Behavioral Ecology

The official journal of the
ISBE
International Society for Behavioral Ecology

Behavioral Ecology (2020), 31(1), 222–231. doi:10.1093/beheco/arz180

Original Article

Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity

David J. Mitchell,[a,b] Antoine M. Dujon,[a] Christa Beckmann,[a,c] and Peter A. Biro[a]

[a]Centre for Integrative Ecology, School of Life & Environmental Sciences, Deakin University, 75 Pigdons Road, Geelong VIC 3216, Australia, [b]Department of Zoology/Ethology, Stockholm University, Svante Arrheniusväg 18B. SE-10691, Stockholm, Sweden, and [c]School of Science, Western Sydney University, Locked Bag 1797, Penrith, NSW, 2751, Australia

Received 26 May 2019; revised 3 September 2019; editorial decision 4 September 2019; accepted 23 September 2019; Advance Access publication 1 November 2019.

Quantifying individual variation in labile physiological or behavioral traits often involves repeated measures through time, so as to test for consistency of individual differences (often using repeatability, "*R*") and/or individual differences in trendlines over time. Another form of temporal change in behavior is temporal autocorrelation, which predicts observations taken closely together in time to be correlated, leading to nonrandom residuals about individual temporal trendlines. Temporal autocorrelation may result from slowly changing internal states (e.g., hormone or energy levels), leading to slowly changing behavior. Autocorrelation is a well-known phenomenon, but has been largely neglected by those studying individual variation in behavior. Here, we provide two worked examples which show substantial temporal autocorrelation ($r > 0.4$) is present in spontaneous activity rates of guppies (*Poecilia reticulata*) and house mice (*Mus domesticus*) in stable laboratory conditions, even after accounting for temporal plasticity of individuals. Second, we show that ignoring autocorrelation does bias estimates of *R* and temporal reaction norm variances upwards, both in our worked examples and in separate simulations. This bias occurs due to the misestimation of individual-specific means and slopes. Given the increasing use of technologies that generate behavioral and physiological data at high sampling rates, we can now study among- and within-individual changes in behavior in more detailed ways, including autocorrelation, which we discuss from biological and methodological perspectives and provide recommendations and annotated R code to help researchers implement these models on their data.

Key words: **endogenous plasticity, individual gambit, intraindividual variability, pseudorepeatability, temporal plasticity, serial correlation, slowly-changing states.**

INTRODUCTION

Evolutionary and behavioral ecologists are fundamentally interested in understanding the existence and maintenance of phenotypic variation. As such, researchers studying labile physiological and behavioral traits often aim first to test whether the trait of interest differs between individuals. To do so, repeated samples are taken from individuals in order to partition the variance due to among- versus within-individual sources of variance, standardized to a proportion given by repeatability (*R*, also called the "intraclass correlation coefficient"; Bell et al. 2009; Wolak et al. 2012). Thus, a substantial *R* estimate indicates among-individual variation in predicted mean values—often assumed to reflect some underlying intrinsic and stable factors, such as genetics or induced developmental effects (Falconer 1981; Dohm 2002; Wilson 2018).

This assumption can easily be violated by confounding variables which exaggerate the estimated *R*, sometimes termed "pseudo-repeatability" (Dingemanse and Dochtermann 2013; Zsebők et al. 2017). This upward bias of *R* can result from environmental differences between locations where individuals are observed (Niemelä and Dingemanse 2017), or from environmental change through time (Zsebők et al. 2017), both internal and external. In order for our estimates of *R* to provide meaningful inferences on the extent of intrinsic individual differences, we should therefore control for extrinsic factors affecting individual variation (Niemelä and Dingemanse 2017; Wilson 2018) and assess the temporal stability of labile traits (Biro and Stamps 2015).Temporal change in labile traits has typically involved modeling individual temporal trendlines as "reaction norms"

Address correspondence to D.J. Mitchell. E-mail: david.mitchell@zoologi.su.se.

(Biro and Stamps 2015). This approach has proven particularly useful in considering the effect of habituation or acclimation (Martin and Réale 2008; Biro 2012; Dingemanse et al. 2012), ontogeny (Biro et al. 2014; Brommer and Class 2015), or seasonal and diurnal change (Carter et al. 2012), where assumed intrinsic differences lead to individual differences in temporal change (i.e., an IDxTime interaction). For instance, rates of development and senescence predict the rate at which a trait should change as a function of age (Brommer and Class 2015).

Behavior may also be temporally autocorrelated. Many potential processes might cause two measures of behavior to be correlated. Some of these, such as individual identity or environmental correlates that change more slowly than the time course of measurements (e.g., season), may already be included in a model, but if residuals of successive data points remain correlated, then additional forces are at work and estimates of included factors may be biased. Thus, temporal autocorrelation can be necessary for unbiased parameter estimates and also could provide clues as to previously undetected and potentially interesting processes. From a statistical perspective, (positive) autocorrelation predicts temporal clustering of residuals around individual trendlines (Figure 1a), in contrast to temporal reaction norms which predict individual means to change over time (black line Figure 1a,b), and residuals to be random (Figure 1b). Thus, autocorrelation represents a form of inertia or lag, which results in slowly changing trait values about an individuals' temporal trendline. So, while reaction norms model a predictable and directional change in behavior through time, autocorrelation models a nonpredictable, slowly changing states of behavior.

To date, autocorrelation has received little consideration in the behavioral literature (but see; Westneat et al. 2011; Nakayama et al. 2016; Villegas-Ríos et al. 2018). Indeed, in two thorough reviews dealing with residual variance and temporal change in labile traits, temporal autocorrelation was not mentioned (Westneat et al. 2015; Stamps 2016). Yet, this within-individual patterning of trait values ("residuals" in statistical terms) can reveal important biological processes, and has implications for study design and data analysis, as we will discuss in this paper.

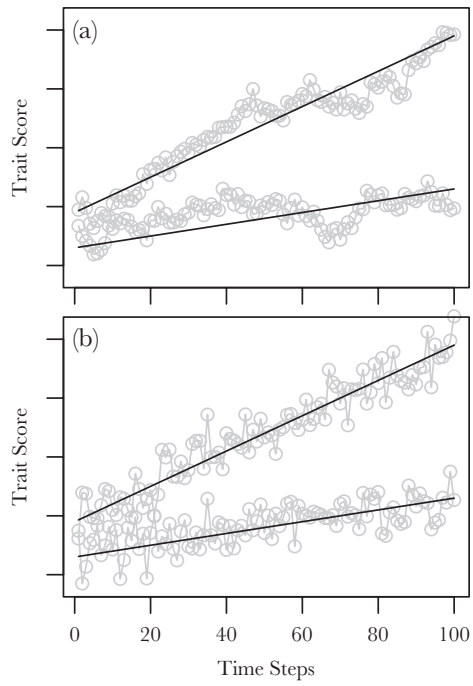

Figure 1
Depicted are hypothetical "random walks" of repeated measures through time. Two hypothetical individuals are shown differing in their temporal reaction norms (black lines). The trait scores are then shown in grey. Panel (a) showing individuals displaying obvious autocorrelation and (b) no autocorrelation (residual trait scores are random about trendlines).

