Partitioning variance in a signaling trade-off under sexual selection reveals among-individual covariance in trait allocation

Michael S. Reichert[1*], Iván de la Hera[1], Maria Moiron[2]

1. Department of Integrative Biology, Oklahoma State University, Stillwater, OK, 74078, USA

2. Institute of Avian Research, An der Vogelwarte 21 26386, Wilhelmshaven, Germany

*Correspondence: michael.reichert@okstate.edu

ABSTRACT

Understanding the evolution of traits subject to trade-offs is challenging because phenotypes can (co)vary at both the among- and within-individual levels. Among-individual covariation indicates consistent, possibly genetic, differences in how individuals resolve the trade-off, while within-individual covariation indicates trait plasticity. There is also the potential for consistent among-individual differences in behavioral plasticity, although this has rarely been investigated. We studied the sources of (co)variance in two characteristics of an acoustic advertisement signal that trade off with one another and are under sexual selection in the gray treefrog, *Hyla chrysoscelis*: call duration and call rate. We recorded males on multiple nights calling spontaneously and in response to playbacks simulating different competition levels. Call duration, call rate, and their product, call effort, were all repeatable within nights, but only call duration was repeatable across nights. Call duration and call rate covaried negatively, and the largest covariance was at the among-individual level. There was extensive plasticity in calling both within and across nights, but no evidence that the plastic response to competition was repeatable. The significant negative among-individual covariance in trait values is perpendicular to the

primary direction of sexual selection in this species, indicating potential limits on the response to selection.

Key words: random regression, mixed effects model, reaction norm, individual variation, plasticity, repeatability

INTRODUCTION

A fundamental tenet of life history theory is that variation in the allocation of resources to different phenotypic characteristics is subject to trade-offs, resulting in a negative covariance between traits (Stearns 1989; Agrawal et al. 2010; Garland 2014). Understanding how individuals resolve these trade-offs is key to understanding strategies related to important fitness-related traits such as mating, foraging decisions, growth, and parental care (Giraldeau and Caraco 2000; Badyaev and Ghalambor 2001; Simmons and Emlen 2006; Evans 2010). For trade-offs involving behavioral traits, the short duration and repeated expression of behaviors means that there are two fundamental levels at which trade-offs can be expressed. First, trade-offs can be expressed at the among-individual level, in which case individuals differ consistently in the extent to which they favor one trait over the other. Second, trade-offs can be expressed at the within-individual level, where individuals adjust the expression of each trait from one instance of the behavior to the next. These sources of variation are not mutually exclusive and often there are both among- and within-individual contributions to trait variation. Nevertheless, it is important to quantify the relative contribution of among- and within-individual effects on trait covariance because these effects do not necessarily operate in the same direction (Dingemanse and Dochtermann 2013; Moiron et al. 2016; Berberi and Careau 2019), thus potentially leading to misinterpretations of patterns of covariance measured at only one level (van Noordwijk and de Jong 1986; Reznick et al. 2000; Careau and Wilson 2017). Furthermore, these sources of variation are not necessarily driven by the same evolutionary processes. Among-individual covariance potentially indicates underlying heritable covariance that could respond to or constrain correlated selection on those traits (Price and Langen 1992; Sinervo and Svensson 2002). Within-individual covariance indicates phenotypic plasticity and may reflect short-term responses to changes in the costs and benefits of different levels of trait variation (Baugh et al. 2014; Katz and Naug 2015). Although trade-offs are

recognized as being critical forces in trait evolution and organismal performance, there is still mixed empirical evidence for their actual evolutionary consequences in natural systems, in part because the sources of trait covariance are rarely quantified (Roff and Fairbairn 2007; Laskowski et al. 2021).

Even for traits that vary at the within-individual level via plastic responses to environmental conditions, there may nevertheless be among-individual variance in plasticity itself (Nussey et al. 2007; Dingemanse et al. 2010; Dingemanse and Wolf 2013; Stamps 2016; O'Dea et al. 2022). In other words, some individuals may consistently make a large change in the trait expression in response to changing environmental conditions, while other individuals consistently make smaller changes or none at all (Brommer et al. 2005; Nussey et al. 2005; Araya-Ajoy and Dingemanse 2017). Variation in individual plasticity could reflect underlying differences in state variables such as metabolic scope or condition (Biro and Stamps 2010; Biro et al. 2018; Dubois 2019), responsiveness to environmental stimuli (Wolf et al. 2011), personality differences (Highcock and Carter 2014), or differences in individual strategies for balancing the costs and benefits of plasticity in general (Snell-Rood 2013). If the costs and benefits of different levels of trait expression vary with environmental conditions, then these consistent among-individual differences in plasticity will also affect individual fitness, and could result in a response to selection if there is an underlying genetic component (Nussey et al. 2007).

Sexual signaling in Cope's gray treefrog, *Hyla chrysoscelis*, is an excellent example of these complex sources of trait variation and covariation in the face of a strong trade off. Male *H. chrysoscelis* produce energetically costly advertisement calls to attract females. Females are especially attracted to calls with higher total acoustic energy ("call effort") (Tanner et al. 2017). Call effort is allocated into two distinct traits that trade off: the duration of individual calls, and the rate at which calls are produced (Ward et al. 2013). Males tend to increase call durations and decrease call rates in response to escalating

78 competition (Ward et al. 2013). When call efforts are held constant, there is some evidence that female
79 *H. chrysoscelis* prefer longer calls with lower rates to shorter calls at a higher rate (Gerhardt et al. 1996),
80 although another study found no such preference (Ward et al. 2013).
81
82 The variance and covariance components of the characteristics involved in this signaling trade-off have
83 not been quantified in *H. chrysoscelis*, and indeed some of these have received little attention in
84 sexually selected traits in any species. First, the extent to which there are consistent among-individual
85 differences (i.e. the repeatability) in each call characteristic has not been measured. In the closely
86 related Eastern gray treefrog, *Hyla versicolor*, both call duration and call rate have been shown to be
87 significantly repeatable, although the repeatability of call rate depended on year in some studies
88 (Sullivan and Hinshaw 1992; Runkle et al. 1994; Gerhardt et al. 1996). Second, while there is evidence
89 for within-individual covariance in call duration and call rate (Ward et al. 2013), it is unknown whether
90 there is also a significant among-individual covariance component. This is likely to be the case because in
91 *H. versicolor*, there was significant covariance between call duration and call period (the inverse of call
92 rate) at both the phenotypic and genetic level (Welch et al. 2014), the latter indicating an among-
93 individual component to the covariance. Third, while call duration and call rate are both plastic, it is
94 unknown whether there is consistent among-individual variation in the magnitude of this plasticity in
95 response to competition.
96
97 In this study, we investigated the role that different sources of (co)variance play in a performance trade-
98 off that is under sexual selection by making repeated recordings in the field of male *H. chrysoscelis* both
99 calling spontaneously and in response to playbacks simulating two different levels of competition, to
100 address the following questions: 1. Are male call characteristics repeatable? And if so, are repeatabilities
101 context-dependent? 2. Is there repeatable plasticity in male calling in response to a change in acoustic

competition? 3. What are the relative contributions of within- and among-individual variation to the expected negative covariance between call duration and call rate?

METHODS

We studied a population of *H. chrysoscelis* at a pond on the McPherson Preserve, Payne County, Oklahoma, USA (36.100575, -97.204080). Recordings were made during the breeding season (May – July) in 2019 and 2020, from 21:00-24:00. We searched for calling males as test subjects and removed any additional calling individual within 2 m to reduce the effects of intensive competition during the playback recordings. Each male was recorded using the procedure described below. Male body temperature was measured at the beginning of the recording session using an infrared thermometer (Fluke 62 Max+) pointed at the body of the frog. Males were then captured and marked with a unique visible implant elastomer for individual identification before being released the same night back to the study pond. All procedures were carried out with the approval of the Institutional Animal Care and Use Committee at Oklahoma State University (protocol number AS-19-4) and with permits from the Oklahoma Department of Wildlife Conservation.

