model 8
   # Homogeneous random effect + Homogeneous error variance by treatment
      # + Homogeneous error variance random effect

    for(j in 1:n.ids) {
      alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a)
      alpha.e[j] ~ dnorm(mu.e, tau.e)
    }

    for(q in 1:n.ttt) {
mu.a[q] ~ dnorm(0, 0.0001)
}

mu.e ~ dnorm(0, 0.0001)
      tau.e <- pow(sigma.y, -2)
      sigma.y ~ dunif(0, 100)
      tau.a <- pow(sig.a, -2)
      sig.a ~ dunif(0,100)
```

C) JAGS model for the snail behavior "foraging observations"

```
foraging_JAGS.data <- list("y", "n", "n.ids", "ttt", "n.tank", "n.ttt",
      "TTT", "indiv", "tank", "cue", "day")

  # y       = data vector
  # n       = total sample size
  # indiv   = individual vector
  # TTT     = vector recording the treatment each individual is found in
  # ttt     = vector recording the treatment at each data point
  # cue     = cue vector
  # day     = day vector
  # tank    = tank vector
  # n.ids   = number of individuals
  # n.tank  = number of tanks
  # n.ttt   = number of treatments

foraging_JAGS.inits <- function() {list(alpha = rnorm(n.ids),
                              alpha.tank = rnorm(n.tank),
                                     sig.a = runif(n.ttt),
                                    disp = runif(n),
beta = rnorm(2),
                                     mu.a = rnorm(n.ttt),
 sigma.mu.overdisp = runif(n.ttt),
                                     sig.tank = runif(1),
                                     sigma.tau.overdisp = runif(1))}


  # alpha               = individual intercept
# alpha.tank            = tank intercept
  # sig.a               = individual random effect
  # disp                = additive overdispersion
  # beta                =fixed effects coefficient vector
# mu.a                  = treatment mean
  # sigma.mu.overdisp   = treatment unique Normally distributed additive
                              # overdispersion
  # sig.tank            = tank random effect
# sigma.tau.overdisp    = treatment unique Normally distributed
                              # individual overdispersion random effect


foraging_JAGS.parameters <- c("sigma.ttt", "mu.ttt",
      "sigma.mu.overdisp", "sigma.tau.overdisp", "sig.tank", "beta")

foraging_JAGS.output <- jags(foraging_JAGS.data, foraging_JAGS.inits,
foraging_JAGS.parameters,
n.chains=6, n.iter=400000,
                              n.burnin=40000, n.thin=20,
                              model.file = foraging_JAGS.model_1, refresh =
                              100)


foraging_JAGS.model_1 <- function() {
```

```
  for(i in 1:n) {

    y[i] ~ dpois(lambda[i])

    log(lambda[i]) <- alpha[indiv[i]] + alpha.tank[tank[i]] + disp[i]
          + beta[1]*day[i] + beta[2]*cue[i]

    disp[i] ~ dnorm(0, overdisp[i])

    overdisp[i] <- overdisp.mu[ttt[i]] + overdisp.tau[indiv[i]]

    }

  for(k in 1:2) {
    beta[k] ~ dnorm(0, 0.0001)
  }

  for (j in 1:n.ids) {
    alpha[j] ~ dnorm(mu.a[TTT[j]], tau.a[TTT[j]])
    overdisp.tau[j] ~ dnorm(0, overdisp.tau.indiv)
  }

  for(q in 1:n.tank) {
    alpha.tank[q] ~ dnorm(0, tau.tank)
  }

  tau.tank <- pow(sig.tank, -2)
  sig.tank ~ dunif(0, 100)

  for (k in 1:n.ttt) {
    mu.a[k] ~ dnorm(0, 0.0001)
    overdisp.mu[k] <- pow(sigma.mu.overdisp[k], -2)
    sigma.mu.overdisp[k] ~ dunif(0, 100)
    tau.a[k] <- pow(sig.a[k], -2)
    sig.a[k] ~ dunif(0, 100)
  }

  overdisp.tau.indiv <- pow(sigma.tau.overdisp, -2)
  sigma.tau.overdisp ~ dunif(0, 100)