The biology underlying autocorrelation in behavior

The expression of behavior is affected by the underlying internal state of individuals (Stamps 2016), such as their energetic state (Krause et al. 2017; Mitchell and Biro 2017) or circulating hormone levels (Gerall et al. 1973; Girard and Garland 2002), just to name two factors. These internal factors often have a degree of inertia, as they tend not to change instantly, and it is this lagging effect which constrains the rate of change in behavior and creates autocorrelation. For instance, thermal inertia prevents large swings in body temperature across time intervals of minutes or more (Bell 1980), which affects the energetic state and behavior of the animal (Pruitt et al. 2011; Briffa et al. 2013). Similarly, hormones cannot be absorbed instantly, and after they are perturbed from their homeostatic state a lag period follows before they return to allostasis (Romero et al. 2009). These two examples (among a multitude of other potential physiological constraints) would likely create temporal autocorrelation over relatively short periods (e.g., minutes to hours).

Other factors, however, could affect behavior over longer time intervals. For instance, muscle size and aerobic capacity develop in response to training (Houle-Leroy et al. 2000; Girard and Garland 2002) and could affect behavior over days to weeks. Similarly, environmental temperature or prey abundance change slowly across hours to weeks or more, which will lead to temporal patterns in the residuals (if the variable is omitted from the analysis) (Nakayama et al. 2016; Allegue et al. 2017). Therefore, if assays of an animal's behavior are taken with a relatively short interval between them, both the internal physiological state and the external environmental conditions are likely to be similar. This will therefore add a confound which will lead to similar behavioral scores.

Autocorrelation is largely ignored by behaviorists

While rarely considered in the physiological and behavioral literatures, temporal autocorrelation is a phenomenon commonly accounted for in movement ecology (e.g., Wittemyer et al. 2008; Dray et al. 2010) and when using dataloggers which collect data at high temporal resolution and over prolonged periods of time (e.g., do Amaral et al. 2002). Not surprisingly, short intervals between observations are common in the study of individual variation, creating a situation where problems associated with undetected and/or unaccounted for pseudoreplication seem likely. For example, over 20% of avian studies reporting behavioral and physiological trait repeatability had intervals between observations shorter than 5 days (Holtmann et al. 2016). As more labs adopt the use of tracking software, among other automated technologies (e.g., Nakayama et al. 2016; Bierbach et al. 2017; Mitchell and Biro 2017; Houslay et al. 2018; Villegas-Ríos et al. 2018), the potential (and need) to assess autocorrelation will become more common. Hence, it is particularly timely to highlight this phenomenon for an audience we feel has yet to fully appreciate it. Fortunately, accounting for this in our models is quite easy to implement as we show here.

Problems associated with undetected autocorrelation

When observations are temporally autocorrelated and this is not accounted for in the analysis, R estimates will be inflated. This is because individuals are likely to have been assayed under similar internal states or external conditions. For this reason, from intuition no one would perform many repeated observations at intervals of 1 min and then report a significant R value as being biologically meaningful individual variation; rather, most would separate observations by hours or days or more to avoid implicit autocorrelation. Indeed, the literature seems to support the prediction that R decreases with time, as meta-analyses of the repeatability of behavioral, metabolic and hormone traits observed that studies with observations taken over longer periods of time typically found lower R values (Bell et al. 2009; White et al. 2013; Holtmann et al. 2016; Fanson and Biro 2019), although counter (null) examples exist in hormone levels (Holtmann et al. 2016; Schoenemann and Bonier 2018).

Another potential and related problem of undetected autocorrelation is the possibility for pseudoreplicated observations in the residuals to be erroneously attributed to among individual variation in temporal trajectories. For this reason, it is important to evaluate whether individuals differ in temporal trendlines by modeling time as a random slope effect in conjunction with modeling autocorrelation of the residuals. Otherwise, solely fitting individual temporal reaction norms could lead to misattributing residual variance to an among-individual source. While temporal plasticity describes systematic mean change in labile traits through time (Biro and Stamps 2015) and can take on a diverse range of shapes (Morrissey and Liefting 2016), their use alone in describing autocorrelated data will prove insufficient and at times misleading. This is because autocorrelation is simply defined by the correlation of adjacent time points, with no predictable change in means (Figure 1a).

In this paper, we explore the topic of autocorrelation in labile traits using empirical data sets and through simulations. We first show that substantial temporal autocorrelation is present in behavioral data even under controlled lab conditions, and after accounting for individual variation in temporal plasticity, suggesting that unknown internal state variables may lead to these slow changes in behavior. Next, we show that ignoring this autocorrelation when modeling these datasets results in upwardly biased estimates of R and temporal slope variation, which affects our biological inferences drawn from the data. We then use these empirical estimates of autocorrelation from the fish and mouse studies to inform simulations of data which confirm how temporal autocorrelation biases estimates of R and slope variances upwards. In light of the empirical data and simulations, we discuss how autocorrelation may lead to biases in the analysis of temporal plasticity, biases in correlations between traits, and suggest sampling and analytical tools to account for these effects. We conclude with a discussion about what autocorrelation can tell us about the biology, including likely proximate constraints to behavior and how it contributes to the recent debate and surge of interest in studying residual variance (behavioral predictability).

METHODS

Guppies

Male guppies were randomly sampled from three stock populations, residing in large 1600 L ponds at Deakin University, Geelong, Australia. Male fish were sampled in 3 batches of 40 fish from each population at a time (N = 120). Fish were housed individually in 1.25 L tanks (25 cm × 6 cm × 15 cm) placed on a recirculation system and left to acclimate for 7 days. Fish were kept under stable conditions (Temperature = 24.7 ± 0.2 °C, pH = 7.9–8.1, kH = 100–120 ppm, salinity = 1.1–1.35 ppt) and a 12:12 h light:dark photoperiod (7 am–7 pm day), consistent with the stock pond conditions. Each morning they received enough hydrated commercial flake food to allow some sink to the bottom uneaten and in the afternoon received 0.5 mL of concentrated brine shrimp *nauplii*.

Activity rates were measured using EthoVision XT10 tracking software. To standardize hunger levels, fish received their flake food ration approximately 25 min before a trial. Fish in their home tanks were then picked up and placed on a 12-arena stage, backlit with infrared light. A camera recorded the fish from a side view at a distance of 2.6 m. As the tanks were narrow (see above), swimming movements were restricted to being predominately in two dimensions. Once all fish were in place, the next groups of fish were fed in readiness for their trial, the door to the animal holding room was closed and EthoVision was set to track from an adjacent room where trials could be monitored. Trials ran for 22 min with the first 2 min of the trial discarded to negate residual effects of handling and the experimenter's activity in the room. Activity was measured as the cumulative distance moved in the 20-min period of tracking. The temperature of a tank was checked before and after each trial and was always at 24.7 ± 0.2 °C. At the conclusion of the trial, tanks were returned to the recirculation system.