Playback stimuli

We generated synthetic *H. chrysoscelis* calls with a custom script in R version 3.6.0 software (R Development Core Team 2021) using the seewave version 2.1.3 (Sueur et al. 2008) and tuneR version 1.3.3 (Ligges et al. 2013) packages. Call stimuli were designed to have the characteristics of a typical male *H. chrysoscelis* call, with temporal parameters varied to match the temperature of the test subject and to simulate two different levels of competition, average or high (henceforth, ‘competitive’). The call

consisted of a repeated series of pulses (pulse duty cycle = 50%) with exponential rise and fall (each 50% of pulse duration). Each pulse contained two frequency bands, a low frequency peak at 1.1 kHz and a high frequency peak of 2.2 kHz that was 6 dB louder than the low frequency peak. Call pulse rates vary with temperature (Gerhardt 2005), so we generated stimuli with pulse rates that corresponded to one of four different temperatures (36.4, 49.4, 62.5, and 75.5 pulses/s at 15, 20, 25 and 30 °C, respectively). The 'average' stimulus had 52 pulses (linear call rise over 50 ms), and was repeated every 6 seconds. The 'competitive' stimulus had call characteristics of a male engaged in an especially competitive interaction (Reichert and Gerhardt 2012), with both increased number of pulses (70) and a more rapid repetition rate (one call every 4 s). Otherwise, the two stimuli were identical. Stimuli were repeated to make a playback track of 3 minutes total duration, and saved as .Wav files (44.1 kHz sampling rate).

Playback procedure

A schematic of the experimental design is shown in Figure S1. For each male on each night of recording, there were four consecutive recording periods corresponding to different "social contexts". 1. Prior to playbacks we made a baseline recording of approximately 20 calls given spontaneously by the subject male, using digital audio recorders (Marantz PMD-660, Marantz PMD-661 or Tascam dr-100mkiii; 44.1 kHz sampling rate and 16-bit resolution in all cases) and directional microphones (Sennheiser ME-67). We measured the male's body temperature and chose the stimulus with the pulse rate most closely corresponding to that temperature for playback. 2. We then began playback of the average stimulus and recorded the male's calling responses for the three-minute duration of the playback. 3. We then immediately began playback of the competitive stimulus, again recording the male's responses for three minutes. Both playbacks were broadcast as .Wav files with an MP3 player (FiiO M3K) through a Pignose 7-100 portable speaker, at a sound pressure level of 85 dB (RMS, re: 20 µPa) at 1 m., calibrated at the

beginning of the evening with a sound pressure level meter (Extech 407703A). We always played the two stimuli in the same order to simulate escalating competition. 4. Following the two playbacks we recorded another 20 spontaneous calls as a post-playback baseline. Together, the responses to these four playbacks comprised a 'series', which referred to the set of calls given by an individual on a single night and was used as a random effect in statistical analyses (see below). We returned to the pond on multiple nights and continued to perform playbacks using the procedure described above. This resulted in many males being recorded multiple times on different nights in response to the same playback stimuli.

Acoustic analyses

All advertisement calls given by males during each recording period were analyzed in Raven Pro 1.6 software (Cornell Lab of Ornithology). We measured three characteristics for each call. Call duration was measured as the difference between the onset and offset of the call. Call rate was measured on a call-by-call basis (final call of the recording excluded) as the reciprocal of the amount of time between the onset of a call and the onset of the next call. Call effort was calculated (again, excepting the final call) as the call duration multiplied by the call rate.

Statistical analyses

Statistical model implementation

All statistical models described below were fitted using a Bayesian framework implemented in the statistical software R (v. 3.6.1, R Core Team 2019) using the R-package *MCMCglmm* (Hadfield 2010). For

all models, we used parameter-expanded priors (Hadfield 2010). The number of iterations (1300000) and thinning interval were chosen for each model to ensure that the minimum MCMC effective sample sizes for all parameters were 1000. Burn-in was set to 300000 iterations. The retained effective sample sizes yielded absolute autocorrelation values lower than 0.1 and satisfied convergence criteria based on the Heidelberger and Welch convergence diagnostic. We drew inferences from the posterior modes and 95% Credible Intervals (CI). We considered fixed effects and correlations to be significant if the 95% CI did not include zero; estimates centered on zero were considered to provide support for the absence of an effect. For all models, we used a Gaussian error distribution, therefore call rate and call effort were transformed with a natural log+1 transformation prior to analyses to meet the assumption of a normal distribution.

All models had a common set of fixed and random effects. These fixed effects were year (categorical variable, two levels: 2019 and 2020), time of night (continuous variable, in hours past 21:00), calendar day (continuous variable), and male body temperature (continuous variable). Both time of night and calendar day were centered, by subtracting the mean, and standardized, by dividing by the standard deviation. Additionally, for each recording period (i.e. calls from a single social context on one night), we fitted a “sequence” effect, a continuous variable numbering the calls in order on the recording, to control for potential autocorrelation among the repeated observations. For analyses that included data from recordings in more than one social context, we included social context (categorical variable, four levels: baseline, average playback, competitive playback, post-playback) as a fixed effect. Male identity was included as a random effect. We included an additional random effect called “series”, a composite variable of “male identity” and “calendar day”, which is used in the calculation of short-term repeatability to control for repeated measures of an individual’s calls on the same night (Araya-Ajoy et

al. 2015). The additional effects included in specific models, and the justification for doing so, are described below for those models.

Repeatability of call characteristics

We performed a series of univariate general linear mixed-effects models to obtain estimates of the repeatability (R) of each call characteristic, separately for each social context. We fitted the call characteristic of interest as the dependent variable. By fitting both male identity and series as random effects in our statistical model, we were able to estimate "long-term" repeatability based on variation across different nights (i.e., variance explained by individual identity divided by the total variance) and "short-term" repeatability based on variation within a single night (i.e., variance explained by individual identity and series divided by the total) (Araya-Ajoy et al. 2015). As such, we partitioned the phenotypic variance of each call characteristic into three components: among-males, within-male across-nights, and within-male within a night. For all repeatability analyses, we included data from all males, i.e., also for those males that were only recorded once, for improved statistical power (Martin et al. 2011).

Besides calculating the repeatability of each call characteristic in each social context, we also tested whether call characteristics were repeatable across the social contexts by calculating a single repeatability value for each call characteristic across all contexts. To do so, we fitted the call characteristic of interest as the dependent variable and social context as an additional fixed effect. In this case, repeatability of each call characteristic, conditional to the variance explained by the fixed effects, was estimated as the proportion of the total phenotypic variance explained by individual variance.

221 *Individual differences in call characteristic plasticity*

222

223 To test whether there were differences among individuals in their plastic response to social context
224 (individual by environment interactions, IxE), we used a random regression analysis (Nussey et al. 2007).
225 We used separate models to examine specifically how call characteristics changed with increasing
226 competition from the initial baseline to the average competition playback, and then how they changed
227 from the average competition to the high competition playback. We fitted a series of univariate general
228 linear mixed-effects models with the call characteristic of interest (call duration, call rate or call effort)
229 as the response variable. We included social context as an additional fixed effect (two levels), and the
230 interactions of "individual male identity" and "series" with social context as two additional random
231 slope effects. With these models, we tested for among-individual and within-individual among-night
232 variance in the intercepts ($V_{intercepts}$) and slopes in response to social context (V_{slopes}), while also testing
233 for the covariance and correlation between intercepts and slopes ($COV_{intercepts\text{-}slopes}$ and $r_{intercepts\text{-}slopes}$,
234 respectively). A correlation between intercepts and slopes would indicate that the value of an
235 individual's call characteristics in one context predict how it changes those characteristics in another
236 context. Moreover, by applying this approach, we were able to estimate "intercept" and "slope"
237 repeatabilities (*sensu* Araya-Ajoy et al. 2015). Intercept repeatability indicates consistent differences
238 among individuals in average calling across all contexts, while slope repeatability indicates consistent
239 differences among individuals in the extent to which they changed their calls in response to changes in
240 social context (i.e., the repeatability of individual plasticity).