}
```

4) Censored Snail:Days

As described in the primary manuscript, approximately 20% of climbing snails on all days (snail:days) were removed prior to analysis. This was conducted for reasons outlined in the primary manuscript. The following table describes patterns of censored snail:days across

Table A2.3: Censored Snail:Days. P-values were obtained from a Wald test using summary () in R. Bolded p-values indicate significance at $\alpha = 0.05$. In all cases the probability of being censored does not differ by treatment. In experiment 1 presence of cue decreases the probability of being censored and in experiment 2 both Day and Cue magnitude decrease the probability of being censored.

Experiment 1 (Coarse-Grained Variation)			**Experiment 2 (Fine-Grained Variation)**				
				Control Absent		**Control Present**	
Predictor	**Estimate**	**P-value**	**Predictor**	**Estimate**	**P-value**	**Estimate**	**P-value**
Low	-1.734	**$2.68*10^{-5}$**	Constant	-0.444	0.2577	-0.495	0.1799
High	-1.605	0.6931	High	-0.591	0.6463	-0.642	0.6483
Stochastic	-2.180	0.2463	Stochastic	-0.989	0.1266	-1.040	0.1299
			Control	-	-	-1.044	0.1506
Day	0.005	0.9062	Day	-0.078	**0.0322**	-0.070	**0.0308**
Cue	-0.750	**0.0123**	Cue	-82.19	**0.0412**	-81.47	**0.0311**

treatments, cue magnitudes and days.

As shown in the table, the probability of censoring (the snail reaching its max height at time 0 or at the last measured time point) decreases as cue magnitude increases and across days. Removing a larger proportion of snail climbing behavior on days without cue has two possible implications for data analysis: 1) Cue effect on climbing patterns is anti-conservative ; 2) Variance estimates are deflated, especially for time at max height, because times at the extremes have been removed. **However**, as our data censoringequally effects treatments, inference on the

significance of heterogeneous random effects or heterogeneous error variance by treatment is not expected to be impacted.

5) Summary of all ΔDIC values for all models

Table A2.4: Summary of all ΔDIC values for all models.

Experiment 1 (Coarse-Grained Environmental Variation)								
Variance Structure	1	2	3	4				
	Delta DIC							
Hiding Observations	**-15.56**	**0.00**	**-10.52**	**-8.18**				
Foraging Observations	**0.00**	**-4.01**	**-5.65**	**-14.79**				
Climbing Observations	**-6.07**	**-32.21**	**0.00**	**-35.37**				
Max Height	**-17.04**	**-21.23**	**0.00**	**-2.65**				
Time at Max Height	**-0.52**	**-5.26**	**0.00**	**-4.47**				
Time at First Response	**0.00**	**-2.32**	**-1.62**	**-4.09**				
Response Duration	**-3.90**	**-8.91**	**0.00**	**-4.91**				