Trials took place between 9 am and 1 pm and were repeated daily for 14 days (N_{obs} = 1626). Due to a power outage, trials were not run on day 9 for the second batch of fish. To account for any potential effects of differences in handling due to the missed trial, fish were moved to the stage, then returned to the rack after 10 min, without collecting data on the afternoon of that day. One fish died during the experiment for unknown reasons and was discarded from the analysis (N_{ID} = 119). Data are available on Dryad (Mitchell et al. 2019).

House mice

We used existing data collected as part of a long-term experiment where voluntary wheel running was artificially selected on (Eisenmann et al. 2009). House mice were 6–8 weeks old and placed in an enclosure under a temperature of 22 °C, a 12:12 h light:dark photoperiod and provided food and water *ad libitum*. Mice were provided with a running-wheel which tracked the number of rotations made each day. Data are derived from a control line and an up-selected line, and from each line 10 males and 10 females were sampled (N_{ID} = 39). The number of wheel rotations were tracked daily for 21 days, though no observations were made on day 5 of week 1 (N_{obs} = 779). More information of this experiment can be found in Eisenmann et al. (2009) and the raw data are available on Dryad (Biro et al. 2018).

Statistical methods

The two datasets were analyzed using linear mixed effect models. The guppy data were fit with a random intercept of ID, with both fixed and random slopes of day post arrival on the system and time of day (9:00–14:00). The mice data were fit with the fixed effects of selection line, sex, and day, with a random intercept and slope of day for individuals. Wheel revolutions were log-transformed, and distance moved was square root transformed to achieve normality of the residuals. In the mice data, day was also square-root transformed to aid linearity. Response variables and temporal predictor variables were

z-transformed (set to a mean of 0 and SD of 1) to aid interpretation of variance parameters that are compared across both datasets.

In addition to the random effects, models for both datasets were fit with a temporal autocorrelation parameter, another class of random effects on the residual side, which accounts for temporal dependence of observations. Ignoring autocorrelation would otherwise yield a degree of pseudoreplication and an overestimation of the effective sample size. Data were fit with a first-order auto-regression structure (AR1), which calculates the correlation of residuals taken 1 unit of time apart, in this case 1 day. Simply put, the first-order autocorrelation yields the correlation coefficient between the residual observation of any given day and the residual of the day which preceded it. Thus, it assumes observations to be taking a random walk through the residual distribution (Figure 1) (Zuur et al. 2009).

The AR1 structure is only informed by observations taken 1 unit of time apart, but makes implicit predictions on the correlation between observations taken at a greater interval and the temporal spacing at which two observations become effectively independent (i.e., the correlation has vanished). Autocorrelations can also be set to evaluate observations taken 2 days (AR2) or more apart, which is additive to the AR1 prediction (Zuur et al. 2009; Box et al. 2016). However, we only consider AR1 in this paper for simplicity and due to the relatively short duration of our two datasets (14 and 21 days, respectively).

Models were fit using the R package "nlme" (Pinheiro et al. 2017) and tests of significance of the fixed effects were calculated using *F*-tests derived from models fit with maximum likelihood. Tests of significance of the random effects, including the autoregressive terms, were conducted with likelihood ratio tests from models fit by restricted maximum likelihood. Annotated model code and output can be found in the Supplementary Material 1.

Simulation

We studied the effect of autocorrelation on *R* through simulating data of a known, simplified structure. A simple random intercept model was simulated in R software (R Core Team 2017), with no mean effect of time on the individual or population level trajectory. Simulations each yielded a true $R = 0.4$, consistent with that observed in most behavioral and metabolic studies (Bell et al. 2009; White et al. 2013) and autocorrelations of 0.2, 0.4, and 0.6 were used, in line with estimates from our case studies (see below), and those derived from a field study (Villegas-Ríos et al. 2018). Simulations using different values of *R* can be found in the Supplementary Material 2, Figure S3, which show patterns consistent to the $R = 0.4$ simulation. We chose to simulate data, rather than explore the effect of autocorrelation in the empirical data so as to have data of a known structure.

Each of the three resultant simulated datasets (with AR(1) = 0.2, 0.4, and 0.6) had 51 observations taken 1 unit of time apart from 50 subjects. In order to assess the effect of autocorrelation on biasing our estimates of *R* when not accounted for, the simulated data was fit with random intercept models, ignoring the effect of autocorrelation. We aimed to study the effect of two factors of sampling on the estimated repeatability: 1) the period of time between successive observations, and 2) the total length of the experiment. To do so, we culled the data to follow different sampling designs, by first considering only every second or every fifth observation, thus creating data sets with 26 and 11 observations, respectively. The expected correlation of observations taken at a lag of "x" steps apart can be calculated using the AR1 correlation coefficient to the power of the "x" (i.e., r^x). Thus, from the simulated correlations of 0.2, 0.4, and 0.6, observations would cease to be correlated and become independent samples ($r \leq 0.01$) at lags of 3, 5, and 9 steps, respectively. Thus, the culling to samples taken every fifth step aims to show the diminished effect of autocorrelation after an appropriate lag period. For each of these culling frequencies, we sequentially culled observations until we considered just the first two observations.

We also wished to explore the effect of autocorrelation on the estimate of random slope variances (i.e., the extent to which individual temporal trendlines differed in slope). To do so, we simulated a random slope variance of 0.05, with no covariance with the intercept; the predictor variable, time-steps, was mean-centered. Intercept variance was set to 0.6 and residual variance was set to 0.4. This was done for the three AR correlation values considered (0.2, 0.4, and 0.6). Time was centered so that the 51 observations varied from −25 to 25 and the models were the culled to 5 observations (−2 to 2). Observations were added two at a time to retain the centering of the data considered (i.e., increased to −3 to 3, then −4 to 4, etc.) until all 51 observations were considered (−25 to 25). To help models consistently converge across the 100 iterations, we simulated 200 individuals to increase power.

Models were fit for each culled dataset and each iteration, from which we extracted the *R* estimate or, slope variance and residual variance. Importantly, our study aims to explore the effect of autocorrelation on biasing parameter estimates, rather than their effect on the precision. For this reason, we simulated well sampled data with high numbers of individuals, which we then iterated 1000 times to get to the mean bias. Thus, our study does not consider the precision or power of these sampling regimes, which should also be considered when designing a sampling regime (see Wolak et al. 2012; Dingemanse and Dochtermann 2013; Kain et al. 2015).

RESULTS

Guppies

Fish acclimated to the new conditions and handling, as seen by an increase in activity through time, and were more active later in the morning (Table 1). There was additionally substantial among-individual variation in intercepts and slopes across days. Individuals did not differ in their response to time of day ($\sigma^2_{slope} = 0.002$, $\chi^2_3 = 4.23$, $P = 0.12$), and this parameter was thus discarded from the model. Together, the fixed and random effects explained much of the variance, as indicated by the low residual variance (0.26).