241

242 Additionally, for those models where random regression analyses suggested evidence of among-male
243 differences in the plastic response between contexts, we used a "character-state" approach to confirm
244 the results (Brommer 2013b; Roff and Wilson 2014). The only call characteristic where this was the case

was call duration (see Results). Therefore, we fitted two separate bivariate general linear mixed-effects models: one in which we simultaneously modeled call duration in the baseline and in the average playback as response variables, and one in which we modeled call duration in the average playback and in the competitive playback. We fitted unstructured (co)variance matrices for the random effects of individual identity and series, and did not model the covariance at the residual variance because there can be no cross-context covariance for residual components. For this model structure in which trait expression is compared between two environments, an IxE can be inferred when the 95% CI of the correlation coefficient between the traits does not encompass +1 (Houslay and Wilson 2017).

Sources of (co)variance in the trade-off between call duration and call rate

To determine the level of covariance between call duration and call rate, we fitted a bivariate general linear mixed-effects model and investigated whether and how these two call characteristics covaried at the population (phenotypic), among-individual, within-individual among-night and within-individual within-night levels. To do so, we simultaneously modelled call duration and call rate as response variables, and fitted unstructured (co)variance matrices for the random effects of individual identity, series and residual variance. We included data from all recording periods and therefore fitted social context as an additional fixed effect.

RESULTS

The dataset includes 38 males recorded on more than one night (2 nights: N=24, 3 nights: N=5, 4 nights: N = 6, 5 nights: N=3). In addition, data from 77 males that were recorded on only one night are included.

Repeatability of call characteristics

All male call characteristics were significantly repeatable in the long- (among nights) and short-term (within nights), except that there was no significant long-term repeatability for call rate and call effort (Figure 1). The magnitude of the short-term repeatability was always greater than the long-term repeatability (Figure 1). Social context did not affect the repeatability coefficients, which were generally very similar across all four contexts (Table S1).

When we calculated a single repeatability value for each call characteristic across all social contexts, we found that call duration (posterior mode: 0.40, 95% Credible Interval: [0.298, 0.538]), but not call rate (0.05, [0, 0.126]) or call effort (0, [0, 0.044]), were significantly repeatable. The magnitudes of the repeatability coefficients when pooling together data from the four contexts were approximately the same as the long-term repeatability (Figure 1).

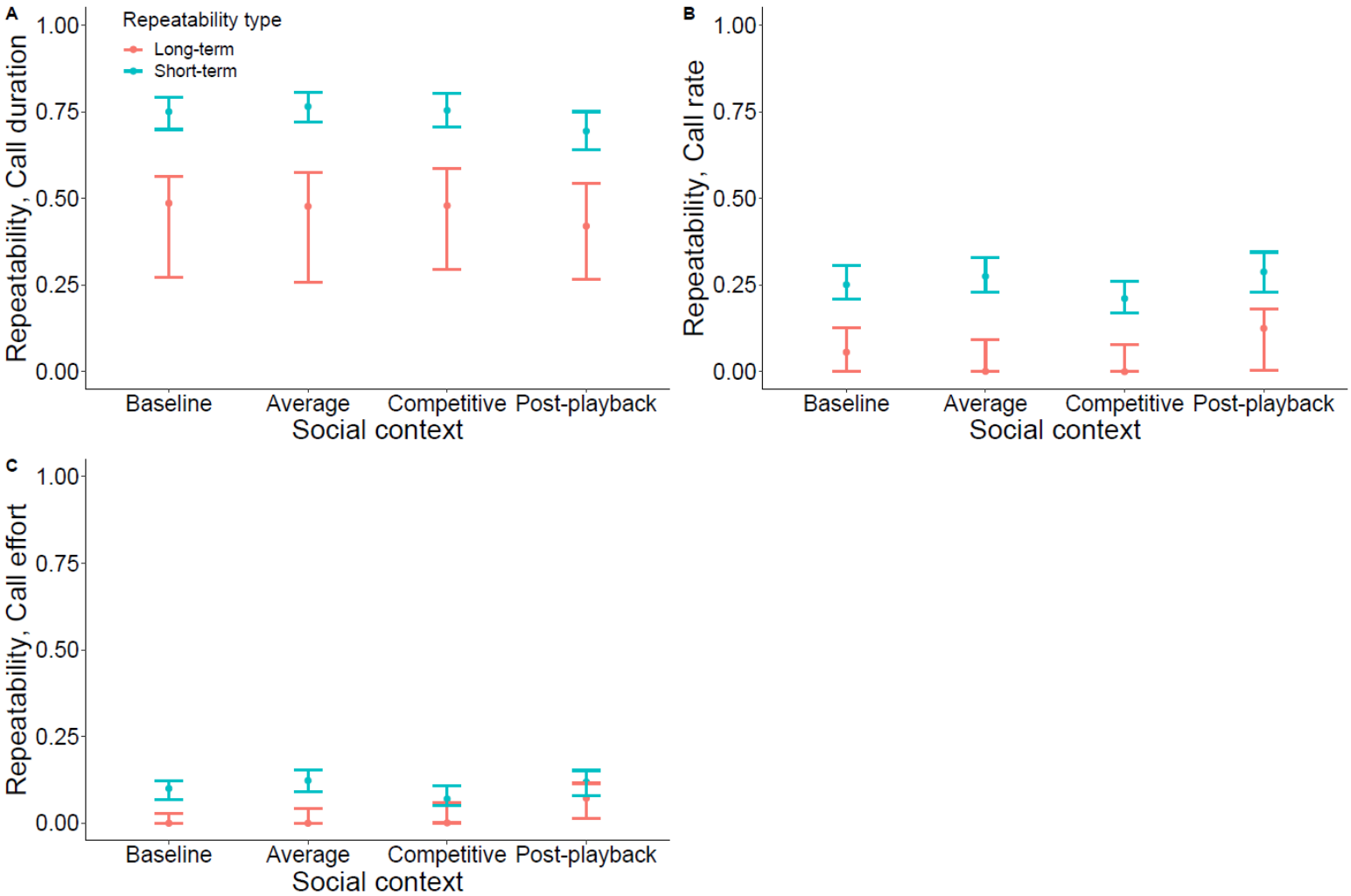

Figure 1. Long- and short-term repeatability of call characteristics in each recording period (social context). Repeatability estimates are shown as posterior modes with 95% Credible Intervals. Long-term repeatability refers to the repeatability of call characteristics among series (i.e.

288 recordings made across multiple nights). Short-term repeatability refers to the repeatability of call characteristics within series (i.e., within the

289 same night).

290

Individual differences in call characteristic plasticity

At the among-individual level, we did not find clear evidence for differences in the plastic responses to any of the two changes in context (i.e., neither from baseline to average playback nor from average playback to competitive playback) for any of the call characteristics, as indicated by the very low to zero among-individual variance in individual slopes in all cases (Tables 1-3). These results indicate that there was no significant variation among individuals for the extent to which they changed call duration, call rate or call effort as social competition increased. There was also no evidence for a correlation between intercepts and slopes of the reaction norm for any call characteristic at the among-individual level (Tables 1-3).