Total Vertical Movement	**-15.00**	**-18.03**	**0.00**	**-2.41**				
Fixed Effects Model	1	2	3	4	5	6	7	8
	D+C+T	**D+C**	**D+T**	**C+T**	**D**	**C**	**T**	**-**
	Variance Model.Fixed Effects Model							
	Delta DIC							
Hiding Observations	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
	-6.15	**0.00**	**-8.51**	**-15.08**	**-3.70**	**-1.40**	**-3.20**	**-2.61**
Foraging Observations	4.1	4.2	4.3	4.4	4.6	4.5	4.7	4.8
	-15.82	**-13.73**	**0.00**	**-39.56**	**-17.40**	**-43.58**	**-29.47**	**-37.12**
Climbing Observations	4.1	4.2	4.3	4.4	4.6	4.5	4.7	4.8
	-40.21	**-37.25**	**-26.72**	**-37.56**	**0.00**	**-19.51**	**-5.51**	**-13.84**
Max Height	2.1	2.2	2.3	2.4	2.6	2.5	2.7	2.8
	-38.90	**-38.66**	**-0.36**	**-41.81**	**-0.71**	**-41.55**	**0.00**	**-1.84**
Time at Max Height	2.1	2.2	2.3	2.4	2.6	2.5	2.7	2.8
	-14.84	**-16.04**	**0.00**	**-16.77**	**-1.62**	**-17.80**	**-0.81**	**-2.23**
Time at First Response	4.1	4.2	4.3	4.4	4.6	4.5	4.7	4.8
	-2.82	**-6.69**	**0.00**	**-4.85**	**-4.11**	**-8.85**	**-1.81**	**-5.77**
Response Duration	2.1	2.2	2.3	2.4	2.6	2.5	2.7	2.8
	-24.96	**-23.64**	**-0.71**	**-21.17**	**0.00**	**-19.31**	**-1.68**	**-0.76**
Total Vertical Movement	2.1	2.2	2.3	2.4	2.6	2.5	2.7	2.8
	-74.97	**-68.84**	**-3.54**	**-76.84**	**-0.43**	**-71.15**	**-3.61**	**0.00**

Experiment 2 (Fine-Grained Environmental Variation) Control Absent								
Variance Structure	1	2	3	4				
	Delta DIC							
Hiding Observations	**-8.85**	**-7.37**	**-3.86**	**0.00**				
Foraging Observations	**-6.28**	**0.00**	**-2.78**	**-8.50**				
Climbing Observations	**-2.65**	**-15.06**	**0.00**	**-5.53**				
Max Height	**0.00**	**-6.93**	**-4.27**	**-11.03**				
Time at Max Height	**-15.95**	**-17.00**	**0.00**	**-1.63**				
Time at First Response	**-12.18**	**-16.44**	**0.00**	**-4.73**				
Response Duration	**-0.73**	**0.00**	**-3.91**	**-2.69**				
Total Vertical Movement	**0.00**	**-3.24**	**-2.69**	**-6.45**				
Fixed Effects Model	1	2	3	4	5	6	7	8
	D+C+T	**D+C**	**D+T**	**C+T**	**D**	**C**	**T**	**-**
	Variance Model.Fixed Effects Model							
	Delta DIC							
Hiding Observations	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
	-1.38	**-1.10**	**-6.55**	**-8.13**	**0.00**	**-7.83**	**-6.27**	**-5.66**
Foraging Observations	4.1	4.2	4.3	4.4	4.6	4.5	4.7	4.8
	-30.43	**-28.04**	**-40.43**	**-2.07**	**-26.71**	**0.00**	**-3.66**	**-7.67**
Climbing Observations	2.1	2.2	2.3	2.4	2.6	2.5	2.7	2.8
	-11.06	**-8.98**	**-11.13**	**-14.34**	**0.00**	**-0.33**	**-4.59**	**-8.04**
Max Height	4.1	4.2	4.3	4.4	4.6	4.5	4.7	4.8
	-3.61	**-3.47**	**-2.19**	**-2.29**	**-1.82**	**-1.74**	**-0.61**	**0.00**
Time at Max Height	2.1	2.2	2.3	2.4	2.6	2.5	2.7	2.8