Table 1

Shown are all parameter estimates of the guppy data, with associated 95% confidence intervals. Significant parameters are shown in bold. Annotated model code can be found in the Supplementary Material 1

Parameter	Estimate (95% CI)	$F_{numDF,denDF}$	*P*
Fixed effects			
Intercept	0.004 (−0.147, 0.156)	$0.003_{1,1505}$	0.954
Days on rack	**0.111 (0.055, 0.167)**	$\mathbf{14.2_{1,1505}}$	**<0.0001**
Time of Day	**0.025 (0.004, 0.046)**	$\mathbf{5.7_{1,1505}}$	**0.017**
	Estimate (95% CI)	χ^2	*P*
Variance-Covariance parameters			
Intercept	**0.666 (0.506, 0.938)**	$\mathbf{98.5_2}$	**<0.0001**
Day on rack	**0.061 (0.04, 0.094)**	$\mathbf{34.9_2}$	**<0.0001**
Cor(*Int*,Day)	0.083 (−0.153, 0.309)	4.23_1	0.12
Cor(AR1)	**0.458 (0.383, 0.523)**	$\mathbf{177.1_1}$	**<0.0001**
Residual	0.26 (0.227, 0.278)	-	-

Importantly, after accounting for the effects described above, there was pronounced temporal autocorrelation of the residuals (Figure 2a), indicating that any given individual exhibited similar residual activity rates on concurrent days, relative to the individual's temporal reaction norm. When the effect of autocorrelation was dropped from the model, the variance in temporal slopes across days increased by 34% to 0.082 (from 0.061), the intercept variance increased 6% to 0.71 (from 0.67), and the residual variance decreased 22% to 0.20 (from 0.26).

Mice

The up-selected line ran more on the wheel than the control line and females were more active than males (Table 2). Activity rates increased over the duration of the 21 days. There was again substantial among-individual variance in intercepts and slopes, and these factors again explained most of the variance in the data.

Most notably, and alike the guppy data, there was a strong effect of temporal autocorrelation (Figure 2b). When the effect of autocorrelation was ignored in the model, the variance in temporal slopes across days increased by 23% to 0.055 (from 0.044), the intercept variance increased only slightly to 0.53 (from 0.52), and the residual variance decreased 18% to 0.099 (from 0.12).

Table 2

Shown are all parameter estimates of the mice data, with associated 95% confidence intervals

Parameter	Estimate (95% CI)	$F_{numDF,denDF}$	*P*
Fixed effects			
Intercept	−0.112 (−0.501, 0.277)	$0.32_{1,739}$	0.57
Sex *(ref Female)*	**−0.657 (−1.12, −0.192)**	**$9.5_{1,36}$**	**0.004**
Line	**0.89 (0.425, 1.35)**	**$16.2_{1,36}$**	**0.0003**
Day	**0.14 (0.066, 0.214)**	**$14.2_{1,739}$**	**0.0002**
	Estimate (95% CI)	χ^2	*P*
Variance-Covariance parameters			
Intercept	**0.518 (0.323, 0.832)**	**82.2_2**	**<0.0001**
Day	**0.044 (0.025, 0.078)**	**35.4_2**	**<0.0001**
Cor(*Int*,Day)	−0.224 (−0.54, 0.147)	1.37_1	0.24
Cor(AR1)	**0.454 (0.363, 0.537)**	**111.9_1**	**<0.0001**
Residual	0.116 (0.099, 0.136)	-	-

Significant parameters are shown in bold.

Random intercept simulation

The simulations reveal the extent to which autocorrelation can bias estimates of *R*. The lack of temporal independence between observations leads to a situation where variance occurring within individuals (residual variance) is misattributed to occurring among individuals, so that among individual variance is overestimated and residual variance is underestimated, thereby increasing *R* (Figure 3). Therefore, while variance is misattributed between the residual variance and random intercept variance, the total variance (i.e., "$\sigma^2_{ID} + \sigma^2_{\varepsilon}$") remains constant. The bias decreases when considering longer time series (i.e., longer total length of the experiment), though the effect of autocorrelation was persistent when AR1 was high (Figure 3). Increasing the lag between observations was more effective at reducing the bias: at a lag of 5 units of time there was no remaining bias created when AR1 was 0.2 or 0.4 (Figure 3c), as was expected given the correlations approached 0 at this lag (i.e., $r^5 \approx 0$). These patterns are highly consistent across a range of *R* values (Supplementary Material 2, Figure S3).

Random slope simulation

The presence of autocorrelation elevated the estimated variance in the slopes when not accounted for in the statistical model (Figure 4a). This was due to the pseudoreplicated observations at the extreme ends of the time considered, leading to greater leverage. The change in the slope variance was additive, with the bias created being uniform regardless of the slope variance (see Supplementary Material 2). There was an additional large effect on the residual variance, with high AR leading to a large underestimation of the residual variance that persisted for a long duration (Figure 4a).

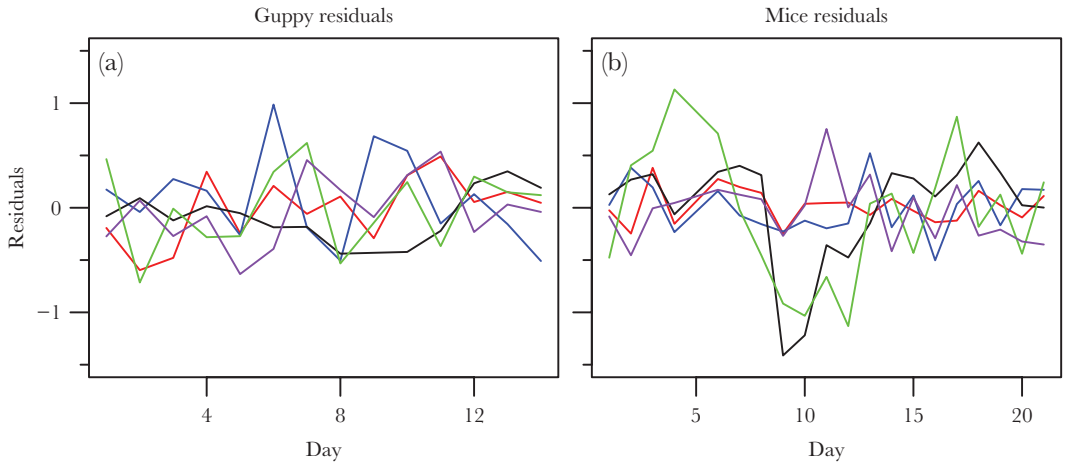

Figure 2

Displayed are residuals of five example individuals from each dataset, with (a) showing the residuals of randomly chosen individuals from the guppy data (AR(1): $r = 0.46$) and (b) residuals from the mice data (AR(1): $r = 0.45$).

DISCUSSION

Here, we present results showing substantial autocorrelation exists among samples of activity rates taken on successive days, for two disparate model species held under controlled lab conditions, and after accounting for individual variation in temporal change. In each of the two datasets, observations of activity rates spaced one day apart were correlated (approximately 0.4). Further, we show that ignoring autocorrelation leads to exaggerations of individual differences, both in mean and temporal reaction norms, due to misestimating individual predicted mean values, in both the empirical data sets and via simulations. The biases in estimates of variance parameters may in turn affect other parameters of biological significance, such as estimates of trait correlations and reaction norm variances, which we discuss below.