At the within-male among-night level, we did find some evidence for variation in plasticity in call duration for both the baseline to the average playback and the average playback to the competitive playback (Table 1). In other words, there was variation in how the same male tested on different nights plastically changed its call characteristics in response to the same playback stimuli. Additionally, there was a significant negative intercept-slope correlation for call rate in both contexts, for call duration from baseline to the average playback and for call effort from the average to the competitive playback (Figure 2; Tables 1-3). This indicates a fanning-in pattern, where a male that initially called with low values of these call characteristics would increase them more in response to playback than when that same male initially called with high values of these call characteristics.

As expected, the intercept repeatabilities, which measure the repeatability of the mean value of call characteristics across the two contexts examined, resembled the repeatability estimates using data from all of the social contexts (Tables 1-3). We found that the slope repeatability of call duration had a

315 relatively large coefficient (~39%), but was not significant, for the change from calling in response to the
316 average playback to the response to the competitive playback (Table 1). Slope repeatability was not
317 significant for call rate or call effort (Tables 2-3). The results of the character-state approach
318 corroborated the random regression model results: there is evidence for within-male among-night
319 variation in plasticity for call duration, but no clear evidence for among-individual variation (i.e., no I x E)
320 (Table 4).
321
322 Fixed effects for the random regression models are given in Tables S2-S4, and for the character state
323 models in Tables S5-S6. There was an effect of context on call duration and call rate, which reflected the
324 overall negative covariance of these traits: with increasing competition, males had a higher call duration
325 but a lower call rate (Table S2, S3). At higher temperatures, males had lower call durations, likely due to
326 increased pulse repetition rates (Gayou 1984), and called at a faster rate. Calendar day also had an
327 effect on call effort in both models and call rate in the model with the average and competitive
328 contexts: males had lower call efforts and call rates later in the season.
329

Table 1. Sources of phenotypic variance in call duration plasticity.

	Baseline vs. average		Average vs. competitive	
Random effects	mode	95% CI	mode	95% CI
Among-individual variance				
Intercept	0.006	[0.004, 0.01]	0.007	[0.004, 0.011]
Slope	0	[0, 0.001]	0.001	[0, 0.002]
Intercept–slope covariance	0	[0, 0.001]	0.001	[-0.001, 0.002]
Intercept–slope correlation	0.376	[-0.482, 0.994]	0.412	[-0.223, 0.853]
Within-individual among-series variance				
Intercept	0.004	[0.003, 0.006]	0.005	[0.003, 0.008]
Slope	0.003	[0.002, 0.004]	0.002	[0.001, 0.003]
Intercept–slope covariance	-0.001	[-0.002, 0]	0	[-0.001, 0.001]
Intercept–slope correlation	-0.221	[-0.465, -0.015]	0.013	[-0.354, 0.238]
Residuals	0.004	[0.004, 0.004]	0.005	[0.005, 0.005]
Intercept repeatability	0.581	[0.435, 0.755]	0.599	[0.373, 0.733]
Slope repeatability	0.001	[0, 0.2]	0.394	[0, 0.61]

Posterior mode and 95% Credible Intervals for each (co)variance component are shown for the among-individual, within-individual among-series (i.e., among-night), and residual levels. In addition, we give the repeatability coefficient for the intercept and slope of the reaction norm (Araya-Ajoy et al. 2015).

Separate models were fitted for the contexts of baseline versus average playback, and average versus competitive playback.

339 Table 2. Sources of phenotypic variance in call rate plasticity.

340

	Baseline vs. average		Average vs. competitive	
Random effects	mode	95% CI	mode	95% CI
Among-individual variance				
Intercept	0	[0, 0.001]	0	[0, 0]
Slope	0	[0, 0]	0	[0, 0]
Intercept–slope covariance	0	[0, 0]	0	[0, 0]
Intercept–slope correlation	-0.944	[-0.999, 0.824]	-0.929	[-0.998, 0.742]
Within-individual among-series variance				
Intercept	0.001	[0.001, 0.001]	0.001	[0.001, 0.002]
Slope	0.001	[0, 0.001]	0	[0, 0]
Intercept–slope covariance	0	[0, 0]	0	[-0.001, 0]
Intercept–slope correlation	-0.266	[-0.478, -0.007]	-0.7	[-0.864, -0.451]
Residuals	0.004	[0.003, 0.004]	0.003	[0.003, 0.003]
Intercept repeatability	0.212	[0, 0.432]	0.001	[0, 0.276]
Slope repeatability	0.001	[0, 0.234]	0.004	[0, 0.629]

341

342 Values as in Table 1. Note that call rate was ln(x+1) transformed prior to analyses.

343

Table 3. Sources of phenotypic variance in call effort plasticity.

	Baseline vs. average		Average vs. competitive	
Random effects	mode	95% CI	mode	95% CI
Among-individual variance				
Intercept	0	[0, 0]	0	[0, 0]
Slope	0	[0, 0]	0	[0, 0]
Intercept–slope covariance	0	[0, 0]	0	[0, 0]
Intercept–slope correlation	-0.325	[-0.873, 0.992]	-0.004	[-1, 0.854]
Within-individual among-series variance				
Intercept	0	[0, 0]	0	[0, 0]
Slope	0	[0, 0]	0	[0, 0]
Intercept–slope covariance	0	[0, 0]	0	[0, 0]
Intercept–slope correlation	-0.136	[-0.374, 0.541]	-0.926	[-0.999, -0.366]
Residuals	0.002	[0.002, 0.002]	0.002	[0.002, 0.002]
Intercept repeatability	0.001	[0, 0.307]	0.002	[0, 0.358]
Slope repeatability	0.004	[0, 0.802]	0.005	[0, 0.887]

Values as in Table 1. Note that call effort was ln(x+1) transformed prior to analyses.

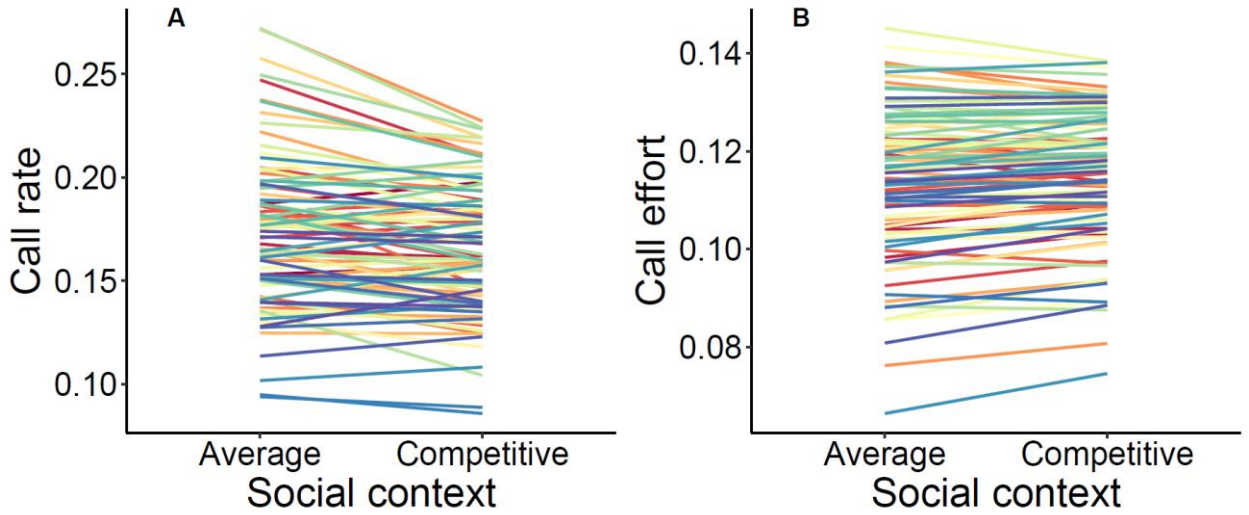

Figure 2. Illustration of intercept-slope correlations for (A) call rate and (B) call effort. Each line represents the mean best linear unbiased predictor of that call characteristic (fitted from the statistical models in Tables 2 and 3) for one male on one night for calling in response to the average playback and in response to the competitive playback. Note that both call rate and call effort were ln(x+1) transformed prior to analyses. Only data from males that were recorded on at least two nights are shown; line colors indicate individual males. For both characteristics there is a clear fanning in pattern, corresponding to the negative intercept-slope correlation at the within-individual among-nights level.