	0.00	**-2.72**	**-1.25**	**-2.89**	**-3.74**	**-5.20**	**-3.10**	**-6.01**
Time at First Response	2.1	2.2	2.3	2.4	2.6	2.5	2.7	2.8
	0.00	**-1.37**	**-2.77**	**-0.83**	**-2.93**	**-3.23**	**-3.08**	**-4.01**
Response Duration	3.1	3.2	3.3	3.4	3.6	3.5	3.7	3.8
	0.00	**-3.70**	**-0.61**	**-1.22**	**-4.11**	**-5.41**	**-0.14**	**-5.71**
Total Vertical Movement	4.1	4.2	4.3	4.4	4.6	4.5	4.7	4.8
	-19.55	**-19.72**	**-15.32**	**-4.64**	**-15.12**	**-4.50**	**0.00**	**-0.56**

Table A2.4 Continued.

Experiment 2 Control Present								
Variance Structure	1	2	3	4				
				Delta DIC				
Hiding Observations	**-2.11**	**-4.10**	**-1.03**	**0.00**				
Foraging Observations	**-11.73**	**-22.30**	**0.00**	**-12.85**				
Climbing Observations	**-14.48**	**-15.18**	**0.00**	**-5.86**				
Max Height	**-7.55**	**-14.48**	**0.00**	**-7.83**				
Time at Max Height	**-22.28**	**-24.52**	**0.00**	**-8.55**				
Time at First Response	**-14.39**	**-19.61**	**0.00**	**-7.47**				
Response Duration	**-4.12**	**0.00**	**-0.73**	**-2.46**				
Total Vertical Movement	**-25.41**	**-12.79**	**-14.04**	**0.00**				
Fixed Effects Model	1	2	3	4	5	6	7	8
	D+C+T	**D+C**	**D+T**	**C+T**	**D**	**C**	**T**	**-**
	Variance Model.Fixed Effects Model							
	Delta DIC							
Hiding Observations	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
	0.00	**-4.37**	**-11.49**	**-22.59**	**-5.15**	**-10.26**	**-20.14**	**-8.48**
Foraging Observations	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
	-31.27	**-31.35**	**-24.04**	**-11.51**	**-9.73**	**-22.98**	**0.00**	**-7.63**
Climbing Observations	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
	-0.09	**0.00**	**-2.37**	**-19.99**	**-6.32**	**-4.13**	**-11.42**	**-7.60**
Max Height	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
	-5.39	**-5.47**	**-4.81**	**-3.65**	**-2.58**	**-3.14**	**-1.55**	**0.00**
Time at Max Height	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
	0.00	**-4.69**	**-0.34**	**-1.33**	**-4.57**	**-6.45**	**-2.63**	**-6.78**
Time at First Response	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8
	0.00	**-2.83**	**-0.27**	**-1.71**	**-3.71**	**-4.33**	**-1.86**	**-5.35**
Response Duration	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
	-6.65	**-1.87**	**-8.35**	**-2.91**	**-3.90**	**-0.05**	**-4.94**	**0.00**
Total Vertical Movement	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
	-22.94	**-26.32**	**-20.24**	**-7.90**	**-18.54**	**-10.98**	**-2.25**	**0.00**

Table A2.5: Effect sizes for the most parsimonious model for each response variable in both experiments. Effect sizes listed are ratios relative to the smallest estimate. The treatment with the largest effect size is highlighted. Table 3 provides a description of each model. Table S3 provides ΔDIC values for each model. For treatment homogenous parameter values effect sizes are all listed as *1.000*.