Temporal autocorrelation is rarely considered in the study of individual variation in labile traits. The few studies which have quantified this source of variation in behavior have been field studies (Westneat et al. 2011; Nakayama et al. 2016; Villegas-Ríos et al. 2018), where it was unclear to what degree autocorrelation results from intrinsic versus extrinsic factors. By contrast, we show substantial temporal autocorrelation in behavioral trait variation under lab conditions, where external environmental factors were largely stable. Therefore, temporal autocorrelation likely has a substantial intrinsic component, and may yield novel insights into the proximate constraints of behavioral and physiological variation.

Biological insights from autocorrelation

The presence of temporal autocorrelation could reveal important biological processes affecting behavior. Indeed, some theoretical models of behavioral change explicitly predict change in behavior to be an updating process (Stamps and Krishnan 2014; Sih et al. 2015; Stamps and Krishnan 2017). Recent models of developmental plasticity are based on a Bayesian updating approach, whereby an animal begins development with a prior "belief" of the state of the environment. This "belief" is then sequentially updated by experience, so that the "belief" after event one becomes the prior to event two (Stamps and Frankenhuis 2016).

Similarly, behavioral updating has been discussed in the context of "state-behavior feedback" effects (reviewed by Sih et al. 2015). Here, a behavior is expected to change the internal state of an animal, which is then said to feedback to affect the expression of the behavior (Nathan et al. 2008; Sih et al. 2015). For instance, having low energy reserves should increase the foraging activity rate until the reserves are replenished, at which point foraging rates would decline. Thus, current behavior is a function of past behavior, mediated by internal state. While these two examples of behavioral change as an updating process have focused on their effect on among-individual (or genotypic) variance (Stamps and Krishnan 2014; Sih et al. 2015), updating predicts no particular shape to the temporal trajectories. Rather, these processes predict a correlated random walk, with future behavior predicted by recent past behavior.

In this study, fish were fed a set period before observations, while mice were fed *ad libitum*, so energy availability is unlikely to play as important a role. A possible cause for the autocorrelation in these datasets is a training effect, which would be predicted to lead to a positive state-behavior feedback loop. We might predict mice which are most active to most quickly invest in relevant muscles and metabolic pathways (Houle-Leroy et al. 2000; Girard and Garland 2002), and most quickly increase their familiarity with the previously novel running-wheel. However, a positive feedback loop would also predict a positive intercept–slope covariance among individuals, which we did not observe in either case. This does not rule out the possibility of an effect of fatigue, which may lead to autocorrelation and a negative feedback loop predicting no intercept–slope covariance. Autocorrelation could also result from (unspecified) nonlinear temporal trends in some, or all, individuals. Intrinsic factors, such as estrus, are known to cycle through time and affect activity rates in female mice (Gerall et al. 1973). This could create a cycling effect over a 3–4-day period, though we detected no effect of sex. Female guppies undergo a gestation cycle, which could create a cycling effect of behavior over the course of a 28-day period. However, rather than the random walk assumed by autoregressive structures (see statistical methods and Figure 1), these effects would yield firm predictions of the period and shape of the cycle in the trait of interest (Fanson et al. 2014), and may be better specified as nonlinear reaction norms. The nonlinearity of the mean trajectory should be checked visually by plotting of individual scores through time, though plots revealed no clear patterns in these data.

Autocorrelation could also yield important insights to the understanding of behavioral predictability, which aims to understand potential causes of residual intraindividual variability (Stamps et al.

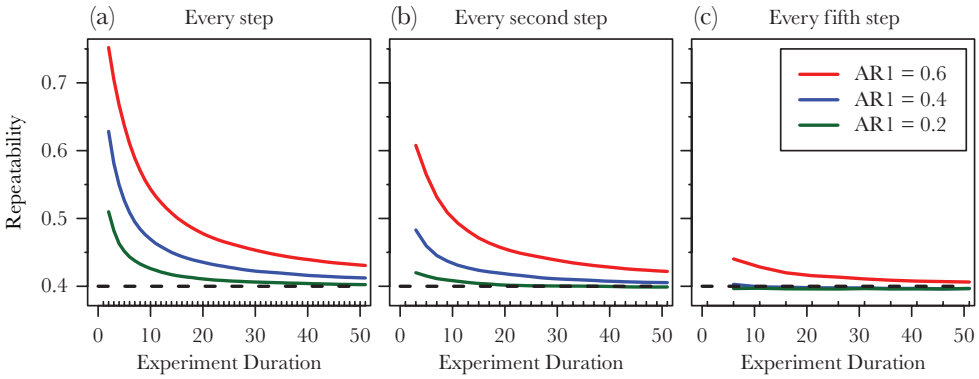

Figure 3

Displayed is the effect of undetected temporal autocorrelation on estimates of repeatability (R) under different sampling regimes. The dashed line shows the true R value of 0.4, and the upward ticks on the *x* axis depict when a sample was considered. Panel a depicts R estimates derived from sampling every step, up to 51 steps, panel b shows a sampling frequency of every second step, up to 26 observations, and panel c shows a sampling frequency of every fifth step up to 11 observations. Effects of using different underlying simulated R values yielded very similar results that are shown in the supplement.

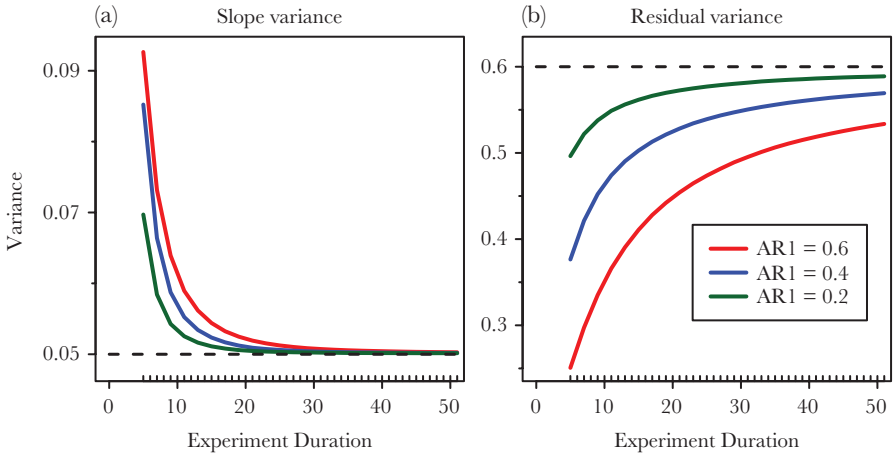

Figure 4
Displayed is the effect of undetected temporal autocorrelation on estimates of random slope variance (a) and residual variance (b). The dashed line shows the true simulated slope variance, and the upward ticks on the *x* axis depict when a sample was considered.

2012; Westneat et al. 2015). To study predictability, analyses typically focus on quantifying individual-specific residual variances (Stamps et al. 2012; Biro and Adriaenssens 2013; Briffa 2013), with high variance individuals said to be of low predictability. However, the existence of temporal autocorrelation implies that individuals are more predictable over short periods of time (Wittemyer et al. 2008) as recent behavior is likely a better predictor of future behavior than the overall variance. Thus, it is critical to acknowledge over what period behavioral predictability is important. For instance, a predator is unlikely to encounter and recognize an individual prey after a long duration and thus predictability is likely not important over this period of time for predator–prey interactions. Conversely, repeated individual predator–prey interactions are more likely over short time periods, when autocorrelation is likely to lead to increased predictability. As such, future work on behavioral predictability should consider the effect of environmental factors and biotic interaction on temporal autocorrelation. These effects can be evaluated using heterogenous autocorrelation structures across treatments or environments to yield valuable biological information (Wittemyer et al. 2008; Boyce et al. 2010).