Table 4. Results of the “character state” model of plasticity in call duration.

	Baseline vs. average			Average vs. competitive	
Random effects	mode	95% CI	Random effects	mode	95% CI
$V_{\text{among-individual average}}$	0.008	[0.005 , 0.012]	$V_{\text{among-individual competitive}}$	0.01	[0.006 , 0.014]
$V_{\text{among-individual baseline}}$	0.006	[0.003 , 0.009]	$V_{\text{among-individual average}}$	0.008	[0.004 , 0.011]
$COV_{\text{among-individual}}$	0.007	[0.004 , 0.01]	$COV_{\text{among-individual}}$	0.008	[0.005 , 0.012]
$r_{\text{among-inidvidual}}$	0.999	[0.932 , 1]	$r_{\text{among-inidvidual}}$	0.978	[0.913 , 1]
$V_{\text{within-individual among-series average}}$	0.005	[0.003 , 0.008]	$V_{\text{within-individual among-series competitive}}$	0.006	[0.004 , 0.01]
$V_{\text{within-individual among-series baseline}}$	0.004	[0.003 , 0.006]	$V_{\text{within-individual among-series average}}$	0.006	[0.004 , 0.008]
$COV_{\text{within-individual among-series}}$	0.003	[0.002 , 0.006]	$COV_{\text{within-individual among-series}}$	0.005	[0.003 , 0.008]
$r_{\text{within-individual among-series}}$	0.745	[0.619 , 0.841]	$r_{\text{within-individual among-series}}$	0.845	[0.769 , 0.924]
$V_{\text{within-individual within-series average}}$	0.004	[0.004 , 0.004]	$V_{\text{within-individual within-series competitive}}$	0.006	[0.005 , 0.006]
$V_{\text{within-individual within-series baseline}}$	0.004	[0.004 , 0.004]	$V_{\text{within-individual within-series average}}$	0.004	[0.004 , 0.004]

Posterior mode and 95% Credible Intervals for the variance (V) and covariance (COV) components, along with the correlation coefficient (*r*), are shown for three levels of variation: among-individual, within-individual among-series (i.e. among night), and within-individual within-series (i.e. within night). Separate models were run for call duration for the baseline and average contexts, and for the average and competitive contexts.

Sources of (co)variance in the trade-off between call duration and call rate

There was a strong trade-off between call duration and call rate; longer duration calls were given at lower rates at the phenotypic level (Table 5). When we tested whether this phenotypic negative correlation was also present at other hierarchical levels of variation, we observed that there was a significant negative covariance between call duration and call rate at all three levels: among-males, within-males among-night, and within-males within-night (residual variance) (Table 5). The negative correlation among individuals indicates that some males consistently resolved the trade-off by producing longer-duration calls at a lower rate than others. The negative within-individual among-night correlation means that males that had a high call duration on a given night also tended to have a low call rate. Finally, the negative within-individual within-night correlation suggests that on a call-to-call basis, when a male had a high call duration it also had a low call rate. Across the four levels of correlations, the among-individual correlation was the strongest and the within-individual within-night correlation was the weakest (Table 5). Fixed effects on the covariance are shown in Table S7.

Table 5. Models of covariance between call duration and call rate

Random effects	Posterior mode	95% CI
$V_{\text{among-individual CD}}$	0.006	[0.004, 0.01]
$V_{\text{among-individual CR}}$	0	[0, 0.001]
$COV_{\text{among-individual CD-CR}}$	-0.001	[-0.002, -0.001]
$r_{\text{among-individual CD-CR}}$	-0.902	[-0.996, -0.674]
$V_{\text{within-individual among-series CD}}$	0.004	[0.003, 0.006]
$V_{\text{within-individual among-series CR}}$	0.001	[0.001, 0.001]
$COV_{\text{within-individual among-series CD-CR}}$	-0.001	[-0.002, -0.001]
$r_{\text{within-individual among-series CD-CR}}$	-0.648	[-0.78, -0.482]
$V_{\text{within-individual within-series CD}}$	0.006	[0.006, 0.006]
$V_{\text{within-individual within-series CR}}$	0.004	[0.003, 0.004]
$COV_{\text{within-individual within-series CD-CR}}$	0	[0, 0]
$r_{\text{within-individual within-series CD-CR}}$	-0.038	[-0.054, -0.022]
Phenotypic (population-level) *r*	-0.283	[-0.344, -0.234]

Random effect results from a bivariate mixed-effects model of the correlations between call duration and call rate at the phenotypic, among-males, within-males among-series (i.e. among night), and within-males within-series (i.e. within night) levels. Estimates represent the posterior modes for the variance (V), covariance (COV) and correlation (*r*) estimates, with associated 95% Credible Intervals (CI). Note that call rate was ln(x+1) transformed prior to analyses.

DISCUSSION

Variation in call characteristics in *H. chrysoscelis* was due to a combination of phenotypic plasticity in response to competition, and repeatable among-individual differences. Furthermore, for two key temporal call characteristics, call duration and call rate, there was significant negative covariation, confirming the expected trade-off between these two characteristics at the phenotypic level (Ward et al. 2013). Because we had multiple recordings per individual, we could partition the covariance, and we found significant covariation between call duration and call rate at both within- and among-individual levels. Thus, males generally adjusted their calls in accordance with the trade-off between duration and rate in response to competition, exhibiting short-term behavioral plasticity, but some males nevertheless consistently allocated more calling energy to call duration than others. However, there was no evidence for consistent among-individual differences in the plastic response to competition. We discuss below the implications of these results for the understanding of sexual selection and of the evolution of multivariate phenotypes subject to trade-offs.

Implications for sexual selection

Call duration was highly repeatable across nights and across social contexts representing different levels of competition. This indicates that even though males vary their call duration in response to competition, males that produce the longest duration calls in low levels of competition are the same males that produce the longest duration calls in high levels of competition. Thus, males should maintain more or less the same rank order of call duration as levels of competition within the chorus fluctuate, as has also been noted in *H. versicolor* (Schwartz et al. 2002). In contrast to call duration, call rate and call effort were not repeatable across nights. The lack of repeatability for call rate is somewhat surprising,

given that call rate trades off strongly with call duration, and that call duration is significantly repeatable. Apparently, despite this relationship, call rate maintains enough independent variance that the within-individual variance component for call rate predominates. Call effort is the product of call rate and call duration, so the lack of repeatability of call rate may explain why call effort was not repeatable. Furthermore, the strong upper limit on call effort (Reichert and Gerhardt 2012) may restrict the among-male variance for this characteristic. The fact that call duration shows relatively greater among-male variance may explain why call duration often outweighs call rate in female mate choice decisions (Gerhardt et al. 1996). Females sample male calls over a limited time period (Schwartz et al. 2004), thus it makes sense that call duration, which has very high short-term repeatability and likely can be reliably assessed by attending to just a few calls, predominates over call rate, which is much less repeatable and probably more difficult to assess (Tanner and Bee 2020).