Experiment 1 (Coarse-Grained Environmental Variation)					**Experiment 2 (Fine-Grained Environmental Variation) Control Absent**				
		Parameter					**Parameter**		
	Treatment	**Treatment Mean**	**Individual Random Effect**	**Dispersion**		**Treatment**	**Treatment Mean**	**Individual Random Effect**	**Dispersion**
Hiding Observations	Low	1.000	1.000	1.440	Hiding Observations	Constant	1.280	3.230	1.055
Model 1.4	High	5.150	1.526	1.000	Model 1.4	High	1.000	1.000	1.139
	Stochastic	3.178	1.106	1.161		Stochastic	2.061	1.670	1.000
Foraging Observations	Low	*1.000*	*1.000*	*1.000*	Foraging Observations	Constant	1.000	*1.000*	*1.000*
Model 4.6	High	*1.000*	*1.000*	*1.000*	Model 4.3	High	1.236	*1.000*	*1.000*
	Stochastic	*1.000*	*1.000*	*1.000*		Stochastic	1.120	*1.000*	*1.000*
Climbing Observations	Low	1.242	*1.000*	*1.000*	Climbing Observations	Constant	1.313	1.938	*1.000*
Model 4.1	High	1.327	*1.000*	*1.000*	Model 2.4	High	1.132	1.000	*1.000*
	Stochastic	1.000	*1.000*	*1.000*		Stochastic	1.000	2.754	*1.000*
Max Height	Low	1.035	1.115	*1.000*	Max Height	Constant	1.060	1.000	*1.000*
Model 2.4	High	1.101	2.030	*1.000*	Model 4.1	High	1.217	1.000	*1.000*
	Stochastic	1.000	1.000	*1.000*		Stochastic	1.000	1.000	*1.000*
Time at Max Height	Low	*1.000*	1.436	*1.000*	Time at Max Height	Constant	*1.000*	1.040	*1.000*
Model 2.6	High	*1.000*	1.854	*1.000*	Model 2.8	High	*1.000*	1.000	*1.000*
	Stochastic	*1.000*	1.000	*1.000*		Stochastic	*1.000*	1.711	*1.000*
Time at First Response	Low	*1.000*	*1.000*	*1.000*	Time at First Response	Constant	*1.000*	1.588	*1.000*
Model 4.6	High	*1.000*	*1.000*	*1.000*	Model 2.8	High	*1.000*	1.000	*1.000*
	Stochastic	*1.000*	*1.000*	*1.000*		Stochastic	*1.000*	1.081	*1.000*
Response Duration	Low	1.128	1.318	*1.000*	Response Duration	Constant	*1.000*	*1.000*	1.214
Model 2.1	High	1.275	1.000	*1.000*	Model 3.8	High	*1.000*	*1.000*	1.000
	Stochastic	1.000	1.893	*1.000*		Stochastic	*1.000*	*1.000*	1.507
Total Vertical Movement	Low	1.000	1.296	*1.000*	Total Vertical Movement	Constant	1.154	*1.000*	*1.000*
Model 2.4	High	1.134	1.000	*1.000*	Model 2.4	High	1.122	*1.000*	*1.000*
	Stochastic	1.010	1.005	*1.000*		Stochastic	1.000	*1.000*	*1.000*

Table A2.5 Continued.

Experiment 2 (Fine-Grained Environmental Variation) Control Present				
		Parameter		
	Treatment	Treatment Mean	Individual Random Effect	Dispersion
Hiding Observations	Constant	1.067	3.438	*1.000*
Model 2.4	High	1.000	1.011	*1.000*
	Stochastic	1.154	1.810	*1.000*
	Control	1.522	1.000	*1.000*