How autocorrelation may bias analyses

Our simulations demonstrate how temporal autocorrelation can inflate *R* and random slope variances. This occurs to the greatest extent when observations are taken at short temporal lags and the total duration of the experiment is short (Figure 3), and these results appear consistent with the empirical data which showed modest effects across the 14-day and 21-day observation periods. The upward bias in *R* is due to a misestimation of individual predicted means, leading to residual variance being erroneously attributed to occur among individuals. The effect of this is that the extent to which individuals differ, and their consistency through time are overestimated. This may mean that the extent to which individuals differ in their behavior, as estimated by *R*, and the inferences to heritability and selectability of behavioral traits may be overstated.

In turn, the misestimation of individual-specific mean values due to autocorrelated (pseudoreplicated) observations will lead to problems when relating those predicted mean values to other factors of interest. Large bodies of theory predict among-individual trait covariances due to underlying proximate constraints, for instance along life-history gradients (Réale et al. 2010) or cognition (Griffin et al. 2015). If ignored, autocorrelation may also affect the analysis of such covariances, as the estimates of the among-individual correlation will be confounded by the within-individual correlation, in ways analogous to studies which lack repeated measures (Brommer 2013). The effect will be dependent on the lability of the second trait. When the correlation is between a labile, autocorrelated trait and a nonlabile trait (e.g., morphology), the inflated misestimations of individual means will be a result of random noise, which would bias the correlation coefficient towards 0.

Conversely, analyses often aim to estimate the among-individual correlations in behaviors (i.e., behavioral syndromes) or correlations with other labile traits such as metabolic or endocrine traits (Careau et al. 2014; Nakayama et al. 2016; Royauté et al. 2018). Such analyses typically aim to decipher the among-individual correlation (due to intrinsic differences) from the within-individual correlations (Dingemanse and Dochtermann 2013; Careau et al. 2014). If observations are clustered together in time, among-individual variance is confounded by within-individual variance (as above), the covariance between traits will then depend on the within-individual correlation; so that individual correlation will be estimated to be similar to the residual correlation.

Temporal autocorrelation may also lead to biases in analyses of individual plasticity. Researchers are often interested in individual differences in temporal plasticity, to understand how individuals differ in their learning, habituation or acclimation rates (Martin and Réale 2008; Biro 2012), ontogenetic trajectories (Biro et al. 2014; Brommer and Class 2015), or seasonality (Carter et al. 2012). However, if data are temporally autocorrelated, one could be led to the spurious or exaggerated conclusion that individuals differ in these temporal trajectories due to a lack of independence of observations in close temporal proximity. Indeed, this effect was evident in both empirical datasets and a simulation. Autocorrelation upwardly biased the slope variance in the simulation, which was additive and independent of the true slope variance. This effect will have occurred due to the statistical "leverage" of pseudoreplicated observations at each end of the time series for any given individual.

While we kept the analysis of the simulation simple to focus on first-order autocorrelation, as evident in the two case studies, field studies have shown temporal autocorrelation can persist for

protracted periods of time. In an analysis of the activity rates of Eurasian perch (*Perca fluviatilis*), autocorrelation terms up to the sixth order (6-day lag) were shown to be significant and positive (Nakayama et al. 2016). Higher order auto-regression (AR) terms act additively to the lower order AR predictions. Thus, where higher order autocorrelation exists, the bias affecting R estimates (Figure 3) would operate over prolonged time periods. For instance, using the best estimates of AR to the seventh order reported by Nakayama et al. (2016) would predict observations taken 14 days apart to remain correlated at $r > 0.2$. Thus, under a fortnightly sampling regime, we would predict bias to be created in accordance with Figures 3a and 4 (green line).

In addition to biasing estimates, autocorrelation will also affect required sampling rates of traits. The inclusion of an AR structure suggests that the effective number of repeated measures will be reduced, due to the lack of independence between observations (i.e., pseudoreplication). Thus, the denominator degrees of freedom of tests of fixed effects will be overestimated if ignored. Therefore, even greater sampling rates will be required when autocorrelation is present to meet the levels prescribed by analyses of the power and precision when estimating individual variation using mixed effect models (e.g., Wolak et al. 2012; Dingemanse and Dochtermann 2013)

Modeling options for time-series

Here we focused on the simple case of an AR1 process in our models, and have briefly discussed higher order autocorrelation (e.g., AR2). However more complex autocorrelation structures are sometimes required to accurately model the residuals through time (Zuur et al. 2009; Box et al. 2016). The fitting of more complex models should be preceded by a comprehensive exploration of the data's autocorrelation structure by a combination of calculating and visualizing the autocorrelation and partial autocorrelation functions (see the "acf" function in the "stats" package of "R"; R Core Team 2017) which will inform modeling choices. Similarly, modeling individual or treatment specific autocorrelation coefficients should also be motivated by a priori biological predictions where possible.

To confidently fit an AR model to the residuals, both the mean (ideally equal to zero) and the variance of the residuals should remain constant for the duration of the experiment (Box et al. 2016). If these assumptions are violated, a gradual change in the mean over time can be modeled by incorporating a moving average term in the model equation, which then becomes an autoregressive moving average model. It should also be noted that if cyclic trends are detected (e.g., due to a biological cycle, Boyce et al. 2010; Fanson et al. 2014), the residuals may be modeled by fitting a harmonic regression (Boyce et al. 2010).

The case of an irregularly sampled dataset adds another layer of complexity as the different parameters of the autoregressive model have to be estimated by fitting a stochastic differential equation to the data (Brockwell 2001; see also corCAR1 in "nlme" Pinheiro et al. 2017), and therefore, autocorrelation should be considered when designing the sampling regime. Further, we warn that focusing on fitting a complex time-series model to the residuals may also become a red herring, as the assumptions underlying these parameters may not always be biologically realistic and may lead to convergence issues (Bates et al. 2015). Therefore, prior to fitting an autoregressive model, it is often advisable to check for explanatory variables (e.g., different shapes of reaction norms Fanson et al. 2014; Morrissey and Liefting 2016), which may have been omitted from the model (Zuur et al. 2009).

Recommendations for sampling autocorrelated traits

It is of course difficult to make firm recommendations for how to deal with potential autocorrelation in sampling designs and the best practice will naturally depend on the system, questions of interest, and constraints of the project. Generally, the best practice will be to acquire enough data to quantify and statistically account for autocorrelation. However, due to expense, time constraints, and potential animal ethics concerns of collecting sufficient and frequent enough repeated samples, this often will not be possible. Under such constraints, the most apparent and effective means to deal with autocorrelation is to take samples at infrequent time intervals. By observing animals at five lag-steps (in our case, 5 days), the simulations confirmed that the bias due to autocorrelation would have dropped to near 0. However, we caution that the extent of autocorrelation may extend over greater periods of time if higher order AR terms are present (e.g., Nakayama et al. 2016).