It remains unclear whether the typical response of males to competition, increasing call duration while decreasing call rate, is aimed at competing males, mate searching females, or both. Males are certainly capable of detecting differences in these call characteristics; indeed, in addition to all of the levels of covariance discussed above, we also found that the trade-off between call rate and call duration was present at the level of short-term changes in calling in response to competition. Males significantly increased their call duration and decreased their call rate as competition increased (Table S2, S3). Nevertheless, whether longer calls are actually more effective at repelling rival males is unknown. If this is the case, then the repeatability of call duration and the strong among-individual component of the covariance between call duration and call rate would result in some males being consistently more competitive than others. In terms of female preferences, there is mixed evidence that call duration predominates over call rate in female decisions (Gerhardt et al. 1996; Ward et al. 2013), and preference strength for either characteristic weakens as its magnitude increases, which should weaken selection

towards either extreme of the trade-off axis (Tanner et al. 2017). There may be other advantages for males to resolve the trade-off by producing longer calls at lower rates, for instance if longer calls could stand out better against the background noise of the chorus, although evidence for such a benefit is limited (Bee and Schwartz 2013).

Individual variation in plasticity

As for variation in individual traits, repeatedly measuring plasticity in response to changing environmental contexts and then partitioning the variance in plasticity into its among- and within-individual components gives important insights into the causes and evolutionary consequences of limited plasticity (Dingemanse et al. 2010; Stamps and Biro 2016). Few studies have examined among-individual variance in behavioral plasticity, but of those that have, there is evidence that plasticity can be repeatable (Mathot et al. 2011; Morand-Ferron et al. 2011; Mitchell and Biro 2017). If this repeatability indeed reflects genetic variation, then there is potential for selection to act on variation in individual plasticity (Roff 1997; Nussey et al. 2007; Dingemanse et al. 2012; Brommer 2013a). However, we found no evidence for consistent among-individual variation in plasticity in calling in response to changes in the level of social competition. Thus, although the ability to respond to a competitor through plastic changes in calls likely affects success in mate competition for male *H. chrysoscelis*, there is little potential for a response to selection on calling plasticity.

Given that we found that call duration both showed long-term repeatability, and plastic expression with respect to the level of competition, we expected that plasticity in call duration in response to competition would also be repeatable. It is unclear why this was not the case. One reason may be that we measured male responses in the field, and therefore social conditions were not standardized. On any

462 given night, males may have experienced different levels of competition, and both prior experience and
463 current levels of competition are known to affect male frogs' responses to competitors (Humfeld et al.
464 2009; Reichert 2010; Heap et al. 2012). We attempted to remove the nearest competitors prior to
465 recordings, but there was certainly still variation in background calling levels that could not be
466 controlled for, and this could have introduced noise in our measurements. The magnitude of the short-
467 term repeatability (within one social context on one night) was much higher than the long-term
468 repeatability (across nights) for call duration, which was likely because of the greater variation in
469 background competition levels across than within nights. Our findings are typical of repeatability
470 studies, in which repeatability usually decreases with increasing intervals between measurements (Bell
471 et al. 2009; Boulton et al. 2014). Furthermore, we found significant intercept-slope correlations at the
472 within-male among-series level for call rate and sometimes call duration and call effort, indicating that
473 the magnitude of a male's response to competition depended on its calling behavior prior to being
474 exposed to our playbacks, which was likely determined in part by the natural level of competition it was
475 experiencing at that time. Although recording males in standardized conditions (e.g. a quiet laboratory
476 chamber) may have given us greater power to detect among-individual variation in male responses to
477 competition, this is not the biologically-relevant environment in which sexual selection is acting. Another
478 potential explanation of why we did not detect individual differences in plasticity is limited statistical
479 power. Detecting repeatable plasticity using random regressions requires large sample sizes (Martin et
480 al. 2011; van de Pol 2012), and it is possible that our sample was not sufficient for this purpose.
481
482 There was a strong negative intercept-slope correlation for call effort from the average to the
483 competitive playback. Males that were calling with high call effort in response to the average playback
484 actually often reduced their call effort in response to the competitive playback, while those with an
485 initially low call effort did the opposite. This fanning-in pattern was apparently due to a reduction in call

rate by those individuals, because a similar intercept-slope correlation was seen for call rate. Because sexual selection strongly favors higher call efforts (Tanner et al. 2017), it was unexpected that some males reduced their call efforts as competition increased. One potential explanation is that males can only achieve a high call effort for a very limited time period (Fellers 1979; Reichert and Gerhardt 2012). If that is the case, then those males that gave high effort calls already during the average playback may have been unable to maintain this performance for the subsequent competitive playback, while those who had a more moderate initial response may have had more flexibility to adjust their calls as competition continued to increase. At the population level, there was no change in call effort across these two contexts, which underscores the importance of examining behavioral variation at the individual level. Further study of the temporal dynamics of call effort across different levels of competition are needed to better characterize strategies of energy allocation in this species. At a different temporal scale, call effort was negatively related to calendar day, suggesting that males reduced energy expenditure on calling as the season progressed, perhaps due to decreased reserves. In *H. versicolor*, males did not reduce call effort across the season, but spent less time calling per night, which would similarly reduce energetic expenditures (Runkle et al. 1994).

Trade-offs and sources of covariance

A negative phenotypic correlation between traits is the hallmark of a trade-off, but these data alone are insufficient to determine at what level(s) of variation the trade-off takes place, and therefore what the consequences are for behavioral expression and trait evolution. Indeed, a negative covariance between traits measured at the population level may have zero or even positive covariance for specific variance components. Therefore, in the absence of pedigreed or genetic data, it is necessary to take repeated measurements of all traits involved in a trade-off to partition among the different levels of covariance

510 (Dingemanse and Dochtermann 2013; Moiron et al. 2016; Berberi and Careau 2019). In our case, we
511 found negative covariance between call duration and call rate at every level of analysis. This likely
512 reflects the fact that the high energetic costs of increasing both call duration and call rate are applicable
513 to both within- and among-individual variation, ultimately imposing a trade-off at both levels.
514 Nevertheless, we found that the covariance was strongest at the among-individual level, and was much
515 weaker at the level of within-male variation from one call to the next. One caveat is that because we
516 recorded males in natural choruses, some among-individual covariance could be environmental in
517 origin, if males consistently differ in their tendency to call from more or less competitive areas of the
518 pond (Aplin et al. 2015). We controlled for some social factors in our models, but until genetic
519 covariances are measured in this species we cannot rule out that environmental sources could have
520 inflated estimates of among-individual covariance (Niemelä and Dingemanse 2017).
521
522 In the closely-related *H. versicolor*, there was a significant genetic correlation between call duration and
523 call rate that was of a similar magnitude to their phenotypic correlation (Welch et al. 2014). Our finding
524 of significant among-individual phenotypic covariance between call duration and call rate in *H.*
525 *chrysoscelis* therefore indicates the potential for genetic covariance in this species as well. The
526 evolutionary significance of a negative among-individual covariance is that, if this indeed has a genetic
527 component, there is potential for a correlated response to selection (Lande and Arnold 1983; Blows and
528 Hoffmann 2005; Walsh and Blows 2009). However, the strongest female preferences in this species are
529 for calls with both longer call durations and higher call rates, and thus higher call efforts (Tanner et al.
530 2017), and we found no evidence for among-individual variance in call effort. Thus, the response to
531 selection on call characteristics may be limited, because most genetic variance would be perpendicular
532 to the major axis of sexual selection (McGuigan et al. 2008). One final consideration is that the male
533 advertisement call is an integrated phenotype made up of many components (Reichert and Höbel 2018),

and our analyses focused on only two of these. We did not examine a third characteristic, call amplitude, which is also under sexual selection (Fellers 1979) and contributes to the energetic cost of calling (Prestwich 1994). Although within-individual variation in call amplitude appears to be minimal (Gerhardt 1975; Love and Bee 2010), whether there are consistent among-individual differences in amplitude, and whether it covaries with other call traits, is unknown. An important goal for future studies is therefore to attempt to characterize more fully the interrelationships among many potentially covarying traits, and determine at what levels they (co)vary (Blows and Hoffmann 2005; Walsh and Blows 2009).