Foraging Observations	Constant	1.000	5.386	*1.000*
Model 2.2	High	1.000	3.205	*1.000*
	Stochastic	1.000	4.265	*1.000*
	Control	1.000	1.000	*1.000*
Climbing Observations	Constant	1.579	2.407	*1.000*
Model 2.4	High	1.361	1.216	*1.000*
	Stochastic	1.207	3.425	*1.000*
	Control	1.000	1.000	*1.000*
Max Height	Constant	*1.000*	2.419	*1.000*
Model 2.2	High	*1.000*	2.035	*1.000*
	Stochastic	*1.000*	3.259	*1.000*
	Control	*1.000*	1.000	*1.000*
Time at Max Height	Constant	*1.000*	1.522	*1.000*
Model 2.8	High	*1.000*	1.448	*1.000*
	Stochastic	*1.000*	2.436	*1.000*
	Control	*1.000*	1.000	*1.000*
Time at First Response	Constant	*1.000*	2.118	*1.000*
Model 2.8	High	*1.000*	1.328	*1.000*
	Stochastic	*1.000*	1.500	*1.000*
	Control	1.000	1.000	*1.000*
Response Duration	Constant	1.150	1.000	1.464
Model 1.3	High	1.191	1.285	1.146
	Stochastic	1.197	1.224	1.851
	Control	1.000	1.825	1.000
Total Vertical Movement	Constant	*1.000*	4.742	1.075
Model 1.2	High	*1.000*	4.064	1.000
	Stochastic	*1.000*	4.395	1.061
	Control	*1.000*	1.000	1.138

6) Additional analysis for the "Coarse-Grained" variation experiment

In the primary manuscript we state that in the "Coarse-Grained" variation experiment, a high frequency of environmental switching (high variation treatment) resulted in the largest mean behavior for each behavior in which treatment unique means was supported. To determine if these responses were possibly due to cue from the previous day, we analyzed snail hiding observations, climbing observations, mean height, and total vertical movement in the three observations prior to cue addition using likelihood ratio tests between [G]LMMs with and without the treatment fixed effect. Mean height ($p = 0.1392$) and hiding behavior ($p = 0.4965$) did not differ by treatment; however, both climbing observations ($p = 0.049$) and total vertical movement ($p = 0.002$) did differ by treatment, with the high variation treatment having the largest estimate in both cases. While it is possible these results are due to lingering cue, a combination of patchy significance and support for no differences in behavioral responses between filtered cue water and tap water (unpublished results), we suggest that the small time window increases the saliency of the cue and results in more active individuals (see primary manuscript for further discussion).

7) Figure A2.1

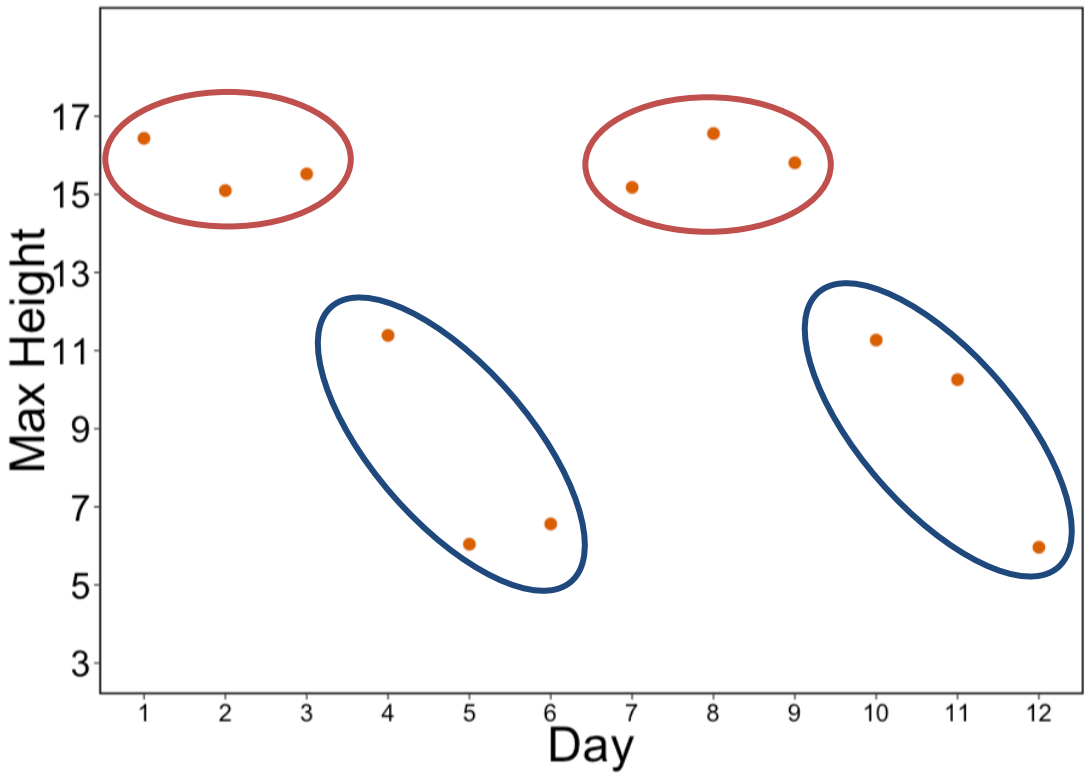

Figure A2.1: Example of larger anti-predator response behavior on non-cue days directly following cue days (days 4 and 10) than on non-cue days following other non-cue days (days 5, 6, 11, and 12). Example from *CGV* low treatment. Red ovals designate cue days and blue ovals designate non-cue days