Another way to sample labile traits to assess temporal stability is through a multiple burst sampling design, where intensive intervals of repeated observation are interspersed with periods of rest (Stamps et al. 2012; Biro and Adriaenssens 2013). This also allows us to appraise the temporal dependence of observations taken in close proximity (Araya-Ajoy et al. 2015; Mitchell et al. 2016; Araya-Ajoy and Dingemanse 2017; Mitchell and Biro 2017) to assess potential temporal dependence of observations. In contrast to autocorrelation, multiple burst analyses fit an intercept of a burst of observations taken in close proximity, nested within individual (Araya-Ajoy et al. 2015). Together, this yields an estimate of the long-term repeatability, analogous to the unbiased R estimate here, and a short-term repeatability estimate, similar to an R estimate with a degree of unaccounted for autocorrelation. Such a burst sampling design also allows for replication of individual reaction norms, even when the gradient is inherently temporal (e.g., food deprivation) (Araya-Ajoy et al. 2015; Mitchell and Biro 2017).

Concluding remark

Here we have shown that activity rates in two commonly used lab species can be substantially temporally autocorrelated. This autocorrelation was present despite observations being taken under highly controlled situations and after accounting for individual differences in temporal trajectories. Thus, autocorrelation appears to reflect an intrinsic state of individuals and may prove insightful in elucidating proximate constraints to behavioral, among other labile traits. Further, the existence of autocorrelation in studies of individual variation in behavior is rarely considered, though can significantly affect how we analyze and interpret our data.

SUPPLEMENTARY MATERIAL

Supplementary data are available at *Behavioral Ecology* online.

We wish to thank Theodore Garland Jr. for access to the mice data used, and David Westneat and a second anonymous reviewer for constructive reviews. Equipment for the fish data was funded by an ARC Discovery grant awarded to P.A.B.

Author contributions: D.J.M. carried out guppy data collection, analysis, simulation of data and wrote the first draft. Ideas were conceived by D.J.M., and refined with P.A.B. and C.B. A.M.D. advised on the analyses and simulation and wrote the modeling options section. All authors further revised the manuscript.

Data accessibility: Analyses reported in this article can be reproduced using the data provided by Mitchell et al. (2019).

Handling editor: Niels Dingemanse

REFERENCES

Allegue H, Araya-Ajoy YG, Dingemanse NJ, Dochtermann NA, Garamszegi LZ, Nakagawa S, Réale D, Schielzeth H, Westneat DF. 2017. Statistical Quantification of Individual Differences (SQuID): an educational and statistical tool for understanding multilevel phenotypic data in linear mixed models. Meth Ecol Evol. 8:257–267.

do Amaral JPS, Marvin Glenn A, Hutchison Victor H. 2002. The influence of bacterial lipopolysaccharide on the thermoregulation of the box turtle *Terrapene carolina*. Physiol Biochem Zool. 75:273–282.

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

Araya-Ajoy YG, Mathot KJ, Dingemanse NJ. 2015. An approach to estimate short-term, long-term, and reaction norm repeatability. Meth Ecol Evol. 6:1462–1473.

Bates D, Kliegl R, Vasishth S, Baayen H. 2015. Parsimonious mixed models. arXiv. doi: arXiv:1506.04967

Bell C. 1980. The scaling of the thermal inertia of lizards. J Exp Biol. 86:79–85.

Bell AM, Hankison SJ, Laskowski KL. 2009. The repeatability of behaviour: a meta-analysis. Anim Behav. 77:771–783.

Bierbach D, Laskowski KL, Wolf M. 2017. Behavioural individuality in clonal fish arises despite near-identical rearing conditions. Nat Commun. 8:15361.

Biro PA. 2012. Do rapid assays predict repeatability in labile (behavioural) traits? Anim Behav. 83:1295–1300.

Biro PA, Adriaenssens B. 2013. Predictability as a personality trait: consistent differences in intraindividual behavioral variation. Am Nat. 182:621–629.

Biro PA, Adriaenssens B, Sampson P. 2014. Individual and sex-specific differences in intrinsic growth rate covary with consistent individual differences in behaviour. J Anim Ecol. 83:1186–1195.

Biro P, Garland Jr T, Post J, Ujvari B, Thomas F, Beckmann C. 2018. Data from: Metabolic scope as a proximate constraint on individual behavioral variation: effects on 'personality', plasticity, and predictability. Dryad Data Repository. http://dx.doi.org/10.5061/dryad.jn33140

Biro PA, Stamps JA. 2015. Using repeatability to study physiological and behavioural traits: ignore time-related change at your peril. Anim Behav. 105:223–230.

Box GEP, Jenkins GM, Reinsel GC. 2016. Time series analysis: forecasting and control, 5th ed. Hoboken (NJ): John Wiley & Sons.

Boyce MS, Pitt J, Northrup JM, Morehouse AT, Knopff KH, Cristescu B, Stenhouse GB. 2010. Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data. Philos Trans R Soc Lond B Biol Sci. 365:2213–2219.

Briffa M. 2013. Plastic proteans: reduced predictability in the face of predation risk in hermit crabs. Biol Lett. 9:20130592.

Briffa M, Bridger D, Biro PA. 2013. How does temperature affect behaviour? Multilevel analysis of plasticity, personality and predictability in hermit crabs. Anim Behav. 86:47–54.

Brockwell P. 2001. Continuous-time ARMA processes. Handbook of Statistics. 19:249–276.

Brommer JE. 2013. On between-individual and residual (co)variances in the study of animal personality: are you willing to take the "individual gambit"? Behav Ecol Sociobiol. 67:1027–1032.

Brommer JE, Class B. 2015. The importance of genotype-by-age interactions for the development of repeatable behavior and correlated behaviors over lifetime. Front Zool. 12 (Suppl 1):S2.

Careau V, Hoye BJ, O'Dwyer TW, Buttemer WA. 2014. Among- and within-individual correlations between basal and maximal metabolic rates in birds. J Exp Biol. 217:3593–3596.

Carter A, Goldizen A, Heinsohn R. 2012. Personality and plasticity: temporal behavioural reaction norms in a lizard, the Namibian rock agama. Anim Behav. 84:471–477.

Dingemanse NJ, Bouwman KM, van de Pol M, van Overveld T, Patrick SC, Matthysen E, Quinn JL. 2012. Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. J Anim Ecol. 81:116–126.

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

Dohm MR. 2002. Repeatability estimates do not always set an upper limit to heritability. Funct Ecol. 16:273–280.

Dray S, Royer-Carenzi M, Calenge C. 2010. The exploratory analysis of autocorrelation in animal-movement studies. Ecol Res. 25:673–681.

Eisenmann JC, Wickel EE, Kelly SA, Middleton KM, Garland T Jr. 2009. Day-to-day variability in voluntary wheel running among genetically differentiated lines of mice that vary in activity level. Eur J Appl Physiol. 106:613–619.

Falconer DS. 1981. Introduction to quantitative genetics. 2nd ed. London (UK): Longman.