ACKNOWLEDGEMENTS

Nicole Clapp, James Erdmann, A.J. Hager and Cheyenne Smith assisted with the field work. Rachel Atherton contributed to the acoustic analyses. M.M. was funded by an Alexander von Humboldt Research Fellowship for Postdoctoral Researchers.

547 REFERENCES

548

549 Agrawal, A. A., J. K. Conner, and S. Rasmann. 2010. Tradeoffs and negative correlations in evolutionary
550 ecology. Pp. 243–268 *in* M. Bell, W. Eanes, D. Futuyma, and J. Levinton, eds. Evolution after
551 Darwin: The first 150 years. Sinauer, Sunderland, MA.

552 Aplin, L. M., J. A. Firth, D. R. Farine, B. Voelkl, R. A. Crates, A. Culina, C. J. Garroway, C. A. Hinde, L. R.
553 Kidd, I. Psorakis, N. D. Milligan, R. Radersma, B. L. Verhelst, and B. C. Sheldon. 2015. Consistent
554 individual differences in the social phenotypes of wild great tits, *Parus major*. Anim. Behav.
555 108:117–127.

556 Araya-Ajoy, Y. G., and N. J. Dingemanse. 2017. Repeatability, heritability, and age-dependence of
557 seasonal plasticity in aggressiveness in a wild passerine bird. J. Anim. Ecol. 86:227–238.

558 Araya-Ajoy, Y. G., K. J. Mathot, and N. J. Dingemanse. 2015. An approach to estimate short-term, long-
559 term and reaction norm repeatability. Methods Ecol. Evol. 6:1462–1473.

560 Badyaev, A. V., and C. K. Ghalambor. 2001. Evolution of life histories along elevational gradients: Trade-
561 off between parental care and fecundity. Ecology 82:2948–2960.

562 Baugh, A. T., K. Van Oers, N. J. Dingemanse, and M. Hau. 2014. Baseline and stress-induced
563 glucocorticoid concentrations are not repeatable but covary within individual great tits (*Parus*
564 *major*). Gen. Comp. Endocrinol. 208:154–163.

565 Bee, M. A., and J. J. Schwartz. 2013. Calling in gray treefrog choruses: modifications and mysteries. Proc.
566 Meet. Acoust. 19:010054–010054.

567 Bell, A. M., S. J. Hankison, and K. L. Laskowski. 2009. The repeatability of behaviour: a meta-analysis.
568 Anim. Behav. 77:771–783.

569 Berberi, I., and V. Careau. 2019. Performance trade-offs in wild mice. Oecologia 191:11–23.

570 Biro, P. A., and J. A. Stamps. 2010. Do consistent individual differences in metabolic rate promote

consistent individual differences in behavior? Trends Ecol. Evol. 25:653–659.

Biro, P., T. Garland, C. Beckmann, B. Ujvari, F. Thomas, and J. Post. 2018. Metabolic scope as a proximate constraint on individual behavioral variation: effects on personality, plasticity, and predictability. Am. Nat. 192:142–154.

Blows, M. W., and A. A. Hoffmann. 2005. A reassessment of genetic limits to evolutionary change. Ecology 86:1371–1384.

Boulton, K., A. J. Grimmer, G. G. Rosenthal, C. A. Walling, and A. J. Wilson. 2014. How stable are personalities? A multivariate view of behavioural variation over long and short timescales in the sheepshead swordtail, *Xiphophorus birchmanni*. Behav. Ecol. Sociobiol. 68:791–803.

Brommer, J. E. 2013a. Phenotypic plasticity of labile traits in the wild. Curr. Zool. 59:485–505.

Brommer, J. E. 2013b. Variation in plasticity of personality traits implies that the ranking of personality measures changes between environmental contexts: Calculating the cross-environmental correlation. Behav. Ecol. Sociobiol. 67:1709–1718.

Brommer, J. E., J. Merilä, B. C. Sheldon, and L. Gustafsson. 2005. Natural selection and genetic variation for reproductive reaction norms in a wild bird population. Evolution 59:1362–1371.

Careau, V., and R. S. Wilson. 2017. Of uberfleas and krakens: Detecting trade-offs using mixed models. Integr. Comp. Biol. 57:362–371.

Dingemanse, N. J., I. Barber, J. Wright, and J. E. Brommer. 2012. Quantitative genetics of behavioural reaction norms: genetic correlations between personality and behavioural plasticity vary across stickleback populations. J. Evol. Biol. 25:485–496.

Dingemanse, N. J., and N. A. Dochtermann. 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. J. Anim. Ecol. 82:39–54.

Dingemanse, N. J., A. J. N. Kazem, D. Réale, and J. Wright. 2010. Behavioural reaction norms: animal personality meets individual plasticity. Trends Ecol. Evol. 25:81–89.

595 Dingemanse, N. J., and M. Wolf. 2013. Between-individual differences in behavioural plasticity within
596 populations: causes and consequences. Anim. Behav. 85:1031–1039.
597 Dubois, F. 2019. Why are some personalities less plastic? Proc. R. Soc. B Biol. Sci. 286:20191323.
598 Evans, J. P. 2010. Quantitative genetic evidence that males trade attractiveness for ejaculate quality in
599 guppies. Proc. R. Soc. B Biol. Sci. 277:3195–3201.
600 Fellers, G. M. 1979. Aggression, territoriality, and mating behaviour in North American treefrogs. Anim.
601 Behav. 27:107–119.
602 Garland, T. 2014. Trade-offs. Curr. Biol. 24:R60–R61.
603 Gayou, D. C. 1984. Effects of temperature on the mating call of *Hyla versicolor*. Copeia 1984:733–738.
604 Gerhardt, H. C. 2005. Acoustic spectral preferences in two cryptic species of grey treefrogs: implications
605 for mate choice and sensory mechanisms. Anim. Behav. 70:39–48.
606 Gerhardt, H. C. 1975. Sound pressure levels and radiation patterns of the vocalizations of some North
607 American frogs and toads. J. Comp. Physiol. A 102:1–12.
608 Gerhardt, H. C., M. L. Dyson, and S. D. Tanner. 1996. Dynamic properties of the advertisement calls of
609 gray tree frogs: patterns of variability and female choice. Behav. Ecol. 7:7–18.
610 Giraldeau, L. A., and T. Caraco. 2000. Social Foraging Theory. Princeton University Press, Princeton.
611 Hadfield, J. D. 2010. MCMCglmm: MCMC methods for multi-response GLMMs in R. J. Stat. Softw. 33:1–
612 22.
613 Heap, S., D. Stuart-Fox, and P. Byrne. 2012. Variation in the effect of repeated intrusions on calling
614 behavior in a territorial toadlet. Behav. Ecol. 23:93–100.
615 Highcock, L., and A. J. Carter. 2014. Intraindividual variability of boldness is repeatable across contexts in
616 a wild lizard. PLoS One 9:e95179.
617 Houslay, T. M., and A. J. Wilson. 2017. Avoiding the misuse of BLUP in behavioural ecology. 00:1–5.
618 Humfeld, S. C., V. T. Marshall, and M. A. Bee. 2009. Context-dependent plasticity of aggressive signalling

619 in a dynamic social environment. Anim. Behav. 78:915–924.

620 Katz, K., and D. Naug. 2015. Energetic state regulates the exploration-exploitation trade-off in

621 honeybees. Behav. Ecol. 26:1045–1050.

622 Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution

623 37:1210–1226.

624 Laskowski, K. L., M. Moiron, and P. T. Niemelä. 2021. Integrating behavior in life-history theory:

625 Allocation versus acquisition? Trends Ecol. Evol. 36:132–138.

626 Ligges, U., S. Krey, O. Mersmann, and S. Schnackenberg. 2013. tuneR: Analysis of music and speech. URL:

627 https://CRAN.R-project.org/package=tuneR.