Fanson KV, Biro PA. 2019. Meta-analytic insights into factors influencing the repeatability of hormone levels in agricultural, ecological, and medical fields. Am J Physiol Regul Integr Comp Physiol. 316:R101–R109.

Fanson KV, Keeley T, Fanson BG. 2014. Cyclic changes in cortisol across the estrous cycle in parous and nulliparous Asian elephants. Endocr Connect. 3:57–66.

Gerall AA, Napoli AM, Cooper UC. 1973. Daily and hourly estrous running in intact, spayed, and estrone implanted rats. Physiol Behav. 10:225–229.

Girard I, Garland T Jr. 2002. Plasma corticosterone response to acute and chronic voluntary exercise in female house mice. J Appl Physiol (1985). 92:1553–1561.

Griffin AS, Guillette LM, Healy SD. 2015. Cognition and personality: an analysis of an emerging field. Trends Ecol Evol. 30:207–214.

Holtmann B, Lagisz M, Nakagawa S, Moore I. 2016. Metabolic rates, and not hormone levels, are a likely mediator of between-individual differences in behaviour: a meta-analysis. Funct Ecol. 31:685–696.

Houle-Leroy P, Garland T Jr, Swallow JG, Guderley H. 2000. Effects of voluntary activity and genetic selection on muscle metabolic capacities in house mice *Mus domesticus*. J Appl Physiol (1985). 89:1608–1616.

Houslay TM, Vierbuchen M, Grimmer AJ, Young AJ, Wilson AJ. 2018. Testing the stability of behavioural coping style across stress contexts in the Trinidadian guppy. Funct Ecol. 32:424–438.

Kain MP, Bolker BM, McCoy MW. 2015. A practical guide and power analysis for GLMMs: detecting among treatment variation in random effects. PeerJ. 3:e1226.

Krause JS, Pérez JH, Meddle SL, Wingfield JC. 2017. Effects of short-term fasting on stress physiology, body condition, and locomotor activity in wintering male white-crowned sparrows. Physiol Behav. 177:282–290.

Martin JGA, Réale D. 2008. Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. Anim Behav. 75:309–318.

Mitchell DJ, Biro PA. 2017. Is behavioural plasticity consistent across different environmental gradients and through time? Proc R Soc B. 284:20170893.

Mitchell DJ, Fanson BG, Beckmann C, Biro PA. 2016. Towards powerful experimental and statistical approaches to study intraindividual variability in labile traits. R Soc Open Sci. 3:160352.

Mitchell D, Dujon A, Beckmann C, Biro P. 2019. Data from: Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity. Behav Ecol. doi:10.5061/dryad.vk8dk4k

Morrissey MB, Liefting M. 2016. Variation in reaction norms: statistical considerations and biological interpretation. Evolution. 70:1944–1959.

Nakayama S, Laskowski KL, Klefoth T, Arlinghaus R. 2016. Between- and within-individual variation in activity increases with water temperature in wild perch. Behav Ecol. 27:1676–1683.

Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE. 2008. A movement ecology paradigm for unifying organismal movement research. Proc Natl Acad Sci USA 105:19052–19059.

Niemelä PT, Dingemanse NJ. 2017. Individual versus pseudo-repeatability in behaviour: lessons from translocation experiments in a wild insect. J Anim Ecol. 86:1033–1043.

Pinheiro J, Bates D, DebRoy S, Sarkar D. 2017. {nlme}: linear and nonlinear mixed effects models. R package version 3.1-131.

Pruitt JN, Demes KW, Dittrich-Reed DR. 2011. Temperature mediates shifts in individual aggressiveness, activity level, and social behavior in a spider. Ethology. 117:318–325.

R Core Team. 2017. R: a language and environment for statistical computing. Version 3.3.3. Vienna, Austria: R Foundation for Statistical Computing.

Réale D, Garant D, Humphries MM, Bergeron P, Careau V, Montiglio PO. 2010. Personality and the emergence of the pace-of-life syndrome concept at the population level. Philos Trans R Soc Lond B Biol Sci. 365:4051–4063.

Romero LM, Dickens MJ, Cyr NE. 2009. The Reactive Scope Model - a new model integrating homeostasis, allostasis, and stress. Horm Behav. 55:375–389.

Royauté R, Berdal MA, Garrison CR, Dochtermann NA, 2018. Paceless life? A meta-analysis of the pace-of-life syndrome hypothesis. Behav Ecol Sociobiol. 72:64.
Schoenemann KL, Bonier F. 2018. Repeatability of glucocorticoid hormones in vertebrates: a meta-analysis. PeerJ. 6:e4398.
Sih A, Mathot KJ, Moirón M, Montiglio PO, Wolf M, Dingemanse NJ. 2015. Animal personality and state-behaviour feedbacks: a review and guide for empiricists. Trends Ecol Evol. 30:50–60.
Stamps JA. 2016. Individual differences in behavioural plasticities. Biol Rev Camb Philos Soc. 91:534–567.
Stamps JA, Briffa M, Biro PA. 2012. Unpredictable animals: individual differences in intraindividual variability (IIV). Anim Behav. 83:1325–1334.
Stamps JA, Frankenhuis WE. 2016. Bayesian Models of Development. Trends Ecol Evol. 31:260–268.
Stamps JA, Krishnan VV. 2014. Combining information from ancestors and personal experiences to predict individual differences in developmental trajectories. Am Nat. 184:647–657.
Stamps JA, Krishnan VV. 2017. Age-dependent changes in behavioural plasticity: insights from Bayesian models of development. Anim Behav 126:53–67.
Villegas-Ríos D, Réale D, Freitas C, Moland E, Olsen Esben M. 2018. Personalities influence spatial responses to environmental fluctuations in wild fish. J Anim Ecol. 87:1309–1319.
Westneat DF, Hatch MI, Wetzel DP, Ensminger AL. 2011. Individual variation in parental care reaction norms: integration of personality and plasticity. Am Nat. 178:652–667.
Westneat DF, Wright J, Dingemanse NJ. 2015. The biology hidden inside residual within-individual phenotypic variation. Biol Rev Camb Philos Soc. 90:729–743.
White CR, Schimpf NG, Cassey P. 2013. The repeatability of metabolic rate declines with time. J Exp Biol. 216:1763–1765.
Wilson AJ. 2018. How should we interpret estimates of individual repeatability? Evol Lett. 2:4–8.
Wittemyer G, Polansky L, Douglas-Hamilton I, Getz WM. 2008. Disentangling the effects of forage, social rank, and risk on movement autocorrelation of elephants using Fourier and wavelet analyses. Proc Natl Acad Sci USA 105:19108–19113.
Wolak ME, Fairbairn DJ, Paulsen YR. 2012. Guidelines for estimating repeatability. Meth Ecol Evol. 3:129–137.
Zsebők S, Herczeg G, Blázi G, Laczi M, Nagy G, Szász E, Markó G, Török J, Garamszegi LZ. 2017. Short- and long-term repeatability and pseudo-repeatability of bird song: sensitivity of signals to varying environments. Behav Ecol Sociobiol 71:154.
Zuur A, Ieno EN, Walker N, Saveliev AA, Smith GM, 2009. Mixed effects models and extensions in ecology with R. New York (NY): Springer.