628 Love, E. K., and M. A. Bee. 2010. An experimental test of noise-dependent voice amplitude regulation in

629 Cope's grey treefrog, *Hyla chrysoscelis*. Anim. Behav. 80:509–515.

630 Martin, J. G. A., D. H. Nussey, A. J. Wilson, and D. Réale. 2011. Measuring individual differences in

631 reaction norms in field and experimental studies: A power analysis of random regression models.

632 Methods Ecol. Evol. 2:362–374.

633 Mathot, K. J., P. J. van den Hout, T. Piersma, B. Kempenaers, D. Réale, and N. J. Dingemanse. 2011.

634 Disentangling the roles of frequency-vs. state-dependence in generating individual differences in

635 behavioural plasticity. Ecol. Lett. 14:1254–62.

636 McGuigan, K., A. Van Homrigh, and M. W. Blows. 2008. An evolutionary limit to male mating success.

637 Evolution 62:1528–37.

638 Mitchell, D. J., and P. A. Biro. 2017. Is behavioural plasticity consistent across different environmental

639 gradients and through time? Proc. R. Soc. B Biol. Sci. 284:20170893.

640 Moiron, M., K. J. Mathot, and N. J. Dingemanse. 2016. A multi-level approach to quantify speed-accuracy

641 trade-offs in great tits (*Parus major*). Behav. Ecol. 27:1539–1546.

642 Morand-Ferron, J., E. Varennes, and L. A. Giraldeau. 2011. Individual differences in plasticity and

643 sampling when playing behavioural games. Proc. R. Soc. B 278:1223–1230.
644 Niemelä, P. T., and N. J. Dingemanse. 2017. Individual versus pseudo-repeatability in behaviour: Lessons
645 from translocation experiments in a wild insect. J. Anim. Ecol. 86:1033–1043.
646 Nussey, D. H., E. Postma, P. Gienapp, and M. E. Visser. 2005. Selection on heritable phenotypic plasticity
647 in a wild bird population. Science 310:304–306.
648 Nussey, D. H., A. J. Wilson, and J. E. Brommer. 2007. The evolutionary ecology of individual phenotypic
649 plasticity in wild populations. J. Evol. Biol. 20:831–844.
650 O'Dea, R. E., D. W. A. Noble, and S. Nakagawa. 2022. Unifying individual differences in personality,
651 predictability and plasticity: A practical guide. Methods Ecol. Evol. 13:278–293.
652 Prestwich, K. N. 1994. The energetics of acoustic signaling in anurans and insects. Am. Zool. 34:625–643.
653 Price, T., and T. Langen. 1992. Evolution of correlated characters. Trends Ecol. Evol. 7:307–310.
654 R Development Core Team. 2021. R: A language and environment for statistical computing. R
655 Foundation for Statistical Computing, Vienna, Austria.
656 Reichert, M. S. 2010. Aggressive thresholds in *Dendropsophus ebraccatus*: Habituation and sensitization
657 to different call types. Behav. Ecol. Sociobiol. 64:529–539.
658 Reichert, M. S., and H. C. Gerhardt. 2012. Trade-offs and upper limits to signal performance during
659 close-range vocal competition in gray tree frogs *Hyla versicolor*. Am. Nat. 180:425–437.
660 Reichert, M. S., and G. Höbel. 2018. Phenotypic integration and the evolution of signal repertoires: A
661 case study of treefrog acoustic communication. Ecol. Evol. 8:3410–3429.
662 Reznick, D., L. Nunney, and A. Tessier. 2000. Big houses, big cars, superfleas and the costs of
663 reproduction. Trends Ecol. Evol. 15:421–425.
664 Roff, D. 1997. Evolutionary Quantitative Genetics. Chapman and Hall, New York.
665 Roff, D. A., and D. J. Fairbairn. 2007. The evolution of trade-offs: where are we? J. Evol. Biol. 20:433–447.
666 Roff, D. A., and A. J. Wilson. 2014. Quantifying genotype-by-environment interactions in laboratory

667 systems. Pp. 101–136 *in* J. Hunt and D. Hosken, eds. Genotype-by-Environment Interactions and
668 Sexual Selection. John Wiley & Sons, Ltd.
669 Runkle, L. S., K. D. Wells, C. C. Robb, and S. L. Lance. 1994. Individual, nightly, and seasonal variation in
670 calling behavior of the gray tree frog, *Hyla versicolor*: implications for energy expenditure. Behav.
671 Ecol. 5:318–325.
672 Schwartz, J. J., B. Buchanan, and H. C. Gerhardt. 2002. Acoustic interactions among male gray treefrogs,
673 *Hyla versicolor*, in a chorus setting. Behav. Ecol. Sociobiol. 53:9–19.
674 Schwartz, J. J., K. Huth, and T. Hutchin. 2004. How long do females really listen? Assessment time for
675 female mate choice in the grey treefrog, *Hyla versicolor*. Anim. Behav. 68:533–540.
676 Simmons, L. W., and D. J. Emlen. 2006. Evolutionary trade-off between weapons and testes. Proc. Natl.
677 Acad. Sci. 103:16346–16351.
678 Sinervo, B., and E. Svensson. 2002. Correlational selection and the evolution of genomic architecture.
679 Heredity 89:329–338.
680 Snell-Rood, E. C. 2013. An overview of the evolutionary causes and consequences of behavioural
681 plasticity. Anim. Behav. 85:1004–1011.
682 Stamps, J. A. 2016. Individual differences in behavioural plasticities. Biol. Rev. 91:534–567.
683 Stamps, J., and P. Biro. 2016. Personality and individual differences in plasticity. Curr. Opin. Behav. Sci.
684 12:18–23.
685 Stearns, S. C. 1989. Trade-offs in life-history evolution. Funct. Ecol. 3:259–268.
686 Sueur, J., T. Aubin, and C. Simonis. 2008. Seewave, a free modular tool for sound analysis and synthesis.
687 Bioacoustics 18:213–226.
688 Sullivan, B. K., and S. H. Hinshaw. 1992. Female choice and selection on male calling behaviour in the
689 grey treefrog *Hyla versicolor*. Anim. Behav. 44:733–744.
690 Tanner, J. C., and M. A. Bee. 2020. Inconsistent sexual signaling degrades optimal mating decisions in

691 animals. Sci. Adv. 6:eaax3957.

692 Tanner, J. C., J. L. Ward, R. G. Shaw, and M. A. Bee. 2017. Multivariate phenotypic selection on a
693 complex sexual signal. Evolution 71:1742–1754.

694 van de Pol, M. 2012. Quantifying individual variation in reaction norms: how study design affects the
695 accuracy, precision and power of random regression models. Methods Ecol. Evol. 3:268–280.

696 van Noordwijk, A. J., and G. de Jong. 1986. Acquisition and allocation of resources: Their influence on
697 variation in life history tactics. Am. Nat. 128:137–142.

698 Walsh, B., and M. W. Blows. 2009. Abundant genetic variation + strong selection = multivariate genetic
699 constraints: A geometric view of adaptation. Annu. Rev. Ecol. Evol. Syst. 40:41–59.

700 Ward, J. L., E. K. Love, A. Vélez, N. P. Buerkle, L. R. O. Bryan, and M. A. Bee. 2013. Multitasking males and
701 multiplicative females: dynamic signalling and receiver preferences in Cope's grey treefrog. Anim.
702 Behav. 86:231–243.

703 Welch, A. M., M. J. Smith, and H. Carl Gerhardt. 2014. A multivariate analysis of genetic variation in the
704 advertisement call of the gray treefrog, *Hyla versicolor*. Evolution 68:1629–1639.

705 Wolf, M., G. S. Van Doorn, and F. J. Weissing. 2011. On the coevolution of social responsiveness and
706 behavioural consistency. Proc. R. Soc. B Biol. Sci. 278:440–448.

707

708